


A global view of aging and Alzheimer's pathogenesis-associated cell population dynamics and molecular signatures in human and mouse brains

Received: 16 December 2022

Accepted: 17 October 2023

Published online: 30 November 2023

 Check for updates

Andras Sziraki^{1,2,8}, Ziyu Lu^{1,2,8}, Jasper Lee^{1,8}, Gabor Banyai¹, Sonya Anderson³, Abdulraouf Abdulraouf^{1,4}, Eli Metzner^{1,5}, Andrew Liao^{1,4}, Jason Banfelder⁶, Alexander Epstein^{1,2}, Chloe Schaefer¹, Zihan Xu^{1,2}, Zehao Zhang^{1,2}, Li Gan⁷, Peter T. Nelson³, Wei Zhou^{1,9}✉ & Junyue Cao^{1,9}✉

Conventional methods fall short in unraveling the dynamics of rare cell types related to aging and diseases. Here we introduce EasySci, an advanced single-cell combinatorial indexing strategy for exploring age-dependent cellular dynamics in the mammalian brain. Profiling approximately 1.5 million single-cell transcriptomes and 400,000 chromatin accessibility profiles across diverse mouse brains, we identified over 300 cell subtypes, uncovering their molecular characteristics and spatial locations. This comprehensive view elucidates rare cell types expanded or depleted upon aging. We also investigated cell-type-specific responses to genetic alterations linked to Alzheimer's disease, identifying associated rare cell types. Additionally, by profiling 118,240 human brain single-cell transcriptomes, we discerned cell- and region-specific transcriptomic changes tied to Alzheimer's pathogenesis. In conclusion, this research offers a valuable resource for probing cell-type-specific dynamics in both normal and pathological aging.

Progressive changes in brain cell populations, which can occur during aging, may contribute to functional decline and increased risks for neurodegenerative diseases such as Alzheimer's disease (AD)^{1–4}. Although the recent advances in single-cell genomics have created unprecedented opportunities to explore the cell-type-specific dynamics across the entire mammalian brain^{5–8}, most prior studies relied on a relatively shallow sampling of the brain cell populations and failed to reveal rare aging or AD-associated cell types.

Additionally, they were technically limited in several ways, including failing to recover isoform-level gene expression patterns and the associated chromatin landscape that regulates cell-type-specific alterations across aging stages.

Here, we introduced EasySci, a cost-effective single-cell profiling strategy based on extensive optimization of single-cell RNA sequencing (RNA-seq) by combinatorial indexing⁹. While the original method has been widely used to study embryonic and fetal tissues^{10,11}, it remains

¹Laboratory of Single Cell Genomics and Population Dynamics, The Rockefeller University, New York, NY, USA. ²The David Rockefeller Graduate Program in Bioscience, The Rockefeller University, New York, NY, USA. ³Department of Pathology and Sanders-Brown Center on Aging, University of Kentucky, Lexington, KY, USA. ⁴The Tri-Institutional MD-PhD Program, New York, NY, USA. ⁵The Tri-Institutional PhD Program in Computational Biology and Medicine, New York, NY, USA. ⁶High Performance Computing Resource Center, The Rockefeller University, New York, NY, USA. ⁷Helen and Robert Appel Alzheimer's Disease Research Institute, Weill Cornell Medicine, New York, NY, USA. ⁸These authors contributed equally: Andras Sziraki, Ziyu Lu, Jasper Lee. ⁹These authors jointly supervised this work: Wei Zhou, Junyue Cao. ✉e-mail: wzhou@rockefeller.edu; jcao@rockefeller.edu

restricted to gene quantification proximal to the 3' end and limited in efficiency and cell recovery rate¹¹. EasySci provided improved conditions for cell lysis, fixation, sample preservation, enzymatic reaction, oligonucleotide design, and purification methodologies (Supplementary Table 1). Several test conditions were inspired by optimizations described in recently developed or optimized single-cell techniques^{12,13}. The major features of EasySci include (i) 1 million single-cell transcriptomes were prepared for ~US \$700 (library preparation cost only, not including personnel or sequencing cost; Fig. 1a–c); (ii) reverse transcription (RT) with indexed oligo-dT and random hexamer primers was achieved, thus recovering cell-type-specific gene expression with full gene body coverage (Fig. 1d); (iii) cell recovery rate, as well as the number of transcripts detected per cell, were substantially improved through optimized nuclei storage, enzymatic reactions and improved primer design (Fig. 1e and Extended Data Fig. 1); and (iv) an extensively improved single-cell data processing pipeline was developed for both gene counting and exonic counting using paired-end single-cell RNA-seq data (Methods).

Leveraging the technical innovations during the development of EasySci-RNA, we further optimized the single-cell chromatin accessibility profiling method by combinatorial indexing (sci-ATAC-seq3)^{14,15}. The key optimizations include (i) a tagmentation reaction with indexed Tn5 that are fully compatible with indexed ligation primers of EasySci-RNA; (ii) a modified nuclei extraction and cryostorage procedure to further increase the library complexity. (A comprehensive quality comparison with other single-cell sequencing assay for transposase-accessible chromatin (scATAC) protocols is shown in Extended Data Fig. 2.) It is noteworthy that the assay for transposase-accessible chromatin with sequencing (ATAC-seq) signal specificity of EasySci-ATAC parallels the original sci-ATAC-seq^{14,15}, albeit lower than 10x ATAC-seq, potentially due to the indexed Tn5 used in single-cell combinatorial indexing. The detailed protocols for EasySci are included as supplementary files (Supplementary Protocols 1 and 2) to facilitate individual laboratories to cost-efficiently generate gene expression and chromatin accessibility profiles from millions of single cells.

Results

A single-cell catalog of the mouse brain in aging and AD

We first applied EasySci to characterize cell-type-specific gene expression, and chromatin accessibility profiles across the entire mouse brain sampling at different ages, sexes and genotypes (Fig. 1f). We collected C57BL/6 wild-type (WT) mouse brains at 3 months ($n = 4$), 6 months ($n = 4$) and 21 months ($n = 4$). To gain insight into the early molecular changes associated with the pathophysiology of AD, two mutants from the same C57BL/6 background at 3 months were included: an early-onset AD (EOAD) model (5x*FAD*) that overexpresses mutant human amyloid-beta precursor protein and human presenilin

1 harboring multiple AD-associated mutations¹⁶; and a late-onset AD (LOAD) model (APOE*4/*Trem2**R47H) that carries two of the highest risk factor mutations of LOAD, including a humanized ApoE knock-in allele and missense mutations in the mouse *Trem2* gene^{17,18}.

In brief, nuclei were extracted from the whole brain and then deposited to different wells for indexed RT (RNA) or transposition (ATAC), such that the first index indicated the originating sample and assay type of any given well. The resulting EasySci libraries (RNA and ATAC) were sequenced separately, yielding a total of 20 billion paired-end reads. After filtering out low-quality cells and doublets, we recovered gene expression profiles in 1,469,111 single nuclei (a median of 70,589 nuclei per brain sample; Extended Data Fig. 3a) and chromatin accessibility profiles in 376,309 single nuclei (a median of 18,112 nuclei per brain sample, Extended Data Fig. 3b) across conditions. Despite shallow sequencing depth (~4,340 and ~16,000 raw reads per cell for RNA and ATAC, respectively), we recovered an average of 1,788 unique molecular identifiers (UMIs) (RNA, median of 935 UMIs) and 5,515 unique fragments (ATAC, median of 3,918) per nucleus (Extended Data Fig. 3c–f), comparable to other published datasets^{10,11,14}.

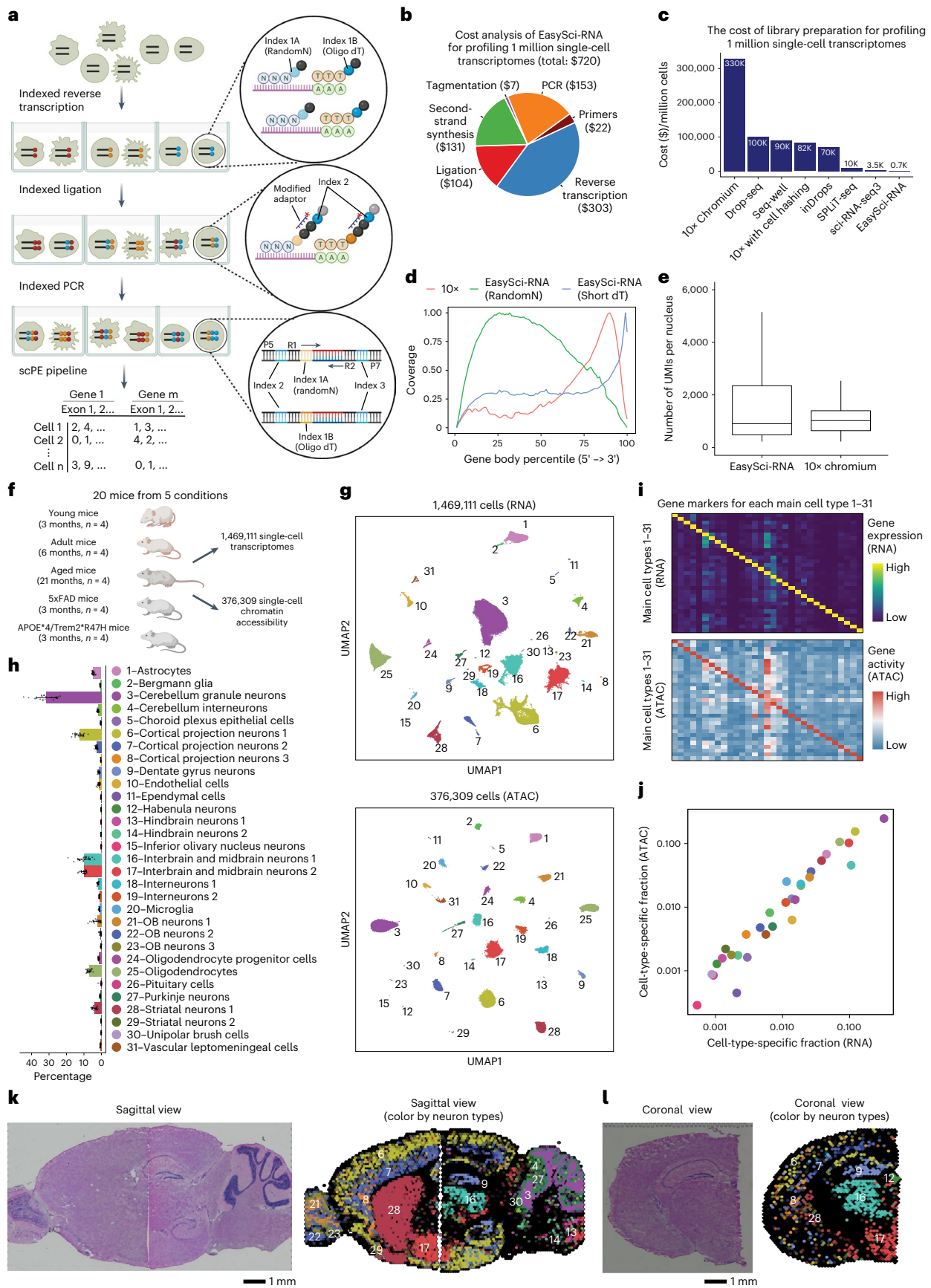
With UMAP visualization¹⁹ and Louvain clustering²⁰, we identified 31 main cell types by gene expression clusters (a median of 16,370 cells per cell type; Fig. 1g), annotated based on cell-type-specific gene markers². Each cell type was present in nearly all individuals, except for rare pituitary cells (0.09% of the population), which were absent in 3 out of 20 individuals (Extended Data Fig. 3g). The cell-type-specific fractions in the global cell population ranged from 0.05% (inferior olivary nucleus neurons) to 32.5% (cerebellum granule neurons) (Fig. 1h). An average of 74 marker genes were identified for each main cell type (defined as at least a twofold expression difference between first- and second-ranked cell types; false discovery rate (FDR) of 5%; and transcripts per million (TPM) > 50 in the target cell type; Supplementary Table 2). In addition to the established marker genes, we identified novel markers that were not previously associated with the respective cell types, such as markers for microglia (e.g., *Arhgap45* and *Wdfy4*), astrocytes (e.g., *Celrr* and *Adamts9*) and oligodendrocytes (e.g., *Sec14l5* and *Galnt5*) (Extended Data Fig. 3h).

Several integration analyses were performed to validate the recovered cell types across different layers. First, we applied a deep-learning-based strategy²¹ to integrate transcriptome and chromatin accessibility profiles, yielding 31 main cell types (Fig. 1g). The gene body accessibility and expression of marker genes across cell types were highly correlated (Fig. 1i), as well as the fraction of each cell type (Pearson correlation $r = 0.95$, $P = 6.68 \times 10^{-16}$) (Fig. 1j). We further investigated the epigenetic controls of the diverse brain cell types through differential accessibility analysis (Extended Data Fig. 4a). We identified a median of 474 differential accessible peaks per cell type (FDR of 5%, TPM > 20 in the target cell type; Extended Data Fig. 4b,c

Fig. 1 | EasySci enables high-throughput and low-cost single-cell transcriptome and chromatin accessibility profiling across the entire mammalian brain. a, EasySci-RNA workflow. Key steps are outlined in the texts. scPE, single-cell paired end. b, Pie chart showing the estimated cost compositions of library preparation for profiling 1 million single-nucleus transcriptomes using EasySci-RNA. c, Bar plot comparing different single-cell RNA-seq methods in terms of their cost of the library preparation for 1 million single-nucleus transcriptomes. The cost of the other techniques (10x Genomics, Drop-seq, Seq-well, inDrops, SPLiT-seq) were calculated using data from previous publications^{13,33,83,84}. d, Density plot showing the gene body coverage comparing single-cell transcriptome profiling using 10x Genomics and EasySci-RNA. Reads from oligo-dT and random hexamers priming are plotted separately for EasySci-RNA. Short dT, oligonucleotides composed of a stretch of 15 thymine nucleotides. RandomN, oligonucleotides composed of 8 random nucleotides. e, Box plot showing the number of unique transcripts detected per mouse brain nucleus comparing 10x Genomics v2 (ref. 83) ($n = 5,351$ cells) and an EasySci-RNA library ($n = 13,440$ cells) at similar sequencing depth (~3,800 raw reads per cell).

For the box plot, middle lines represent medians, upper and lower box edges represent first and third quartiles, respectively, and whiskers represent 1.5 times the interquartile range (IQR). f, Experiment scheme to reconstruct a brain cell atlas of both gene expression and chromatin accessibility across different ages, sexes and genotypes. g, UMAP visualization of mouse brain cells by single-cell transcriptome (top) and chromatin accessibility (bottom), colored by main cell types. h, Bar plot showing the mean and standard error of the cell-type-specific proportions of the brain cell population across samples ($n = 20$ animals) profiled by EasySci-RNA. i, Heatmap showing the aggregated gene expression (top) and gene body accessibility (bottom) of the top 10 marker genes (columns) in each main cell type (rows). j, Scatter plot showing the fraction of each cell type in the global brain population by single-cell transcriptome (x axis) or chromatin accessibility analysis (y axis). k, l, Mouse brain sagittal (k) and coronal (l) sections showing the H&E staining (left) and the inferred localizations of main neuron types through non-negative least squares (NNLS)-based integration (right), colored by main cell types in h.

Nature Genetics | Volume 55 | December 2023 | 2104–2116



and Supplementary Table 3). Key cell-type-specific transcription factor (TF) regulators were discovered by correlation analysis between motif accessibility and expression patterns, such as *Spil* in microglia²², *Nr4a2* in cortical projection neurons 3 (ref. 23) and *Pou4f1* in inferior olivary nucleus neurons²⁴ (Extended Data Fig. 4d).

We next integrated our dataset with a 10x Visium spatial transcriptomics dataset through a modified NNLS approach (Methods). As expected, specific brain cell types were mapped to distinct anatomical locations (Fig. 1k,l), especially for region-specific cell types such as cortical projection neurons (clusters 6–8), cerebellum granule neurons (cluster 3) and hippocampal dentate gyrus neurons (cluster 9). These integration analyses confirmed the annotations and spatial locations of main cell types in our single-cell datasets.

In-depth view of cellular subtypes in the mammalian brain

Rather than performing subclustering analysis with the gene expression alone, we exploited the unique feature of EasySci-RNA (that is, full gene body coverage) by incorporating both gene counts and exonic counts for principal-component analysis followed by unsupervised clustering. The approach substantially increased the clustering resolution, as shown in a microglia subtype example (Fig. 2a,b). Leveraging this subclustering strategy, we identified a total of 359 subclusters, with a median of 1,038 cells in each group (Fig. 2c). All subclusters were contributed by multiple individuals, with a median of nine exonic markers enriched in each subcluster (Extended Data Fig. 5a,b and Supplementary Table 4). Some subtype-specific exonic markers were not detected by conventional differential gene analysis (for example, *Map2-ENSMUSE00000443205.3* in microglia-8; Extended Data Fig. 5c). Notably, our strategy favors detecting extremely rare cell types, such as rare pinealocytes (choroid plexus epithelial cells 7, 21 cells, marked by *Tph1* and *Ddc*²⁵) and tanyocytes (vascular leptomeningeal cells-2, 35 cells, marked by *Fndc3c1*, *Scn7a*²⁶) (Extended Data Fig. 5d–g).

About 75% of the 359 cell subclusters can be validated through integration analysis with other datasets (Fig. 2d). Our initial integration with a single-cell dataset featuring highly detailed cell type annotations² enables the validation of 112 subclusters, each matching with cell types documented in the previous study² (Fig. 2e and Supplementary Table 5). These corresponding cell types were further validated by their cell-type-specific markers, exemplified by neuronal intermediate progenitor cells, vascular smooth muscle cells, and olfactory ensheathing cells (Fig. 2f). Next, we integrated the 10x Visium spatial transcriptomics datasets²⁷ and determined the region-specificity of the recovered cell types or subtypes²⁷ (Extended Data Fig. 5h,i). We then expanded the analysis to include an extensive spatial transcriptomics dataset encompassing 75 coronal sections of the mouse brain^{27,28} and discovered 122 subclusters with high spatial mapping scores (Supplementary Table 5 and Methods). For instance, our analysis revealed that choroid plexus epithelial cells-6 were primarily situated in the

lateral ventricle, whereas cortical projection neurons 1-1 were predominantly found in the amygdala (Fig. 2g). As the third approach to confirm these subclusters, we utilized a deep-learning-based method²¹ to integrate the snRNA-seq and snATAC-seq data from each main cell type and recovered 224 ‘corresponding subclusters’ between the two molecular layers (Fig. 2h,i). As expected, the subclusters validated by ATAC-seq data exhibit more markers than those not validated (Fig. 2j). For example, the chromatin landscape for all 24 subclusters from cortical projection neurons 1 cells was recognized and validated by the significant enrichment of marker gene expression and activity in the target subcluster (Fig. 2k). We further explored *cis*-regulatory elements at the cell subtype resolution by correlation-based linkage analysis and unveiled a global network of putative enhancer-gene pairs shaping brain cell heterogeneity (Extended Data Fig. 6).

We next investigated key molecular programs underlying diverse cellular subtypes by clustering genes based on their expression variance across all 359 cell subclusters (Extended Data Fig. 7). We identified 21 gene modules (GMs), with the largest one (GM1) corresponding to a group of housekeeping genes. Several GMs were enriched in specific cell subtypes, such as the ependymal cell-specific GM²⁹ (GM11), and pituitary cells subtype-6 specific GM (GM9)³⁰. Similar analysis revealed programs in other rare subtypes, such as microglia-13 (GM19), vascular leptomeningeal cells-12 (GM20) and choroid plexus epithelial cells-7 (GM2). Remarkably, rare proliferating cells were identified through a cell-cycle-related GM (GM6), which include both conventional proliferating markers (for example, *Mki67*), and a group of less-studied lncRNAs (for example, *Gm29260* and *Gm37065*) (Extended Data Fig. 7c and Supplementary Table 6).

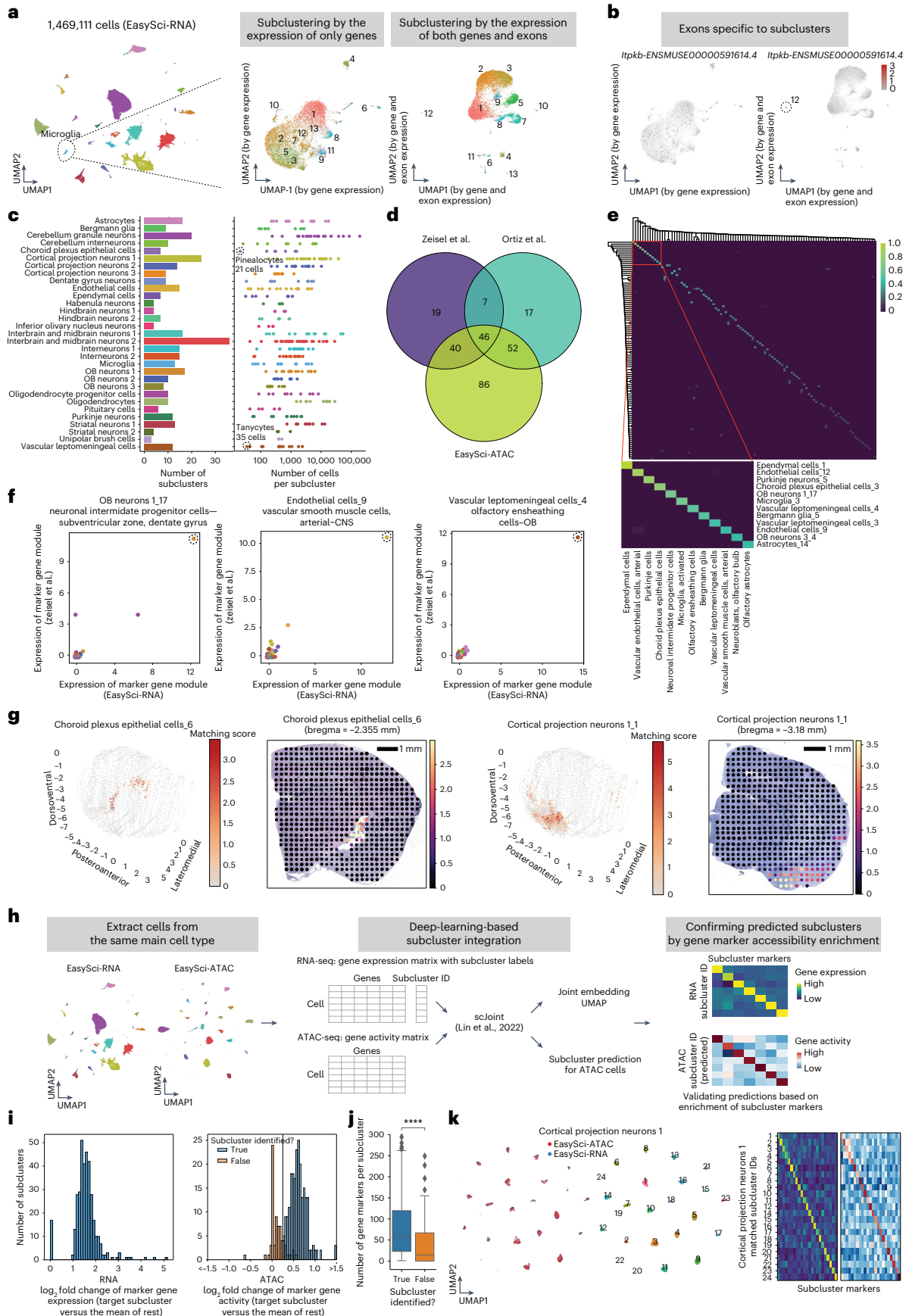
Aging-associated population dynamics at subtype resolution

To obtain a global view of brain cell population dynamics across the adult lifespan, we first quantified the cell-type-specific fractions recovered from each individual mouse. Differential abundance analyses were conducted across all 359 subclusters, yielding 45 and 29 significantly changed subclusters during early growth (between 3 and 6 months) and aging (between 6 and 21 months; Fig. 3a and Supplementary Tables 7 and 8), respectively. Significantly changed cell subtypes were strongly correlated between genders (Fig. 3b).

Consistent with the growth of the olfactory bulb (OB) during early development³¹, we observed significant expansion in all OB neuron subtypes during this phase. Meanwhile, a rare astrocyte subtype (AS-14, *Lyn*⁺ *Adgrb1*⁺) and a vascular leptomeningeal cell subtype (VLC-4, *Sox10*⁺ *Mybpc1*⁺) exhibited significant expansion in the same period (Fig. 3c). AS-14 featured with genes (for example, *BAl1*) involved in the clean-up of apoptotic neuronal debris produced during brain fast growth³², and VLC-4 highly expressed genes (for example, *Sox10* and *Mybpc1* (refs. 33,34)) involved in the growth of axons³⁵. Both subclusters were spatially mapped to the OB region, suggesting their

Fig. 2 | Identification of cellular subtypes in the mouse brain. **a**, Example UMAPs of microglia cells subjected to subclustering analysis based on gene expression alone (middle) or both gene and exon expression (right), colored by subcluster ID derived from combined gene and exon information. **b**, UMAP plots same as **a**, showing the expression of an exonic marker *Itpkb-ENSMUSE00000591614.4* of microglia subcluster 12. **c**, For each main cell type, bar plot showing the number of subclusters (left) and dot plot showing the number of cells from each subcluster (right). Two rare subclusters (choroid plexus epithelial cells-7 and vascular leptomeningeal cells-2) are circled. **d**, Venn diagram showing the number of validated subclusters using integration analysis with Zeisel et al.², Ortiz et al.²⁸ or the EasySci-ATAC dataset. **e**, Heatmap illustrating the similarity score between cell types derived from EasySci (rows) and Zeisel et al.² (columns), colored by the min-max normalized beta values obtained from cell-type correlation analysis (Methods). **f**, Dot plot showing the expression of example GMs unique to subclusters across paired cell types between the EasySci dataset and Zeisel et al.², colored by main cell types

(same as Fig. 2c). **g**, Spatial distributions of example subclusters, inferred using a brain spatial transcriptomics atlas²⁸ and the cell2location approach²⁷, colored by matching score. **h**, Scheme outlining scRNA and scATAC integration analyses. **i**, Histogram showing the log₂ fold change of marker gene expression from RNA-seq (left) and marker gene activity from ATAC-seq (right) per subcluster compared to the rest of cells. Predicted subclusters with log₂ fold change of gene activity > 0.25 between the target subcluster and the rest of cells and contain more than 10 cells are considered matched. **j**, Box plots illustrate the number of markers comparing the subclusters that could be validated in ATAC-seq data ($n = 224$ subclusters) to those that could not ($n = 135$ subclusters). Two-sided Wilcoxon rank-sum test was used. ****, $P < 0.0001$ ($P = 7.8 \times 10^{-8}$). **k**, Example subcluster integration results of cortical projection neurons 1. Left: UMAP plots demonstrating the overlap of two molecular layers, color-coded by subcluster ID. Right: heatmaps showing the enrichment of gene activity and gene accessibility of matching subclusters.



potential involvement in OB expansion (Fig. 3c). In contrast to the early growth stage, most OB neurons remained relatively stable during aging, with only a few subtypes showing significant changes. Key examples include the expansion of an OB neuron subtype corresponding to excitatory neurons in the mitral cell layer of the OB region³⁶ (OBN 3–3, marked by *Cpa6* and *Col23a1*), and the depletion of OB neuroblasts^{2,37} (OBN 1–11, marked by *Robo2* and *Prokr2*). Integration analysis with spatial transcriptomics datasets indicate these cell types were mapped to different regions of the OB (Fig. 3c).

More than twenty brain cell subtypes showed a marked reduction across the adult lifespan. For example, the most depleted populations in the aged brain include OB neuroblasts (OBN 1–11, marked by *Prokr2* and *Robo2*^{2,37}), OB neuronal progenitor cells (OBN 1–17, marked by *Mki67* and *Egfr*³⁸), and dentate gyrus neuroblasts (DGN-8, marked by *Sema3c* and *Igf1*³⁹) (Fig. 3d). DG neuroblasts declined even in the early growth, suggesting an earlier decline of DG neurogenesis compared to OB neurogenesis. In contrast to age-associated depletion of neurogenesis progenitors, oligodendrocyte progenitors (OPC-4, marked by *Pdgfra* and *Mki67*) remained relatively stable. However, newly formed oligodendrocytes (OLG-6, marked by *Prom1* and *Tcf7l1* (ref. 38,40)) and committed oligodendrocyte precursors (OPC-6, marked by *Bmp4* and *Enpp6* (ref. 38,40,41)) decreased during aging, indicating impaired oligodendrocyte differentiation. The age-associated population dynamics were further validated using the scATAC-seq dataset (Fig. 3d and Extended Data Fig. 8a,b) and our companion study in which we tracked cell dynamics via metabolic labeling⁴². Furthermore, we identified subtype-specific TF regulators using both gene expression and TF motif accessibility. This includes recognized regulators of neurogenesis (for instance, *Sox2* and *E2f2*^{43,44}), demonstrating the potential of our datasets to unveil key epigenetic signatures of aging-associated cell subtypes (Fig. 3e).

A total of 14 cell subtypes notably expanded in the aged brain, such as a microglia subtype (MG-9, *ApoE*⁺, *Csf1*⁺) corresponding to a previously reported DAM⁴⁵, and a reactive oligodendrocyte subtype (OLG-7, *C4b*⁺, *Serpina3n*^{46,47}). With the scATAC-seq dataset, we further confirmed its expansion (Fig. 3f and Extended Data Fig. 8b,c) and identified its associated TFs. For example, the OLG-7 associated TF, *Stat3* (Fig. 3e), plays a critical role in regulating inflammation and immunity in the brain⁴⁸. We also performed a spatial transcriptomics experiment using adult and aged mouse brains. Strikingly, we detected a significant enrichment of the reactive oligodendrocyte-specific markers (for example, *C4b* and *Serpina3n*) around the subventricular zone (SVZ) (Fig. 3g,h), indicating an age-related activation of inflammation signaling around the adult neurogenesis niche.

We next explored the subtype manifestation of aging signatures by differentially expressed (DE) gene analysis. We identified 7,135 aging-associated signatures across 359 subclusters (Supplementary Table 9 and Extended Data Fig. 9a). Of the 580 genes significantly

altered in multiple (≥ 3) subtypes, 241 showed consistent directions. For example, *Nr4a3* (genes involved in DNA repair⁴⁹) was significantly decreased in aged neuron subtypes (striatal neurons, OB neurons, and interneurons). *Hdac4*, encoding a histone deacetylase⁵⁰, decreased in aged astrocytes and ependymal cells. Insulin-degrading enzyme (IDE), involved in amyloid-beta clearance⁵¹, also increased in neuron subtypes. We also identified age-related changes in non-coding RNAs, many with high cell-type specificity (for example, *B230209E15Rik* in cortical projection neuron subtypes), but were not well characterized previously (Extended Data Fig. 9b).

AD pathogenesis-associated gene signatures and cell subtypes

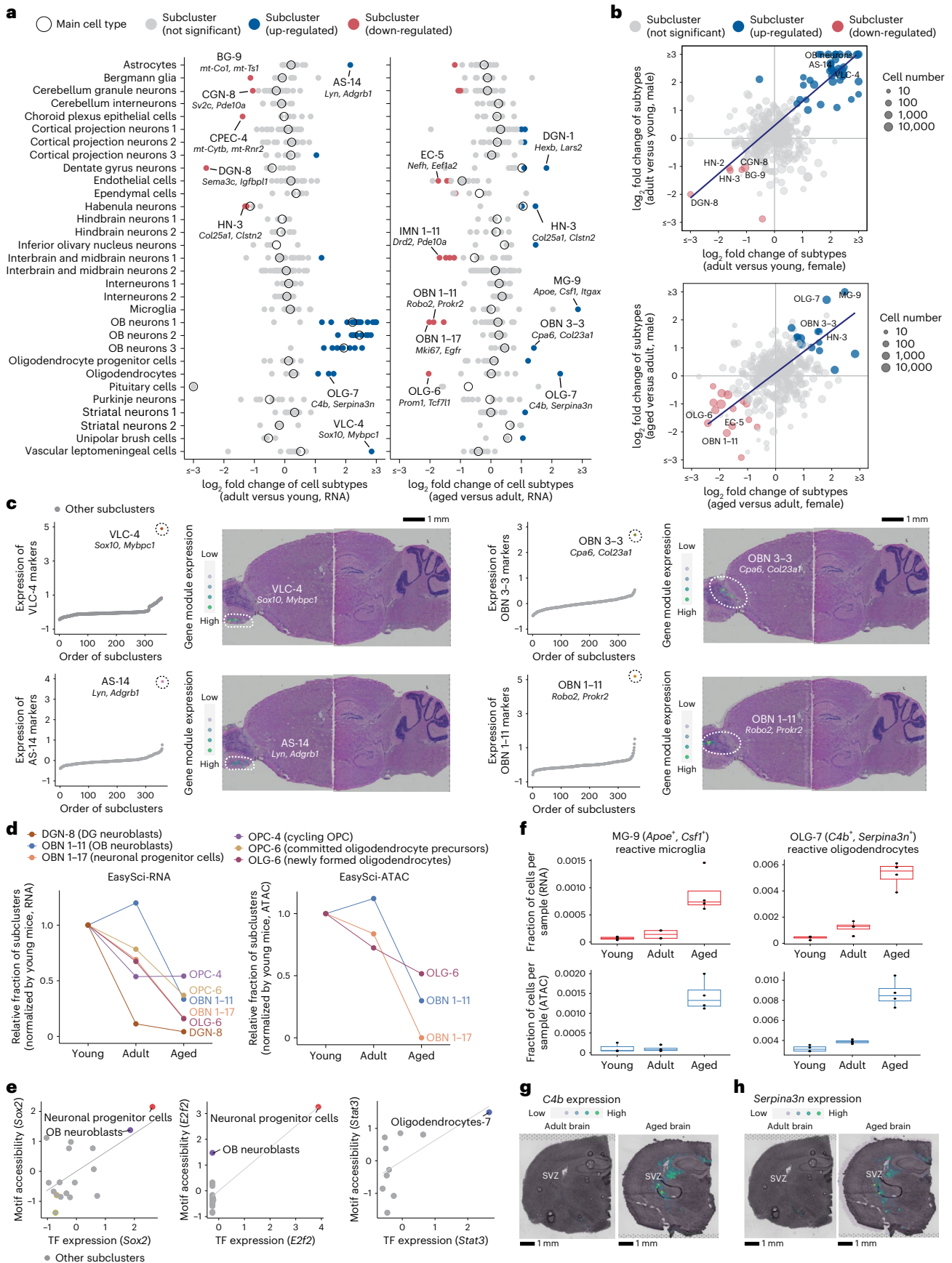
Through comparison of subcluster fractions in two AD models to age-matched WT controls (3 months old), we detected 16 and 14 significantly changed subclusters (FDR of 5%, at least twofold change) in the EOAD (5xFAD) model and LOAD (APOE*4/Trem2*R47H) model, respectively (Fig. 4a and Supplementary Tables 10 and 11). Most significantly altered subtypes correlated between genders (Fig. 4b) and between the two AD models, even though they had distinct genetic perturbations in different cell types (Fig. 4c). For example, a rare choroid plexus epithelial cell subtype (CPEC-4) was strongly depleted (by more than twofold decrease) in both models. This cell type is marked by significant enrichment of multiple mitochondrial genes linked to neuroprotective factors against neurodegeneration (for example, *mt-Rnr2* (ref. 52)) or Tau protein levels in cerebrospinal fluid (*mt-Rnr1* and *mt-Nd5*⁵³). Through spatial transcriptomics analysis, we verified its location around the SVZ and confirmed its depletion in the EOAD (5xFAD) model, suggesting mitochondrial dysfunction in choroid plexus epithelial cells plays a role in neurodegenerative diseases (Fig. 4d,e).

By contrast, another choroid plexus epithelial cell subtype (CPEC-6; marked by *Sptlc3*¹⁺⁵⁴, *Fer1l6*⁺) expanded in both AD models (over twofold increase) (Fig. 4b). A similar expansion was observed in a rare interbrain and midbrain neuron subtype (IMN 1–13, marked by *Col25a1*⁺, *Ndrgr1*⁺) that expresses *Col25a1*, a membrane-associated collagen reported to promote intracellular amyloid formation in mouse models⁵⁵ (Fig. 4c). Spatial transcriptomic analysis confirmed the up-regulation of IMN 1–13 specific gene markers in the thalamus region of the 5xFAD mouse brain (Fig. 4d,e), providing further validation of the AD-related neuron subtype change. Additionally, a septal nuclei neuron subtype IMN 2-9 (marked by *Prdm16* and *Ano2*) that significantly overexpress in both AD model GMs related to axonogenesis (for example, *Nrp1* and *Slit2*) and synaptogenesis (for example, *Ptprd* and *Nrxn1*) (ref. 56) was also significantly expanded in both AD models (Fig. 4f–h), aligning with the observed enlargement of the septal nuclei region several years before the onset of memory decline⁵⁷.

Meanwhile, we observed a significant expansion of microglia subtype 9 (marked by *ApoE* and *Csf1*) in early-onset 5xFAD mice, aligning with previous reports⁴⁵. This disease-associated microglia (DAM) also

Fig. 3 | Identifying brain cell population changes across the lifespan at subtype resolution. a, Dot plots showing the log-transformed cell-type-specific fraction changes of main cell types (circles) and subclusters (dots) between adult versus young (left) and between aged versus adult (right) from EasySci-RNA data. Significantly changed subclusters were colored by the direction of changes. Representative subclusters were labeled along with top gene markers. AS, astrocytes; BG, Bergmann glia; CGN, cerebellum granule neurons; CPEC, choroid plexus epithelial cells; DGN, dentate gyrus neurons; EC, endothelial cells; HN, habenula neurons; IMN 1, interbrain and midbrain neurons 1; MG, microglia; OBN 1, OB neurons 1; OBN 3, OB neurons 3; OLG, oligodendrocytes; OPC, oligodendrocyte progenitor cell; VLC, vascular leptomeningeal cells. **b**, Scatter plots showing the correlation of the subcluster specific fraction changes between males and females in the early growth stage (top) and the aging stage (bottom), with a linear regression line. The most significantly changed subclusters are annotated on the plots. **c**, Examples of development- or aging-associated subclusters and their spatial positions. Left: scatter plots showing the aggregated

expression of subcluster-specific marker genes across all subclusters. Right: plots showing the aggregated expression of subcluster-specific marker genes across a brain sagittal section in 10x Visium spatial transcriptomics data. **d**, Line plots showing the relative fractions of depleted subclusters across three age groups identified from EasySci-RNA (left) and EasySci-ATAC (right). **e**, Scatter plots showing the correlated gene expression and motif accessibility of TFs across subclusters enriched in OB neurons 1–17 (*Sox2* and *E2f2*, left and middle) and oligodendrocytes-7 (*Stat3*, right), together with a linear regression line. **f**, Box plots showing the fractions of the reactive microglia (left) and reactive oligodendrocytes (right) across three age groups (young: $n = 4$ mice, adult: $n = 4$ mice, aged: $n = 4$ mice) profiled by EasySci-RNA (top) and EasySci-ATAC (bottom). **g, h**, Mouse brain coronal sections showing the expression level of *C4b* (g) and *Serpina3n* (h) in the adult (left) and aged (right) brains from spatial transcriptomics analysis. Boxes in box plots indicate the median and IQR with whiskers indicating $1.5 \times$ IQR.



expanded in aged mice but was not evident in the late-onset APOE*4/Trem2*R47H model at 3 months of age (validated by both RNA and ATAC), potentially indicating a correlation with disease onset (Fig. 4i). We further investigated its DE genes (Extended Data Fig. 8d) and key TFs exhibiting consistency between cell-type-specific gene expression and motif accessibility (Fig. 4j). The enriched TFs were reported to be involved in microglia expansion during aging and AD^{58–60}. Additionally, we quantified the enrichment of genetic variants linked to human traits⁶¹ and observed significant enrichment of AD heritability in microglia cells at both the main cell type level and particularly in the microglia-9 subtype, highlighting the role of DAM in AD pathogenesis (Extended Data Fig. 8e,f).

We identified subtype-specific manifestations of key AD-related molecular signatures. In the 5xFAD (EOAD) model, we found 6,792 subcluster-specific DE genes, whereas the APOE*4/Trem2*R47H (LOAD) model had 7,192 subcluster-specific DE genes (Extended Data Fig. 9c,f and Supplementary Tables 12 and 13). The *ApoE* gene was globally down-regulated in the APOE*4/Trem2*R47H mice, possibly due to the replacement of the *ApoE* gene with the human sequence. Many AD-associated gene signatures exhibited consistent changes across cellular subtypes, such as increased stress-related markers (for example, *Hsp90aa1* and *Txnrd1*) in neuron subtypes in the 5xFAD mice. The expression of *Reln*⁶² decreased in various cell types in both models, aligning with previous report of *Reln* depletion before the onset of amyloid-beta pathology in the human frontal cortex⁶³. Other intriguing observations included the down-regulation of *Tlcd4*, a gene involved in lipid trafficking and metabolism⁶⁴ in multiple subclusters in the 5xFAD mice. Interestingly, despite genetic differences and disease onsets in the two AD models, there were remarkably consistent alterations in cell-type-specific molecular profiles. We identified 559 subcluster-specific DE genes shared between both AD mutants, suggesting common molecular mechanisms between early- and late-onset AD models (Extended Data Fig. 9g). We also investigated the connection between aging and AD-associated changes using transcriptomic aging clocks⁶⁵, revealing significantly accelerated biological aging in both AD models (Extended Data Fig. 9h). Although most cell types demonstrated accelerated aging-related molecular changes, specific cell types only exhibited these signs in LOAD (Extended Data Fig. 9i). This is further validated by consistent cell-type-specific changes of aging-associated gene signatures (for example, *Neat1* and *Zfp423*) in aged and AD models (Extended Data Fig. 9k).

Detection of dysregulated gene signatures in human AD brains

To compare molecular signatures associated with AD pathogenesis in mouse models and human patients, we sequenced a total of 118,240 single-nuclei transcriptomes (a median of 5,585 nuclei per sample, with the sequencing depth of 13,850 raw reads and a median of 1,109 UMIs per nucleus; Extended Data Fig. 10a,b) from 24 human brain samples

across two brain regions (hippocampus, superior and middle temporal gyrus (SMTG)), derived from six patients with AD and six age- and gender-matched controls (Supplementary Table 14). Thirteen main cell types were identified through integration analysis with the mouse dataset and validated by the specific expression of known markers (Fig. 5a and Extended Data Fig. 10c–e).

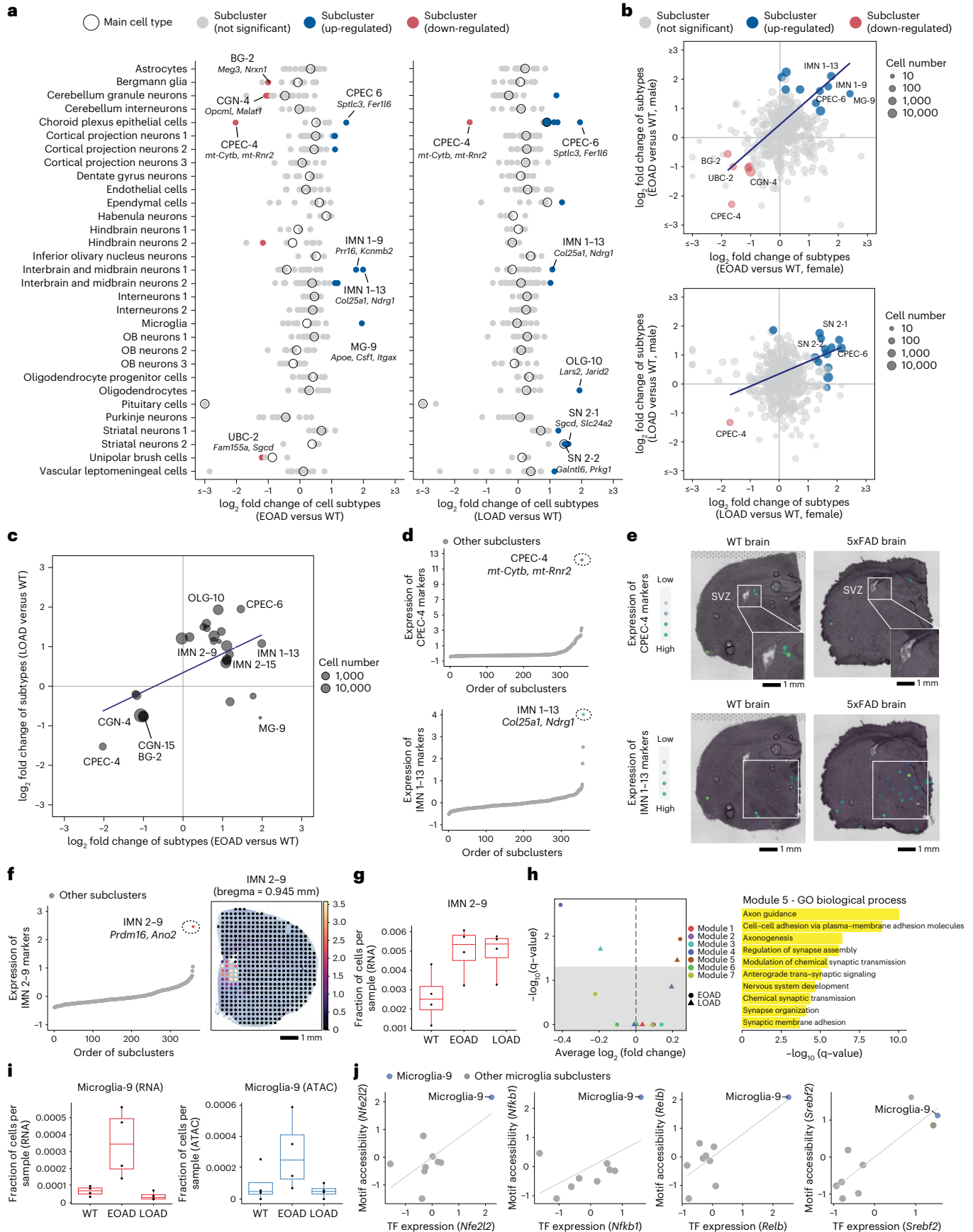
A total of 4,171 and 2,149 cell-type-specific DE genes were identified in the hippocampus and SMTG, respectively (Fig. 5b and Supplementary Table 15). Exactly 349 genes were significantly changed in the same cell type from two distinct regions, among which 332 were altered consistently (Fig. 5c). For example, oligodendrocytes in AD samples from both regions exhibited decreased expression of the oligodendrocyte terminal differentiation factor *OPALIN*⁶⁶ and the oxidation stress protector *OXR1* (ref. 67). Concurrently, we observed an up-regulation of genes related to programmed cell death (for example, *FLCN* and *RASSF2*)^{68,69}, suggesting an elevated stress in oligodendrocytes from AD brains. Other examples include the microglia-specific up-regulation of *PTPRG*⁷⁰, and astrocyte-specific down-regulation of several transmembrane transporters (for example, *AQP4*) and neurotransmitter metabolism enzymes (for example, *GLUD1*)^{71,72}.

Interestingly, some AD-associated gene signatures exhibited region-specific expression patterns. For example, *GNMNB*, encoding a transmembrane glycoprotein associated with microglia activation in AD brains⁷³, showed increased expression in the microglia from the hippocampus but not from the SMTG. On the other hand, *MMP24*, encoding a member of the metalloproteinase family implicated in AD pathogenesis⁷⁴, showed increased expression in cortical projection neurons unique within the SMTG (Fig. 5d). Notably, inhibition of *MMP24* has been demonstrated to decrease amyloid-beta levels and promote cognitive functions in mouse models⁷⁵, suggesting its potential role as a novel therapeutic target for AD.

Finally, we explored the human-mice relevance for AD-associated gene signatures and molecular pathways. Despite differences in the species and disease stages between the two datasets, several genes encoding heat shock proteins (for example, *HSP90AA1* and *HSPH1*) were up-regulated across multiple cell types in both species (Fig. 5e). The elevated chaperon system potentially reduces the formation of toxic oligomeric assemblies in AD brains⁷⁶, further validating the dysfunction of proteostasis as a molecular marker of AD⁷⁷. Meanwhile, we identified down-regulated genes in both human and mice. One of the examples, *PLPI*, was reported as a subtype-specific driver gene contributing to AD pathogenesis⁷⁸. Another gene, *PDE10A*, plays a key role in promoting neuronal survival, with its reduction detected in our datasets and multiple neurodegenerative diseases (for example, Huntington's disease⁷⁹ and Parkinson's disease⁸⁰) (Fig. 5f). Importantly, the above-mentioned trends were readily validated by another single-cell dataset investigating AD in the human prefrontal cortex⁶ (Extended Data Fig. 10f). In summary, the human-mice relevance analysis identified species-conserved genetic programs associated with AD pathogenesis.

Fig. 4 | Identifying AD pathogenesis-associated cell subtypes. **a**, Dot plots showing the log-transformed fold changes of main cell types (circles) and subclusters (dots) comparing EOAD versus WT (left) and LOAD versus WT (right). Significantly changed subclusters were colored by the direction of changes. Representative subclusters were labeled along with top gene markers. **b**, Scatter plots showing the correlation of the log fold change of subclusters (top: EOAD versus WT; bottom: LOAD versus WT) between males and females. **c**, Scatter plot showing the correlation of the log fold change of subclusters in two AD models (both compared with the WT). Only subclusters with significant changes in at least one AD model are included. Black/gray reflects overlapping dots. **d**, Scatter plots showing the aggregated expression of gene markers of two cell subtypes (top: CPEC-4; bottom: the IMN1–13) across all subclusters from *EasySci-RNA* data. **e**, Brain coronal sections showing the spatial expression of subtype-specific gene markers of two subtypes (top: CPEC-4; bottom: the IMN1–13) in the WT

and EOAD brains in 10x Visium data. **f**, Scatter plot showing the aggregated expression of gene markers of IMN2–9 across all subclusters from *EasySci-RNA* data (left) and the mapping score per pixels of the IMN2–9 subcluster cells on the bregma 0.945 mm coronal section highlighting the lateral septal nucleus region²⁸ (right). **g**, Box plots showing the fraction of IMN2–9 cells across conditions (WT: $n = 4$ mice, EOAD: $n = 4$ mice, LOAD: $n = 4$ mice) profiled by *EasySci-RNA*. **h**, GM expression differences in both AD models against WT control (left) and the top enriched Gene Ontology (GO) biological processes pathways of module 5 genes (right). **i**, Box plots showing the fraction of microglia-9 cells across conditions (WT: $n = 4$ mice, EOAD: $n = 4$ mice, LOAD: $n = 4$ mice) profiled by *EasySci-RNA* (left) or *EasySci-ATAC* (right). **j**, Scatter plot showing the correlated gene expression and motif accessibility of four TFs enriched in microglia-9, together with a linear regression line. Boxes in box plots indicate the median and IQR with whiskers indicating 1.5× IQR.



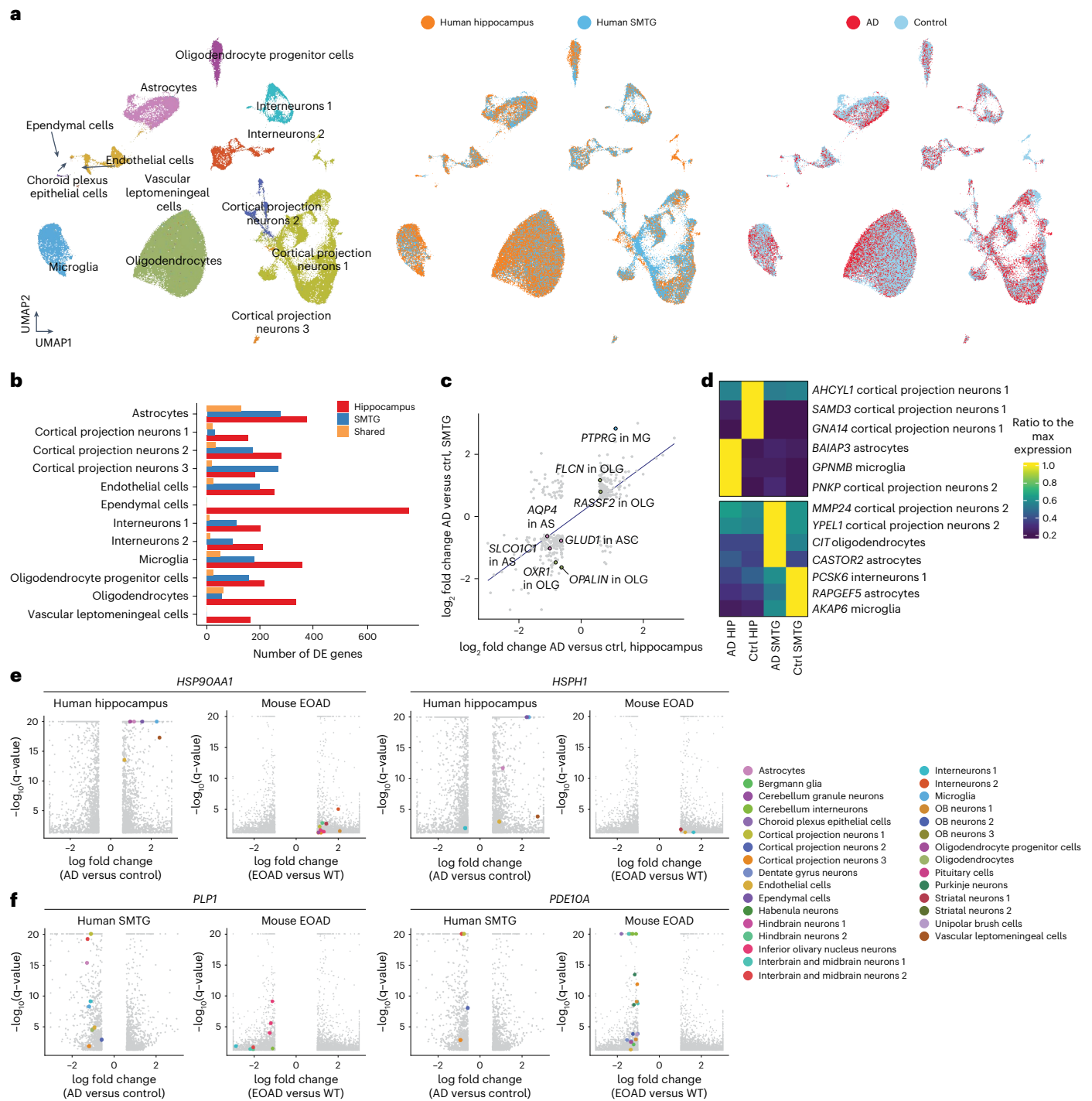


Fig. 5 | Identifying AD pathogenesis-associated gene expression signatures across regions and cell types in human brains. **a**, UMAP visualization of single-cell transcriptomes of all human brain cells, colored by main cell types (left), region (middle) and conditions (right). **b**, Bar plot showing the number of DE genes between AD and control samples in each cell type, colored by whether they are unique to each region or shared between two regions. Choroid plexus epithelial cells and vascular leptomeningeal cells were not included into the differential gene expression analysis in SMTG due to their low cell numbers. **c**, We detected 394 DE genes significantly changed within the same main cell type in both regions. The scatter plot shows the correlation of the \log_2 -transformed

fold changes of these 394 shared DE genes in the hippocampus (x axis) and in SMTG (y axis). Key genes are annotated and colored by their corresponding main cell types. **d**, Heatmaps showing examples of region-specific DE genes for the hippocampus (left) and SMTG (right). Gene expressions were quantified as TPM in the corresponding cell types in each group, and normalized to the maximum expression across groups. **e, f**, Volcano plots showing the examples of top DE genes between the AD and control samples across main cell types in human brains (**e**) or between EOAD and WT samples across cell subclusters in mouse brains (**f**). Highlighted genes are colored by the main cell-type identity.

Discussion

In this study, we introduced EasySci, a cost-effective technical framework for individual laboratories to generate gene expression and

chromatin accessibility profiles from millions of single cells. We used EasySci to analyze 1.5 million single-cell transcriptomes with full gene body coverage and 380,000 chromatin accessibility profiles across

mammalian brains of different ages and genotypes. The datasets enable the identification of over 300 cellular subtypes throughout the brain, including highly rare cell types representing less than 0.01% of the total brain cell population. Furthermore, we discovered region-specific effects attributable to aging and AD and examined the manifestation of molecular signatures associated with aging and AD on a cell-type-specific basis.

As highlighted by our subcluster level analysis, the effects of aging and AD on the global brain cell population are profoundly cell-type specific. Although most brain cell types stay relatively stable under various conditions, we identified over 50 cell subtypes exhibiting over twofold change in brains affected by aging and AD models. Many of these cell subtypes were rare and overlooked in conventional single-cell analysis. For example, the aging brain is characterized by the depletion of both rare neuronal progenitor cells and differentiating oligodendrocytes, associated with the enrichment of a *C4b⁺Serpina3n⁺* reactive oligodendrocyte subtype surrounding the SVZ, suggesting a potential interplay between oligodendrocytes, localized inflammatory signals and the stem cell niche.

The lack of reliable mouse models remains a big challenge in studying late-onset AD. The novel APOE*4/Trem2*R47H model aims to overcome this limitation by introducing two of the strongest late-onset AD-associated mutations⁸¹. We found consistent molecular and cellular population dynamics between the well-established 5xFAD and the novel APOE*4/Trem2*R47H model. For example, we observed shared subtypes that were depleted (for example, *mt-Cytb⁺mt-Rnr2⁺* choroid plexus epithelial cell) or enriched (for example, *Col25a1⁺Ndrgr1⁺* inter-brain and midbrain neuron) in both early- and late-onset AD mutant brains. Meanwhile, differences were also observed between the two AD models, as expected by the different onset times. The absence of an increase in the DAM population in the LOAD model may be due to its lack of amyloid deposition⁸² or by genetic perturbations, as both *Trem2* and *ApoE* play a role in the activation of this cell population⁴⁵.

In addition, we investigated AD-associated gene signatures in human brains by profiling over 100,000 single-nucleus transcriptomes derived from 24 human brain samples from control and AD patients, across two distinct anatomical locations. Although most AD-associated gene dynamics are profoundly cell-type and region specific, we identified dysregulated genetic signatures that are conserved between different locations in the human brains. Moreover, integrating the human and mouse brain datasets further revealed molecular pathways shared between human AD patients and mouse AD models, which suggests that the mouse AD model can serve as a model system to investigate the function and regulation of these conserved features associated with AD or neuronal dysfunction.

Of note, there are several inherent limitations of the study. First, the analysis covers only around 2% of the total mouse brain population (estimated at approximately 100 million cells), which means extremely rare cell subtypes may still be overlooked. Additionally, our relatively shallow sequencing depth might hinder the detection of lowly expressed transcripts or minor aging-related cellular state changes. Nevertheless, the validity of our key biological findings is reinforced by the consistent results across different genders (male versus female), genotypes (EOAD versus LOAD), and orthogonal approaches (such as comparisons between single-cell transcriptome, chromatin accessibility or spatial transcriptomics). This lends significant credence to our discoveries, even when considering the limitations of the study.

In summary, we have showcased the power of highly scalable single-cell genomics to delve into the dynamics of rare cell types, uncovering novel subtypes associated with aging and disease. Though our focus was on brain tissues, the strategic approach could be readily extended to systematically explore cellular states across an entire organism. Such exploration could illuminate the rare vulnerable cell populations to aging and diseases, opening up pathways to develop targeted therapeutic strategies.

Online content

Any methods, additional references, Nature Portfolio 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/s41588-023-01572-y>.

References

1. Erö, C., Gewaltig, M.-O., Keller, D. & Markram, H. A cell atlas for the mouse brain. *Front. Neuroinform.* **12**, 84 (2018).
2. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).
3. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
4. Xia, X., Jiang, Q., McDermott, J. & Han, J.-D. J. Aging and Alzheimer's disease: comparison and associations from molecular to system level. *Aging Cell* **17**, e12802 (2018).
5. Ximerakis, M. et al. Single-cell transcriptomic profiling of the aging mouse brain. *Nat. Neurosci.* **22**, 1696–1708 (2019).
6. Morabito, S. et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021).
7. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
8. Wang, R. et al. Construction of a cross-species cell landscape at single-cell level. *Nucleic Acids Res.* **51**, 501–516 (2022).
9. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
10. Cao, J. et al. A human cell atlas of fetal gene expression. *Science* **370**, eaba7721 (2020).
11. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
12. Ma, S. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116 (2020).
13. Martin, B. K. et al. An optimized protocol for single cell transcriptional profiling by combinatorial indexing. *Nat. Protoc.* **18**, 188–207 (2023).
14. Domcke, S. et al. A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7721 (2020).
15. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* **174**, 1309–1324 (2018).
16. Oakley, H. et al. Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* **26**, 10129–10140 (2006).
17. Desimone, A. et al. The influence of ApoE4 on the clinical outcomes and pathophysiology of degenerative cervical myelopathy. *JCI Insight* **6**, e149227 (2021).
18. Xiang, X. et al. The Trem2 R47H Alzheimer's risk variant impairs splicing and reduces Trem2 mRNA and protein in mice but not in humans. *Mol. Neurodegener.* **13**, 49 (2018).
19. McInnes, L. et al. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).
20. Blondel, V. D. et al. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, 10008 (2008).
21. Lin, Y. et al. scJoint integrates atlas-scale single-cell RNA-seq and ATAC-seq data with transfer learning. *Nat. Biotechnol.* **40**, 703–710 (2022).
22. Yeh, H. & Ikezu, T. Transcriptional and epigenetic regulation of microglia in health and disease. *Trends Mol. Med.* **25**, 96–111 (2019).
23. Watakabe, A. et al. Comparative analysis of layer-specific genes in mammalian neocortex. *Cereb. Cortex* **17**, 1918–1933 (2007).

24. McEvelly, R. J. et al. Requirement for Brn-3.0 in differentiation and survival of sensory and motor neurons. *Nature* **384**, 574–577 (1996).
25. Mays, J. C. et al. Single-cell RNA sequencing of the mammalian pineal gland identifies two pinealocyte subtypes and cell type-specific daily patterns of gene expression. *PLoS ONE* **13**, e0205883 (2018).
26. Campbell, J. N. et al. A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.* **20**, 484–496 (2017).
27. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
28. Ortiz, C. et al. Molecular atlas of the adult mouse brain. *Sci. Adv.* **6**, eabb3446 (2020).
29. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
30. Liu, J. et al. Tbx19, a tissue-selective regulator of POMC gene expression. *Proc. Natl Acad. Sci. USA* **98**, 8674–8679 (2001).
31. Tufo, C. et al. Development of the mammalian main olfactory bulb. *Development* **149**, dev200210 (2022).
32. Sokolowski, J. D. et al. Brain-specific angiogenesis inhibitor-1 expression in astrocytes and neurons: implications for its dual function as an apoptotic engulfment receptor. *Brain Behav. Immun.* **25**, 915–921 (2011).
33. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
34. Tepe, B. et al. Single-cell RNA-seq of mouse olfactory bulb reveals cellular heterogeneity and activity-dependent molecular census of adult-born neurons. *Cell Rep.* **25**, 2689–2703 (2018).
35. Barraud, P. et al. Neural crest origin of olfactory ensheathing glia. *Proc. Natl Acad. Sci. USA* **107**, 21040–21045 (2010).
36. Monavarfeshani, A., Knill, C. N., Sabbagh, U., Su, J. & Fox, M. A. Region- and cell-specific expression of transmembrane collagens in mouse brain. *Front. Integr. Neurosci.* **11**, 20 (2017).
37. Puverel, S., Nakatani, H., Parras, C. & Soussi-Yanicostas, N. Prokineticin receptor 2 expression identifies migrating neuroblasts and their subventricular zone transient-amplifying progenitors in adult mice. *J. Comp. Neurol.* **512**, 232–242 (2009).
38. Pastrana, E., Cheng, L.-C. & Doetsch, F. Simultaneous prospective purification of adult subventricular zone neural stem cells and their progeny. *Proc. Natl Acad. Sci. USA* **106**, 6387–6392 (2009).
39. Kumar, A. et al. Transcriptomic analysis of the signature of neurogenesis in human hippocampus suggests restricted progenitor cell progression post-childhood. *IBRO Rep.* **9**, 224–232 (2020).
40. Marques, S. et al. Transcriptional convergence of oligodendrocyte lineage progenitors during development. *Dev. Cell* **46**, 504–517 (2018).
41. Zhang, Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
42. Lu, Z. et al. Tracking cell-type-specific temporal dynamics in human and mouse brains. *Cell* **186**, 4345–4364.e24 (2023).
43. Graham, V., Khudyakov, J., Ellis, P. & Pevny, L. SOX2 functions to maintain neural progenitor identity. *Neuron* **39**, 749–765 (2003).
44. Li, J. et al. Transcription factors Sp8 and Sp9 coordinately regulate olfactory bulb interneuron development. *Cereb. Cortex* **28**, 3278–3294 (2018).
45. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer’s disease. *Cell* **169**, 1276–1290 (2017).
46. 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).
47. Kenigsbuch, M. et al. A shared disease-associated oligodendrocyte signature among multiple CNS pathologies. *Nat. Neurosci.* **25**, 876–886 (2022).
48. See, A. P. et al. The role of STAT3 activation in modulating the immune microenvironment of GBM. *J. Neurooncol.* **110**, 359–368 (2012).
49. Paillasse, M. R. & de Medina, P. The NR4A nuclear receptors as potential targets for anti-aging interventions. *Med. Hypotheses* **84**, 135–140 (2015).
50. Di Giorgio, E. et al. HDAC4 degradation during senescence unleashes an epigenetic program driven by AP-1/p300 at selected enhancers and super-enhancers. *Genome Biol.* **22**, 129 (2021).
51. Zhang, Y. & Wang, P. Age-related increase of insulin-degrading enzyme is inversely correlated with cognitive function in APPsw/PS1dE9 mice. *Med. Sci. Monit.* **24**, 2446–2455 (2018).
52. Hashimoto, Y. et al. A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer’s disease genes and Aβeta. *Proc. Natl Acad. Sci. USA* **98**, 6336–6341 (2001).
53. Cavalcante, G. C. et al. Mitochondrial genetics reinforces multiple layers of interaction in Alzheimer’s disease. *Biomedicines* **10**, 880 (2022).
54. Mielke, M. M. & Lyketsos, C. G. Alterations of the sphingolipid pathway in Alzheimer’s disease: new biomarkers and treatment targets? *Neuromolecular Med.* **12**, 331–340 (2010).
55. Tong, Y., Xu, Y., Searce-Levie, K., Ptáček, L. J. & Fu, Y.-H. COL25A1 triggers and promotes Alzheimer’s disease-like pathology in vivo. *Neurogenetics* **11**, 41–52 (2010).
56. Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E. & Swarup, V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep. Methods* **3**, 100498 (2023).
57. Butler, T. et al. Volume of the human septal forebrain region is a predictor of source memory accuracy. *J. Int. Neuropsychol. Soc.* **18**, 157–161 (2012).
58. Oeckinghaus, A. & Ghosh, S. The NF-kappaB family of transcription factors and its regulation. *Cold Spring Harb. Perspect. Biol.* **1**, a000034 (2009).
59. Liu, X.-F. et al. Nrf2 as a target for prevention of age-related and diabetic cataracts by against oxidative stress. *Aging Cell* **16**, 934–942 (2017).
60. Bommer, G. T. & MacDougald, O. A. Regulation of lipid homeostasis by the bifunctional SREBF2-miR33a locus. *Cell Metab.* **13**, 241–247 (2011).
61. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
62. Seripa, D. et al. The RELN locus in Alzheimer’s disease. *J. Alzheimers Dis.* **14**, 335–344 (2008).
63. Herring, A. et al. Reelin depletion is an early phenomenon of Alzheimer’s pathology. *J. Alzheimers Dis.* **30**, 963–979 (2012).
64. Attwood, M. M. & Schiöth, H. B. Characterization of five transmembrane proteins: with focus on the Tweety, Sideroflexin, and YIP1 domain families. *Front Cell Dev. Biol.* **9**, 708754 (2021).
65. Buckley, M. T. et al. Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. *Nat. Aging* **3**, 121–137 (2023).
66. de Faria, O. Jr et al. TMEM10 promotes oligodendrocyte differentiation and is expressed by oligodendrocytes in human remyelinating multiple sclerosis plaques. *Sci. Rep.* **9**, 3606 (2019).
67. Volkert, M. R. & Crowley, D. J. Preventing neurodegeneration by controlling oxidative stress: the role of OXR1. *Front. Neurosci.* **14**, 611904 (2020).

68. Schmidt, L. S. & Linehan, W. M. FLCN: the causative gene for Birt-Hogg-Dubé syndrome. *Gene* **640**, 28–42 (2018).
69. Cooper, W. N. et al. RASSF2 associates with and stabilizes the proapoptotic kinase MST2. *Oncogene* **28**, 2988–2998 (2009).
70. Luo, J. et al. PTPRG activates m6A methyltransferase VIRMA to block mitochondrial autophagy mediated neuronal death in Alzheimer's disease. Preprint at medRxiv <https://doi.org/10.1101/2022.03.11.22272061> (2022).
71. Silva, I., Silva, J., Ferreira, R. & Trigo, D. Glymphatic system, AQP4, and their implications in Alzheimer's disease. *Neurol. Res Pr.* **3**, 5 (2021).
72. Kulijewicz-Nawrot, M., Syková, E., Chvátal, A., Verkhatsky, A. & Rodríguez, J. J. Astrocytes and glutamate homeostasis in Alzheimer's disease: a decrease in glutamine synthetase, but not in glutamate transporter-1, in the prefrontal cortex. *ASN Neuro.* **5**, 273–282 (2013).
73. Hüttenrauch, M. et al. Glycoprotein NMB: a novel Alzheimer's disease associated marker expressed in a subset of activated microglia. *Acta Neuropathol. Commun.* **6**, 108 (2018).
74. Zipfel, P., Rochais, C., Baranger, K., Rivera, S. & Dallemagne, P. Matrix metalloproteinases as new targets in Alzheimer's disease: opportunities and challenges. *J. Med. Chem.* **63**, 10705–10725 (2020).
75. Baranger, K. et al. MT5-MMP is a new pro-amyloidogenic proteinase that promotes amyloid pathology and cognitive decline in a transgenic mouse model of Alzheimer's disease. *Cell. Mol. Life Sci.* **73**, 217–236 (2016).
76. Arawaka, S., Machiya, Y. & Kato, T. Heat shock proteins as suppressors of accumulation of toxic prefibrillar intermediates and misfolded proteins in neurodegenerative diseases. *Curr. Pharm. Biotechnol.* **11**, 158–166 (2010).
77. Cornejo, V. H. & Hetz, C. The unfolded protein response in Alzheimer's disease. *Semin. Immunopathol.* **35**, 277–292 (2013).
78. Neff, R. A. et al. Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets. *Sci. Adv.* **7**, eabb5398 (2021).
79. Niccolini, F. et al. Altered PDE10A expression detectable early before symptomatic onset in Huntington's disease. *Brain* **138**, 3016–3029 (2015).
80. Niccolini, F. et al. Loss of phosphodiesterase 10A expression is associated with progression and severity in Parkinson's disease. *Brain* **138**, 3003–3015 (2015).
81. Karch, C. M. & Goate, A. M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol. Psychiatry* **77**, 43–51 (2015).
82. Kotredes, K. P. et al. Uncovering disease mechanisms in a novel mouse model expressing humanized APOEε4 and Trem2^{*R47H}. *Front. Aging Neurosci.* **13**, 735524 (2021).
83. Ding, J. et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).
84. Stoeckius, M. et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Methods

Animals

C57BL/6 WT mouse brains at 3 months ($n = 4$), 6 months ($n = 4$) and 21 months ($n = 4$) were collected in this study. Two AD models at 3 months old from the same C57BL/6 background were added, including an early-onset model (5x*FAD*, JAX stock #034840) that overexpresses mutant human amyloid-beta precursor protein with the Swedish (K670N, M671L), Florida (I716V) and London (V717I) familial AD (FAD) mutations and human presenilin 1 harboring two FAD mutations, M146L and L286V. Brain-specific overexpression is achieved by neural-specific elements of the mouse *Thy1* promoter¹⁶. The second, late-onset AD model (APOE*4/*Trem2**R47H, JAX stock #028709) in this study carries two of the highest risk factor mutations of LOAD⁸¹, including a humanized *APOE* knock-in allele, where exons 2 and 3 and most of exon 4 of the mouse gene were replaced by the human ortholog including exons 2, 3, 4 and some part of the 3' UTR. Furthermore, a knock-in missense point mutation in the mouse *Trem2* gene was also introduced, consisting of an R47H mutation, along with two other silent mutations. Two male and two female mice are included in each condition. Mice were housed socially. All animal procedures were in accordance with institutional, state, and government regulations and approved under institutional animal care and use committee protocols 21049 and 20047.

EasySci-RNA library preparation

Detailed step-by-step EasySci-RNA protocol is included as Supplementary Protocol 1.

Human brain sample

Twenty-four post-mortem human brain samples across two regions (hippocampus and SMTG) and twelve individuals, including six controls and six patients with AD, ranging from 70 to 94 years of age, were collected from the University of Kentucky AD Center Tissue Bank. Each included participant who donated samples for this study signed a relevant consent form (including consent for unrestricted sharing of clinical, pathological and genetic information for dementia research) that was approved by the UK Internal Review Board (UK IRB #44009).

Computational procedures for processing EasySci-RNA libraries

A custom computational pipeline was developed to process the raw fastq files from the EasySci libraries. Similar to our previous studies^{10,11}, the barcodes of each read pair were extracted. Both adaptor and barcode sequences were trimmed from the reads. Second, an extra trimming step is implemented using Trim Galore⁸⁵ with default settings to remove the poly(A) sequences and the low-quality base calls from the cDNA. Afterward, the paired-end sequences were aligned to the genome with the STAR aligner⁸⁶, and the PCR duplicates were removed. Finally, the reads are split into SAM files per cell, and the gene expression is counted using a custom script. The reads from the same cell originating from the short dT and the random hexamer RT primers were counted as independent cells. During the gene counting step, we assigned reads to genes if the aligned coordinates overlapped with the gene locations on the genome. If a read was ambiguous between genes and derived from the short dT RT primer, we assigned the read to the gene with the closest 3' end; otherwise, the reads were labeled as ambiguous and not counted. If no gene was found during this step, we then searched for candidate genes 1,000 bp upstream of the read or genes on the opposite strand. Reads without any overlapped genes were discarded. Similar strategy was used for generating an exon count matrix across cells.

To compare the performance of EasySci-RNA with the commercial 10x Chromium system, we subsampled ~3,800 raw reads/cell from one randomly selected PCR batch of our large-scale mouse brain experiment, a 10x v2 Chromium dataset⁸³, a 10x v3 dataset (<https://www.10xgenomics.com/resources/datasets/5k-adult-mouse-brain-n>

[uclei-isolated-with-chromium-nuclei-isolation-kit-3-1-standard](#)) and a SPLiT-seq dataset⁸³. After the subsampling, the EasySci data were processed with the custom computational pipeline, whereas the 10x Chromium data were processed with 10x Genomics' Cell Ranger software⁸⁷. We removed low-quality cells (unassigned reads >30%, UMIs >20,000 and genes <200) and selected the top 1,000 highest-quality cells from the 10x Chromium dataset⁸³ and a deeply sequenced EasySci-RNA library profiling adult mouse brains. Subsequently, we subsampled these cells to different sequencing depths and quantified the unique transcripts/genes detected per cells. Based on this comparison, we recommend a sequencing depth of no less than 5,000 raw reads per nucleus to ensure adequate coverage and detection of a substantial number of unique molecules.

Cell clustering and annotation analysis

After gene counting, we kept the cells with reads identified by both RT primers. We then merged the reads from the same cells. Low-quality cells were removed based on one of the following criteria: (i) the percentage of unassigned reads > 30%, (ii) the number of UMIs >20,000 and (iii) the detected number of genes <200. We then used the Scrublet⁸⁸ to identify and remove potential doublets. To identify distinct clusters of cells, we subjected the 1,469,111 single-cell gene expression profiles to UMAP visualization and Louvain clustering, similar to our previous study¹⁰. We then co-embedded our data with the published datasets^{2,89,90} through Seurat⁹¹, and clusters were annotated based on overlapped cell types. The annotations were manually verified and refined based on marker genes. DE genes across cell types were identified with the differentialGeneTest() function of Monocle 2 (ref. 92). To identify cell-type-specific gene markers, we selected genes that were DE across different cell types (FDR of 5%, likelihood), with over twofold expression difference between first and second-ranked cell types and TPM >50 in the first-ranked cell types.

Cell subclustering analysis

We selected each main cell type and applied PCA (combined matrix including the 30 principal components derived from the gene-level expression matrix and the first 10 principal components derived from the exon-level expression matrix), UMAP and Louvain clustering similarly to the major cluster analysis. We then merged subclusters that were not readily distinguishable in the UMAP space similar as described before¹⁰. DE genes and exons across cell types were identified with the differentialGeneTest() function of Monocle 2 (ref. 92). To identify subcluster-specific DE genes associated with aging or AD models, we sampled a maximum of 5,000 cells per condition for downstream DE gene analysis using the differentialGeneTest function of the Monocle 2 (ref. 92). The sex of the animals was included as a covariate to reduce sex-specific batch effects.

To detect cellular fraction changes at the subtype level across various conditions, we first generated a cell count matrix by computing the number of cells from every subcluster in each RT well profiled by EasySci-RNA. Each RT well was regarded as a replicate comprising cells from a specific mouse individual. Of note, we repeated the same analysis using the number of cells from each subcluster in each mouse individual (instead of RT well) and the result is highly consistent. We then applied the likelihood ratio test to identify significantly changed subclusters between different conditions, with the differentialGeneTest() function of Monocle 2⁹². Subclusters were removed if they had <20 cells in either the male or female samples. The fold change was calculated by normalizing the number of cells in a cluster by the total number of cells in the corresponding condition, then dividing the normalized values in the case and control conditions after adding a small number (10^{-5}) to reduce the effect of the very small clusters. In addition, we considered subclusters to change significantly only if there was over twofold change between conditions and the q-value was less than 0.05.

Integration analysis with external datasets and to locate the spatial distributions of main cell types and subtypes

To annotate the spatial locations of main cell types, we integrated the EasySci-RNA data with publicly available 10x Visium spatial transcriptomics datasets (<https://www.10xgenomics.com/resources/datasets/mouse-brain-section-coronal-1-standard-1-0-0>, <https://www.10xgenomics.com/resources/datasets/mouse-brain-serial-section-1-sagittal-anterior-1-standard-1-0-0>; <https://www.10xgenomics.com/resources/datasets/mouse-brain-serial-section-1-sagittal-posterior-1-standard-1-0-0>) through a NNLS approach: we first aggregated cell-type-specific UMI counts, normalized by the library size, multiplied by 100,000 and log-transformed after adding a pseudocount. A similar procedure was applied to calculate the normalized gene expression in each spatial spot captured in the 10x Visium dataset. We then applied NNLS regression to predict the gene expression of each spatial spot in 10x Visium data using the gene expression of all cell types recovered in Easy-RNA data, similar to our previous study¹⁰. The same approach¹⁰ was applied to integrate our EasySci-RNA dataset with a large single-cell dataset featuring highly detailed cell type annotations² for identification of shared cellular states in two datasets.

To spatially map EasySci cell subtypes, we first aggregated ~50 single-cell transcriptomes identified by k-means clustering of cells in the UMAP space of subclustering analysis. We then integrated the EasySci-RNA data with the above 10x Visium spatial transcriptomics datasets and a published spatial dataset²⁸, using cell2location²⁷ following the default settings. To establish the corresponding regions of EasySci subclusters, we utilized the regional annotation of the spatial pixels and manually reviewed the anatomical regions of the top 10 pixels with the highest mapping score. To remove low-quality spatial mappings, only mapping scores above 1 were considered.

GM analysis

We performed GM analysis to identify the molecular programs underlying different cell types in the brain. First, we aggregated the gene expression across all subclusters. The aggregated gene count matrix was then normalized by the library size and then log-transformed. Genes were removed if they exhibited low expression (less than 1 in all subclusters) or low variance of expression (that is, the gene expression fold change between the maximum expressed subcluster and the median expression across subclusters is less than 5). The filtered matrix was used as input for UMAP visualization¹⁹ (metric = 'cosine', min_dist = 0.01, n_neighbors = 30). We then clustered genes based on their 2D UMAP coordinates through densityClust package (rho = 1, delta = 1)⁹³.

EasySci-ATAC library preparation and sequencing

The detailed protocol for EasySci-ATAC library preparation is included in Supplementary Protocol 2.

Data processing for EasySci-ATAC

After sequencing, base calls were converted to fastq format and demultiplexed using Illumina's bcl2fastq/v2.19.0.316 tolerating one mismatched base in barcodes (edit distance <2). Downstream sequence processing was similar to sci-ATAC-seq⁹⁴. To compare the performance of EasySci-ATAC with other methods, we extracted reads containing barcodes from cells passing quality control (3,636 cells from one PCR well of the EasySci-ATAC library, 8067 nuclei from the 10x-ATACv2 library and 5,494 nuclei from the sci-ATAC-seq library¹⁵). We normalized for sequencing depth differences by subsampling reads from the 10x-ATACv2 and sci-ATAC-seq library, resulting on average 6,360 raw reads per cell across all three libraries. We processed the data through the same computation pipeline described above. Peak calling was performed on each dataset separately with these parameters: -nomodel -extsize 200 -shift -100 -q 0.05. For peak counting, a union peak set was generated by merging the peaks called from three datasets. Cells

were determined to be accessible at a given peak if a read from a cell overlapped with the peak. The peak-count matrix was generated by a custom python script with the HTseq package⁹⁵.

Cell filtering, clustering and annotation for EasySci-ATAC

We used SnapATAC2/v1.99.99.3^{96,97} to preprocess the EasySci-ATAC dataset. Cells with <1500 fragments and <2 TSS Enrichment were discarded. Potential doublet cells and doublet-derived subclusters were detected using an iterative clustering strategy¹⁰ modified to suit for scATAC-seq data. We then used a deep-learning-based framework scJoint²¹ to annotate main ATAC-seq cell types by using the EasySci-RNA dataset as a reference. First, we subsampled 5,000 cells from each main cell type of the EasySci-RNA dataset, and selected genes detected in more than 10 cells. Then, the gene count matrix and cell type labels of EasySci-RNA, along with the gene activity matrix of EasySci-ATAC were input into the scJoint pipeline with default parameters. Jointed embedding layers calculated from scJoint were used for UMAP visualizations using python package umap/v0.5.3 (ref. 19). Louvain clusters were identified using the Seurat function FindNeighbors() and FindClusters() based on the UMAP coordinates. Cells were assigned to the prediction label with the highest abundance within each louvain cluster. Clusters with low purities (that is, <80% cells were from the highest abundant cell type) were removed. Finally, to validate the integration-based annotations, we selected DE genes identified from the RNA-seq data with the following criteria: fold change between the maximum and the second maximum expressed cell type >1.5, q-value < 0.05, TPM >20 in the maximum RNA group and reads per million >50 in the maximum ATAC group. The top 10 DE genes ranked by fold change were selected using RNA-seq data for each cell type. If there were less than 10 genes passing the cutoff, we selected the top genes ranked by the fold change between the maximum expressed cell type and the mean expression of other cell types. We then calculated the aggregated gene count and gene body accessibility for each cell type. Subcluster level integrations were similar to the main cluster level integrations.

Differential accessible peak analysis

Nonduplicate ATAC-seq reads of cells from each main cell type were aggregated and peaks were called on each group separately with these parameters: -nomodel -extsize 200 -shift -100 -q 0.05 using MACS2/v2.1.1 (ref. 98). To correct for differences in read depth or the number of nuclei per cell type, we converted MACS2 peak scores (-log₁₀(q-value)) to 'score per million'⁹⁹ and filtered peaks by choosing a score-per-million cutoff of 1.3. Peak summits were extended by 250 bp on either side and then merged with bedtools/v2.30.0. Cells were determined to be accessible at a given peak if a read from a cell overlapped with the peak. The peak-count matrix was generated by a custom python script with the HTseq package⁹⁵.

We used R package Signac/v1.7.0 (ref. 100) to perform the dimension reduction analysis using the peak-count matrix. We subsampled 5,000 cells from each main cell type and performed TF-IDF normalization using RunTFIDF(), followed by singular value decomposition using RunSVD() and retained the 2nd to 30th dimensions for UMAP visualizations using RunUMAP(). Differentially accessible peaks across cell types were identified using monocle 2 (ref. 92) with the differentialGeneTest() function. 5,000 cells were subsampled from each cell type for this analysis. Peaks detected in less than 50 cells were filtered out. We selected peaks that were differentially accessible across cell types by the following criteria: 5% FDR (likelihood ratio test), and with TPM >20 in the target cell type.

Transcription factor motif analysis

We used ChromVar/v1.16.0 (ref. 101) to assess the TF motif accessibility using cisBP motif sets curated by chromVARmotifs/v0.2.0 (ref. 101,102). We subsampled 5,000 cells from each main cell type, and calculated the motif deviation score for each single cell using the Signac wrapper

RunChromVAR(). The motif deviation scores of each single cell were rescaled to (0, 10) using R function `rescale()` and then aggregated for each cell type. In addition, we also aggregated the gene expression of each TF in each cell type. We then computed the Pearson correlations between the aggregated motif matrix and aggregated TF expression matrix after scaling across all main cell types. TF analysis at the sub-cluster level was performed similarly with modifications. For each cell type of interest, we selected peaks detected in more than 20 cells and only kept cells with more than 500 reads in peaks. Peaks were resized to 500 bp (± 250 bp around the center) and motif occurrences were identified using `matchMotifs()` function from `motifmatchr/v1.16.0` (ref. 103). The motif deviation matrix was calculated using the `ChromVar` function `computeDeviations()`. Then, the motif deviation scores were rescaled to (0, 10) and aggregated per subcluster. Pearson correlation was calculated between the aggregated motif activity and aggregated TF expression across subclusters after scaling (subclusters with < 20 cells were excluded).

LDSC analysis

The LDSC computational pipeline was modified from Cusanovich et al.¹⁵ and based on the LDSC software¹⁰⁴ (<https://github.com/bulik/ldsc>). Specifically, to integrate human and mouse data, we first used the UCSC utility `liftOver`¹⁰⁵ to lift all GWAS SNPs to the mouse genome. We then took the set of differentially accessible peaks across main clusters and across microglia subclusters, and annotated each SNP according to whether or not it overlapped one of these peaks. We then followed the recommended workflow for running LDSC using HapMap SNPs¹⁰⁶, precomputed files corresponding to 1000 genomes phase 3, excluding the MHC region to generate an LDSC model for each chromosome and peak set. Only main cell types or subclusters containing DE peaks in every chromosome are included in the following analysis.

To calculate enrichments based on each model, we first regenerated the baseline model (version 1.1) provided from the LDSC website and used this as the reference for enrichment calculation. Results for all trait/cluster pairs were gathered into a single file. *P* values were calculated from z-scores assigned to coefficients reported by `ldsc.py` and coefficients were divided by the average per-SNP heritability for traits associated with a given test. Tests were corrected for multiple hypothesis testing using the Benjamini-Hochberg method and only tests with a q-value of 0.05 or lower were considered significant.

Cis-regulatory elements linkage analysis

We first constructed pseudo-cells by aggregating the RNA-seq and ATAC-seq profiles of the same subclusters. Aggregated count matrices were normalized to TPM and log-transformed after adding one pseudocount. We only retained genes and peaks with TPM value greater than 10 in the maximum expressed pseudo-cells. Then, for each gene, we calculated the Pearson correlation coefficient (PCC) between its gene expression and the chromatin accessibility of its nearby accessible sites (± 500 kb from the TSS) across aggregated subclusters. To define a threshold at PCC score, we also generated a set of background pairs by permuting the subcluster ID of the ATAC-seq matrix and with an empirically defined significance threshold of $FDR < 0.01$, to select significant positively correlated *cis*-regulatory element-gene pairs. We only keep the top linked gene with the highest PCC for each peak and distal peaks overlapping with the promoters for other genes were filtered out.

Spatial gene expression profiling of mouse brains

Spatial gene expression analysis experimental protocol was followed according to Visium Spatial Gene Expression User Guide (catalog no. CG000160), Visium Spatial Tissue Optimization User Guide (catalog no. CG000238 Rev A, 10x Genomics) and Visium Spatial Gene Expression User Guide (catalog no. CG000239 Rev A, 10x Genomics).

Transcriptomic aging clock analysis

A ridge regression model was employed to predict the $\ln(\text{age})$ of pseudobulk cells (on average 15 cells merged) utilizing 80% of the pseudobulk cells from 3, 6, and 21-month-old mice. Predicted ages were subsequently calculated for the remaining 20% of WT mice and the entirety of the AD models. Individual models were crafted for each cell type.

Clustering, annotation and differential analysis for human brain samples

A digital gene expression matrix was constructed from the raw sequencing data as described before. To identify distinct clusters of cells corresponding to different cell types in the human brain samples, we co-embedded the human cells from both regions with our mouse brain dataset (up to 5,000 cells randomly sampled from each of 31 cell types), and clusters were annotated based on overlapped cell types. The annotations were manually verified and refined based on marker genes. Following on, the hippocampus and SMTG human dataset were integrated together to construct the same low-dimensional space with only human cells.

DE genes between AD and control samples for each cell type in each region were identified using Monocle 2 (ref. 107,108) with the `differentialGeneTest()` function. Main cell types with less than 50 cells were excluded from the analysis (that is, choroid plexus epithelial cells and vascular leptomeningeal cells in the SMTG). DE genes were filtered based on the following cutoffs: q-value < 0.05 , with fold change (FC) > 1.5 between the maximum and second expressed condition, and with TPM > 50 in the highest expressed condition. To further validate human-mouse shared gene expression changes, we used a recently published AD single-cell dataset from the human prefrontal cortex⁶.

Statistics and reproducibility

Statistical analyses are detailed in figure legends (Fig. 1 and Extended Data Fig. 9) and were performed using R software (version 4.0.1). The number of cells or pseudobulk cells used for the comparisons are detailed in the figure legends and the number of replicates are detailed in Methods. For spatial integration analysis in Fig. 1k,l, Fig. 3g,h, Fig. 4e and Extended Data Fig. 5h,i, each spatial transcriptomic datum includes one section of the experiment.

Reporting summary

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

Data availability

All relevant data generated in this study are deposited to public repositories and are publicly released. Raw and processed data of single-cell RNA-seq/ATAC-seq profiling were deposited at the NCBI Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE212606>).

Code availability

Customized computational scripts of processing EasySci data were deposited in Zenodo¹⁰⁹ (<https://doi.org/10.5281/zenodo.8395492>) and GitHub (<https://github.com/JunyueCaoLab/EasySci>).

References

85. Krueger, F., James, F., Ewels, P., Afyounian, E. & Schuster-Boeckler, B. FelixKrueger/TrimGalore: a wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. *GitHub* <https://doi.org/10.5281/zenodo.5127899> (2021).
86. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

87. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
88. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291 (2019).
89. Yao, Z. et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. *Nature* **598**, 103–110 (2021).
90. Kozareva, V. et al. A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. *Nature* **598**, 214–219 (2021).
91. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
92. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
93. Rodriguez, A. & Laio, A. Machine learning. Clustering by fast search and find of density peaks. *Science* **344**, 1492–1496 (2014).
94. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).
95. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
96. Fang, R. et al. Comprehensive analysis of single cell ATAC-seq data with SnapATAC. *Nat. Commun.* **12**, 1337 (2021).
97. Zhang, K., Zemke, N. R., Armand, E. J. & Ren, B. SnapATAC2: a fast, scalable and versatile tool for analysis of single-cell omics data. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.09.11.557221> (2023).
98. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
99. Corces, M. R. et al. The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
100. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
101. 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).
102. Weirauch, M. T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
103. Schep, A. motifmatchr: Fast Motif Matching in R. *GitHub* <https://github.com/GreenleafLab/motifmatchr/> (2017).
104. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
105. Hinrichs, A. S. et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
106. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
107. 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).
108. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
109. Sziraki, A. & Lu, Z. Computational pipeline for processing EasySci data. *Zenodo* <https://doi.org/10.5281/zenodo.8395492> (2023).

Acknowledgements

We thank all members of the Cao lab for helpful discussions and feedback. We thank J. Shendure (University of Washington) for insightful feedback on this work. We also thank members of the Rockefeller University Genomics Resource Center (SCR_020986), High-Performance Computing Resource Center and Comparative Bioscience Center for their exceptional assistance with library sequencing and animal maintenance. This work was funded by grants from the National Institutes of Health (DP2HG012522, R01AG076932 and RM1HG011014 to J.C.; P30AG072946 and P01AG078116 to P.T.N.; and R01AG072758 to L.G.) and the Sagol Network GerOmic Award (J.C.). This work is partly supported by the Pershing Square Foundation, Bill Ackman and Neri Oxman.

Author contributions

J.C. and W.Z. conceptualized and supervised the project. J.L. and A.S. developed the experimental and computational pipeline for EasySci-RNA profiling of all samples. G.B. and Z.L. developed the experimental and computational pipeline for EasySci-ATAC profiling of all samples. A.A. performed the 10x Visium spatial transcriptomics experiment. S.A. and P.N. processed the human brain samples for single-cell profiling experiments. A.S. and Z.L. performed the downstream analysis with assistance from E.M., A.L., A.E., C.S., Z.X., Z.Z. and J.B. J.C., W.Z., Z.L. and A.S. wrote the paper with input and biological insight from P.N., L.G. and other co-authors.

Competing interests

J.C., W.Z., A.S. and J.L. are inventors on pending patent applications related to EasySci-RNA-seq. The other authors declare no competing interests.

Additional information

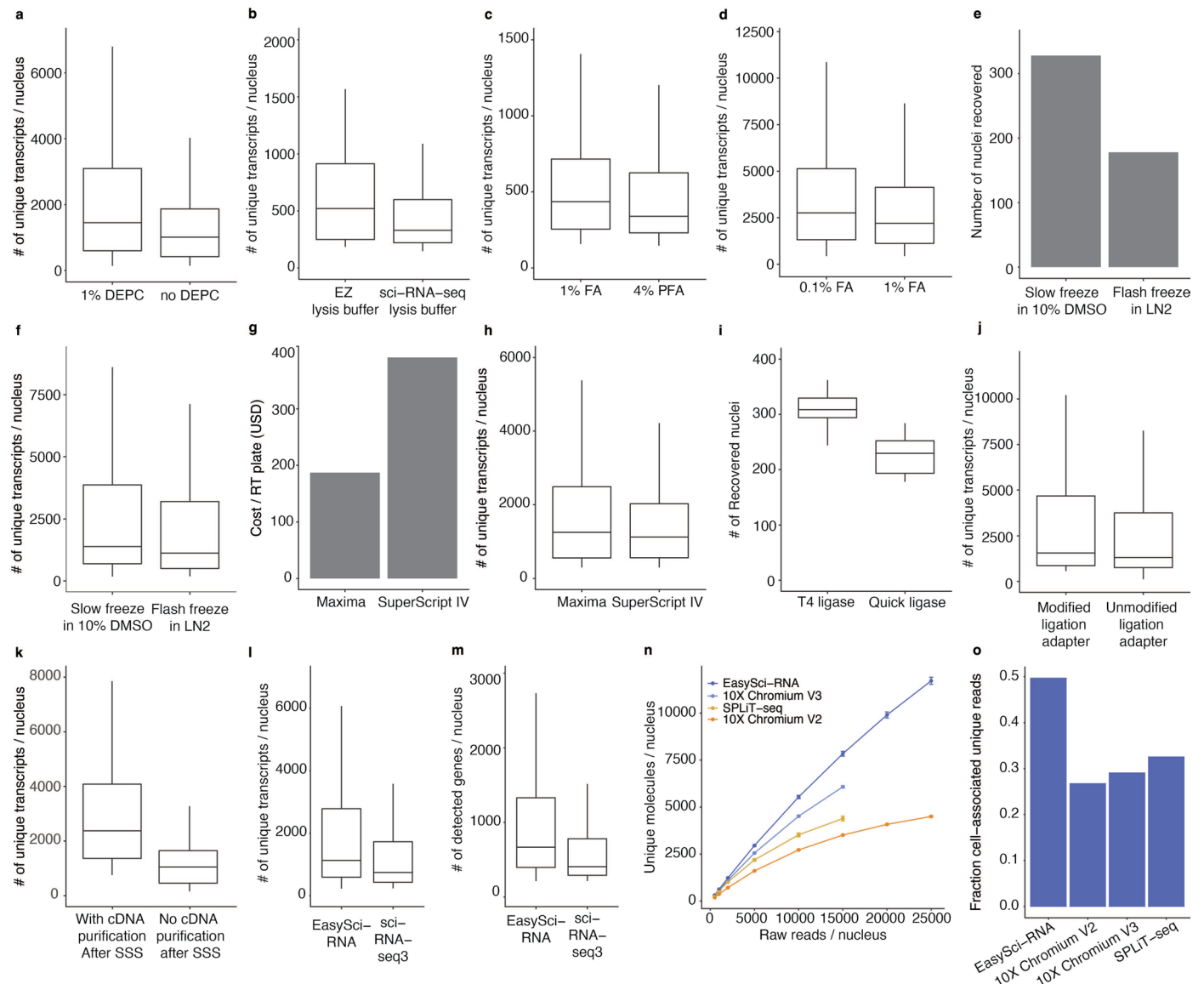
Extended data is available for this paper at <https://doi.org/10.1038/s41588-023-01572-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-023-01572-y>.

Correspondence and requests for materials should be addressed to Wei Zhou or Junyue Cao.

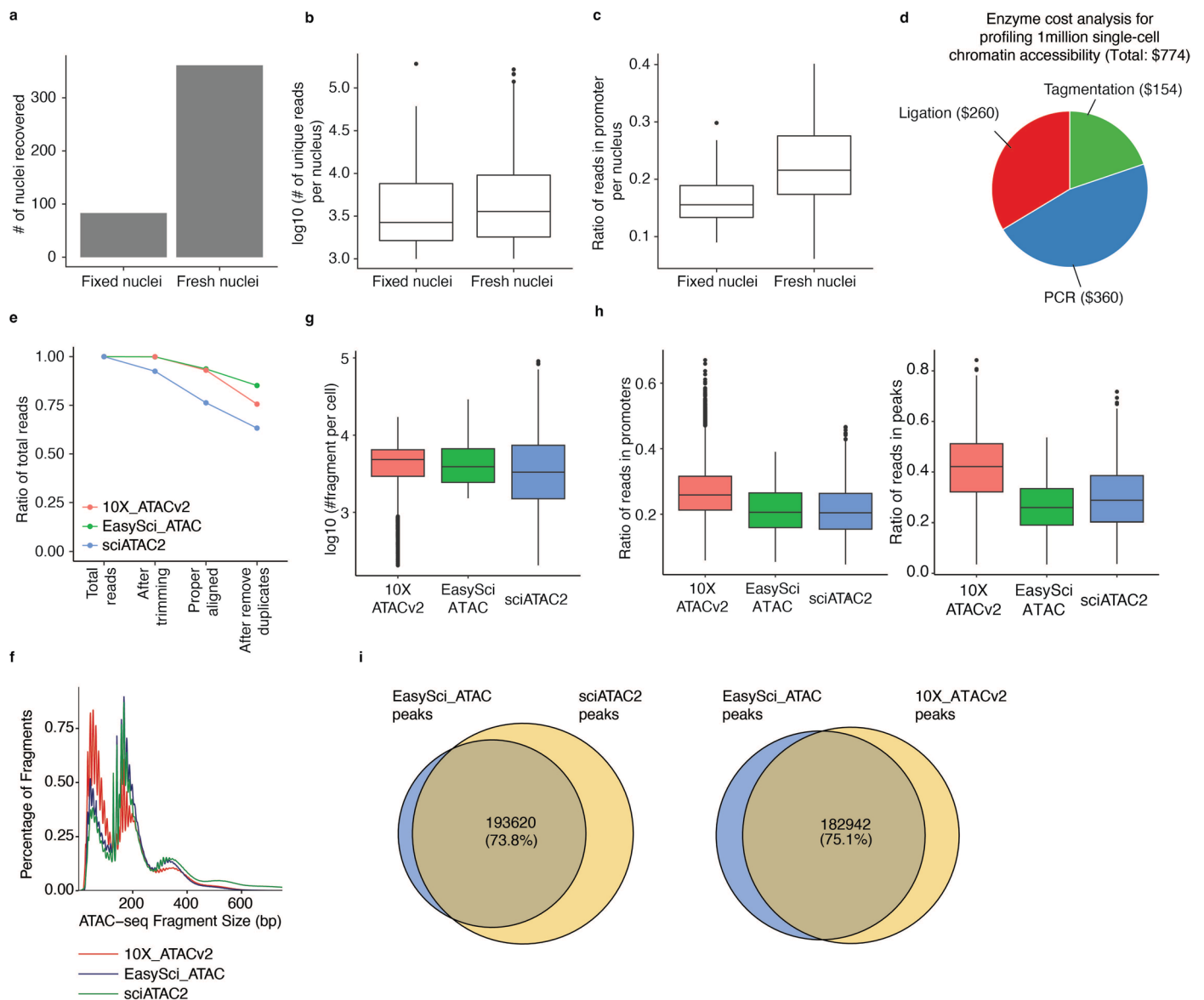
Peer review information *Nature Genetics* thanks Inge Holtman, Luciano Martelotto and Vivek Swarup for their contribution to the peer review of this work.

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



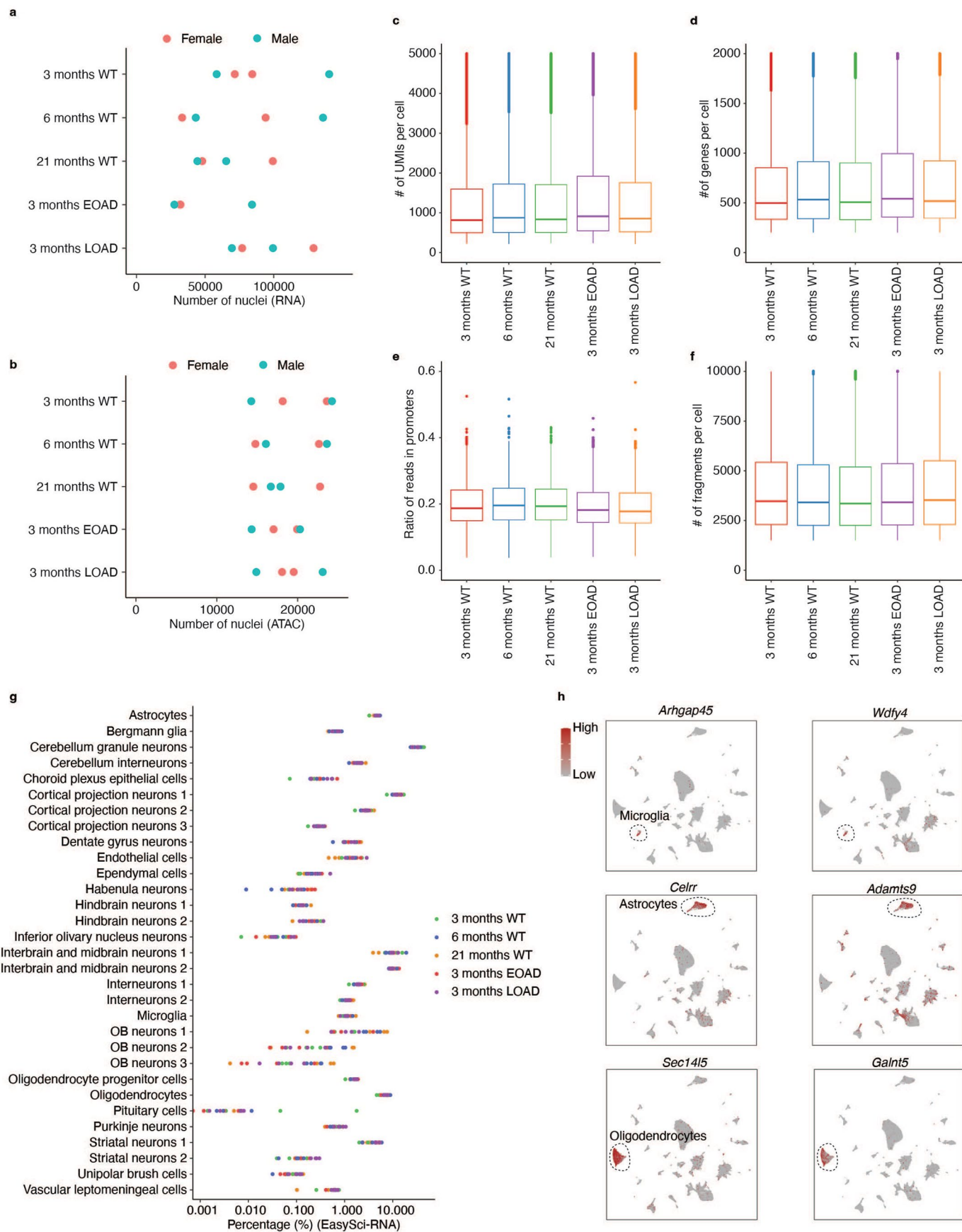
Extended Data Fig. 1 | Representative examples showing the performance of optimized conditions of EasySci-RNA. **a-b.** Box plots showing the number of UMIs per nucleus lysed with 1% DEPC (n = 1,596 cells) vs. no DEPC (n = 3,036 cells) in lysis buffer (a); or EZ lysis buffer (n = 50 cells) vs. sciRNA-seq3 nuclei lysis buffer (n = 96 cells)¹¹ (b). **c-d.** Box plot showing the number of UMIs detected per nucleus across fixation conditions: 1% formaldehyde (n = 403 cells) vs 4% paraformaldehyde (n = 358 cells) (c); 0.1% formaldehyde (n = 2,952 cells) vs. 1% formaldehyde (n = 2,211 cells) (d). **e-f.** Slow freezing condition in 10% DMSO (n = 328 cells) outperformed the flash freezing condition (n = 178 cells) by increasing the number of nuclei recovered (e) and the number of UMIs per nucleus (f). **g-h.** Maxima reverse transcriptase (n = 5,197 cells) reduces the enzyme cost (g) without affecting the number of transcripts detected per nucleus compared to SuperScript™ IV reverse transcriptase (n = 4,071 cells) for profiling mouse brain cells (h). **i.** EasySci-RNA used T4 ligase (n = 40 wells) instead of quick ligase (n = 16 wells) for a higher recovery rate of nuclei. **j.** Box

plot showing the number of unique transcripts detected per nucleus comparing chemically modified ligation primers (n = 499 cells) to unmodified adapters (n = 499 cells). **k.** Additional cDNA purification step after second strand synthesis (n = 90 cells) increased the number of unique transcripts per nucleus compared to without cDNA purification (n = 90 cells). **l-m.** Comparison of the number of unique transcripts and genes in EasySci-RNA (n = 11,501 cells) and sciRNA-seq3 (n = 117 cells) with the same sequencing depth (~2,500 reads/cell). **n.** Line plots showing the median number (with standard error) of unique transcripts per nucleus from a deep sequenced small-scale EasySci-RNA, a 10X V2 library⁸³, a 10X V3 library and a SPLiT-seq library³³, all profiling mouse brains (Methods). **o.** Comparison of the fraction of cell-associated unique transcripts from the total number of raw reads across different techniques at the same sequencing depth (~3,800 raw reads/cell). Boxes in box plots indicate the median and interquartile range (IQR) with whiskers indicating 1.5X IQR.



Extended Data Fig. 2 | Representative examples showing the performance of optimized conditions of EasySci-ATAC and quality comparison with other single-cell ATAC protocols. a-c. We compared two fixation conditions: nuclei were either fixed with 1% formaldehyde for 10 minutes at room temperature ($n = 63$ nuclei) or directly used for tagmentation without fixation ($n = 361$ nuclei). The unfixed condition outperformed the fixed condition by increasing cell recovery (**a**), the number of reads (**b**), and the ratio of reads in promoters (**c**) per nucleus. **d.** Pie chart showing the estimated enzyme cost compositions of library preparation for profiling 1 million single-cell chromatin accessibility profiles using *EasySci-ATAC*. **e.** Lineplot showing the ratio of reads loss during each data processing step comparing *EasySci-ATAC*, 10X-ATACv2 and sci-ATAC-seq (referred as 'sciATAC2' in the figure). **f.** Histogram showing the fragment length

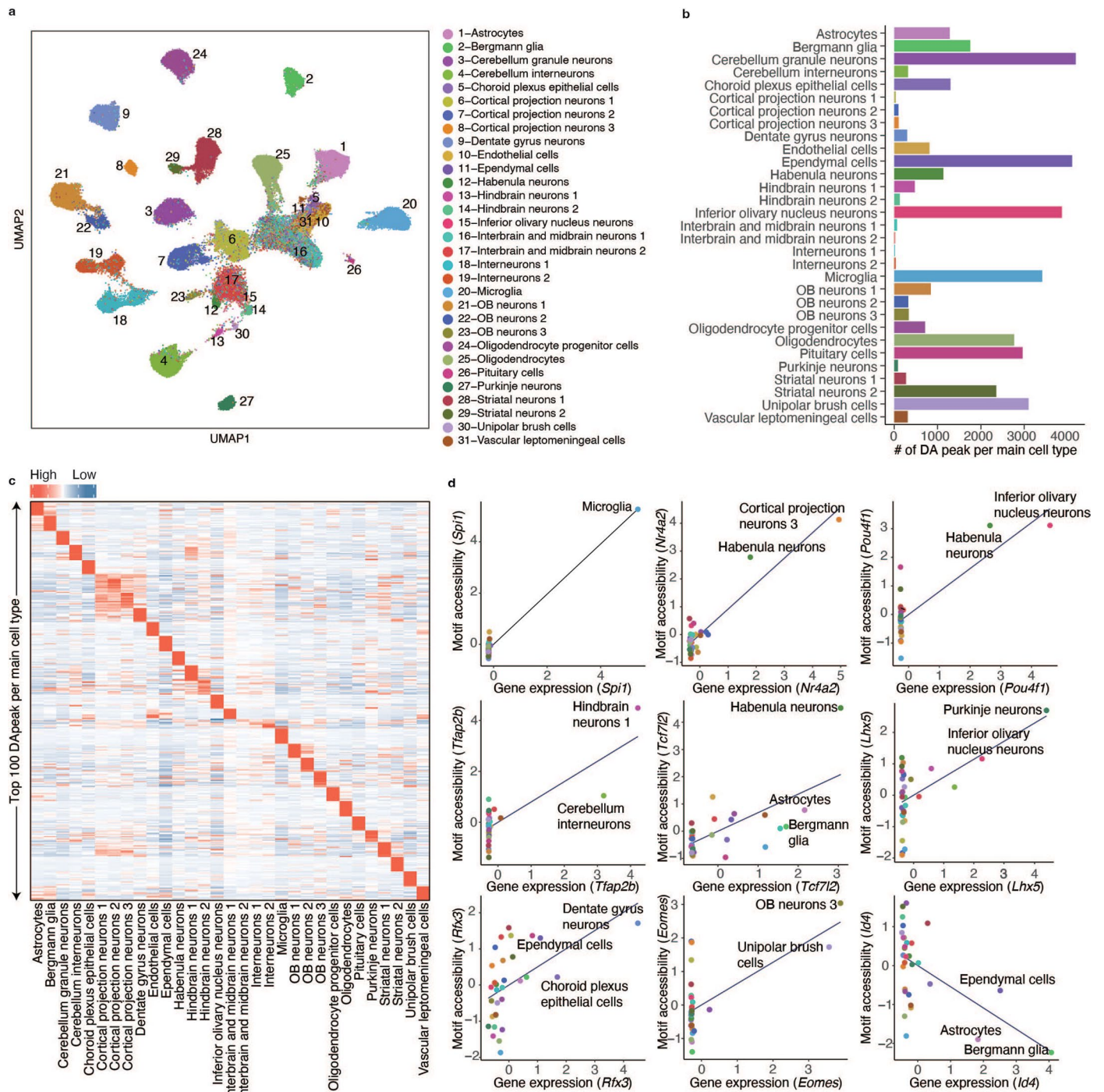
distributions across *EasySci-ATAC*, 10X-ATACv2 and sci-ATAC-seq. **g.** Box plot showing the number of unique fragments comparing *EasySci-ATAC* ($n = 3,636$ cells), 10X-ATACv2 ($n = 6,489$ cells) and sci-ATAC-seq ($n = 5,494$ cells). **h.** Box plots showing the number of reads mapped to promoters (left, defined as ± 1 kb around TSS) and peaks (right) comparing *EasySci-ATAC* ($n = 3,636$ cells), 10X-ATACv2 ($n = 6,489$ cells) and sci-ATAC-seq ($n = 5,494$ cells). Peak calling was performed on each dataset separately and peaks were merged to a union peak set. **i.** Pieplot showing the number of peaks that can be repeatedly identified between *EasySci-ATAC* and sci-ATAC-seq (left) and between *EasySci-ATAC* and 10X-ATACv2 (right). Boxes in box plots indicate the median and interquartile range (IQR) with whiskers indicating 1.5X IQR.



Extended Data Fig. 3 | See next page for caption.

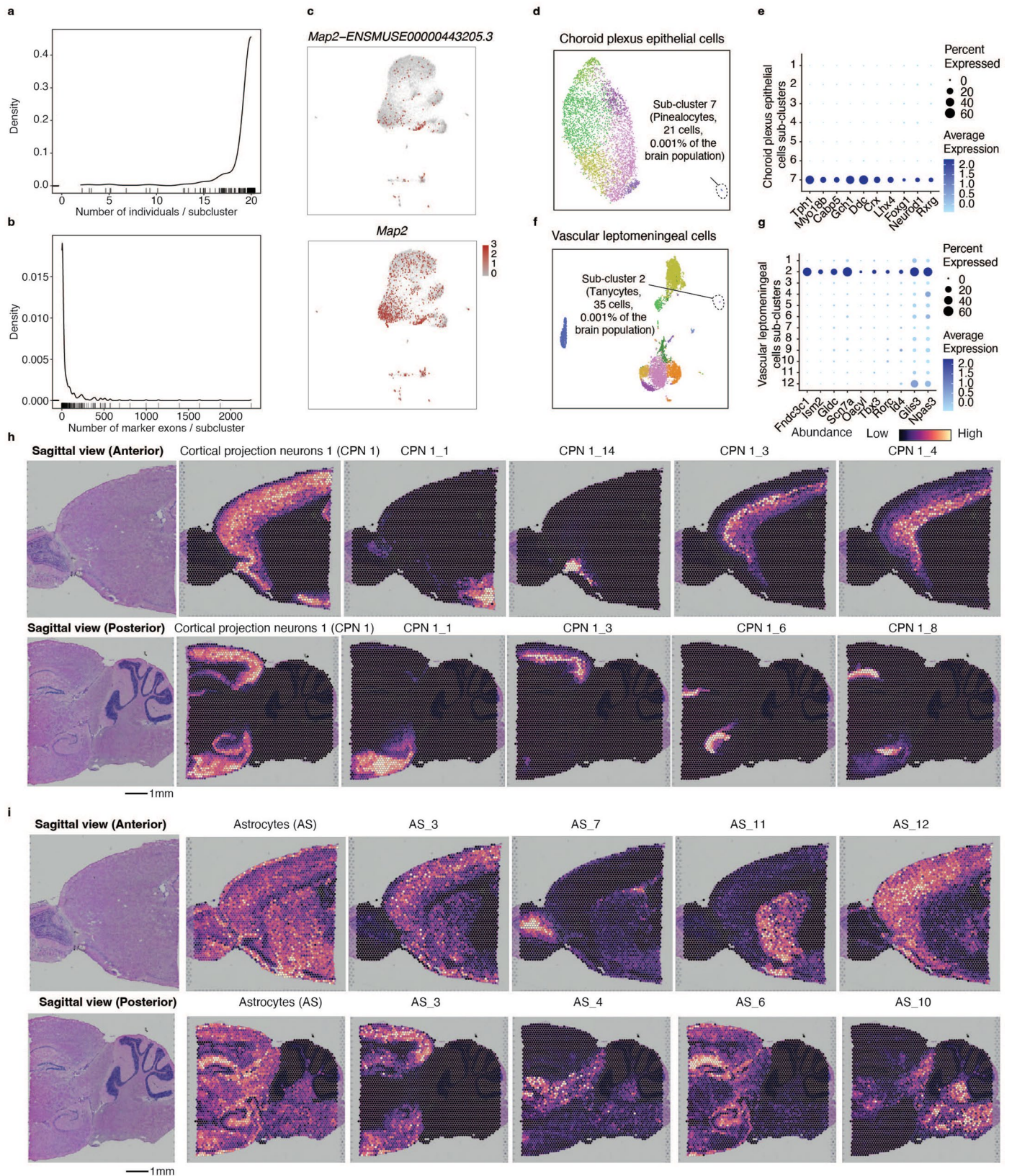
Extended Data Fig. 3 | Performance of EasySci-RNA/ATAC profiling of mouse brain samples and identification of main brain cell types and cell-type-specific markers by EasySci-RNA. a-b. Scatter plots showing the number of single-cell transcriptomes (a) and single-cell chromatin accessibility (b) profiled in each mouse individual across five conditions, colored by sex. **c-d.** Box plots showing the number of unique transcripts (c) and genes (d) detected per nucleus in each condition profiled by *EasySci-RNA* (3-month WT: n = 354,842 cells, 6-month WT: n = 306,537 cells, 21-month WT: n = 257,017 cells, 3-month EOAD: n = 175,673 cells, 3-month LOAD: n = 375,042 cells). **e-f.** Box plots showing the ratio of reads in promoters (e) and the number of unique fragments (f) per cell in

each condition profiled by *EasySci-ATAC* (3-month WT: n = 80,259 cells, 6-month WT: n = 77,101 cells, 21-month WT: n = 71,841 cells, 3-month EOAD: n = 71,548 cells, 3-month LOAD: n = 75,560 cells). **g.** The recovered cell percentage from every main cell type is shown across different replicates. The samples are color-coded based on the condition of origin. **h.** UMAP plots showing the gene expression of identified novel markers for microglia (*Arhgap45*, *Wdfy4*), astrocytes (*Clerr*, *Adamts9*), and oligodendrocytes (*Sec14l5*, *Galnt5*). UMI counts for these genes are scaled by the library size, log-transformed, and then mapped to Z-scores. For all box plots, boxes indicate the median and interquartile range (IQR) with whiskers indicating 1.5X IQR.



Extended Data Fig. 4 | Characterization of cell-type-specific chromatin accessibility and key TF regulators using EasySci-ATAC. a. UMAP plot of the EasySci-ATAC dataset subsampled to 5,000 cells per cell type (or all cells if the number of cells is less than 5,000), colored by main cell types in Fig. 1h. The analysis used the peak-count matrix without integration with the EasySci-RNA dataset. **b.** Bar plot showing the number of cell-type-specific peaks for each main cell type (defined as differential accessible (DA) peaks across main cell types with q-value < 0.05 and TPM > 20 in the target cell type). **c.** Heatmap showing the aggregated accessibility of top 100 DA peaks per cell type (ranked by fold change

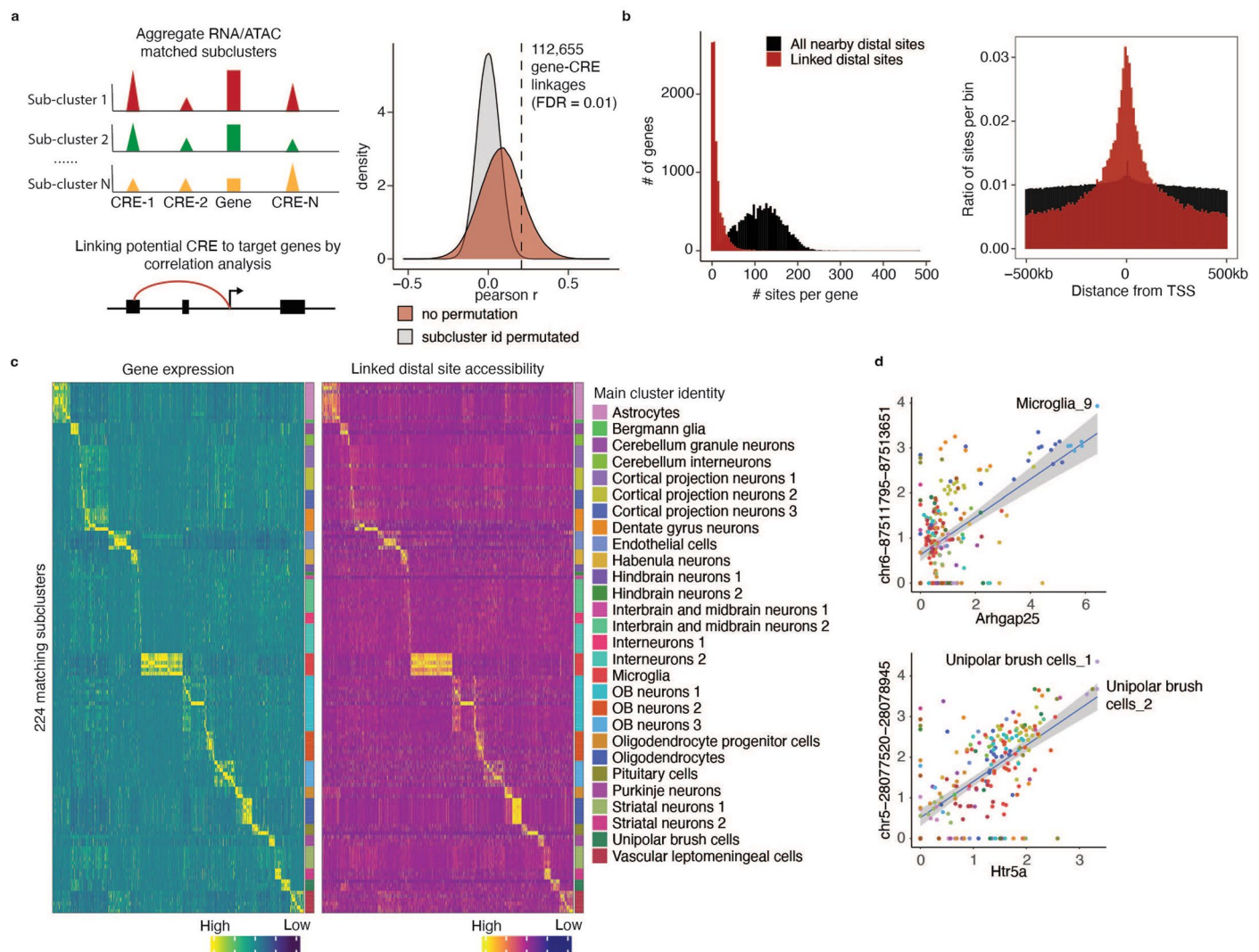
between the maximum and the second accessible cell type). Unique counts for cell-type-specific peaks are first aggregated, normalized by the library size, and then mapped to Z-scores. **d.** Scatter plots showing the correlation between gene expression and motif accessibility of cell-type-specific TF regulators, together with a linear regression line. TF gene expressions are calculated by aggregating scRNA-seq gene counts for each main cluster, normalized by the library size, and then mapped to Z-scores. TF motif accessibilities are quantified by chromVar¹⁰¹, then aggregated per main cell type and mapped to Z-scores (Methods).



Extended Data Fig. 5 | See next page for caption.

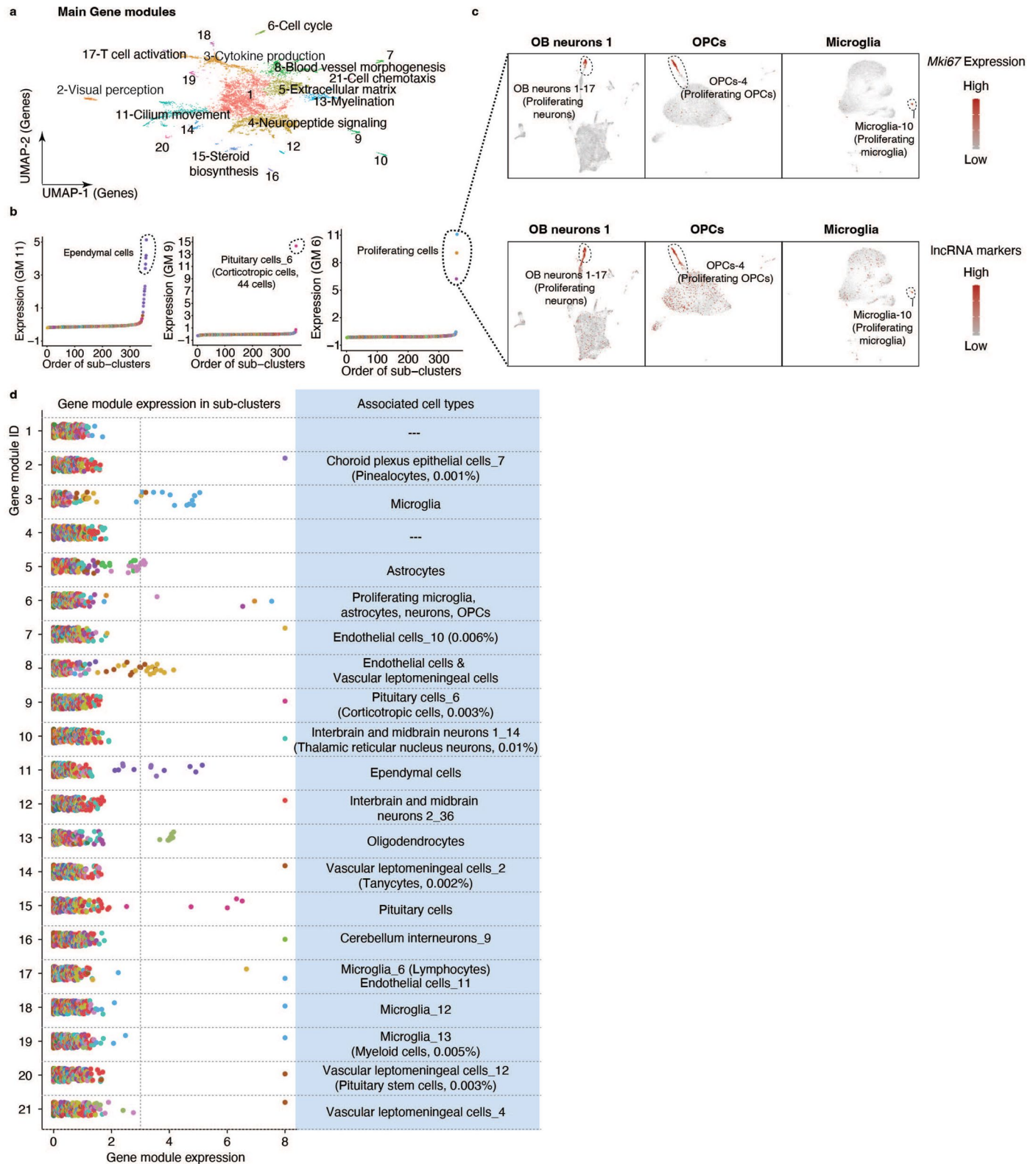
Extended Data Fig. 5 | Characterizations of cellular subtypes in the mouse brain. **a.** Density plot showing the number of individuals per subcluster. The rug plot below the density plot represents the individual subclusters. **b.** Density plot of the number of marker exons per subcluster. The rug plot below the density plot represents the individual subclusters. **c.** UMAP plots, same as Fig. 2a based on both gene and exon-level expression, showing the specific expression of an example exon marker *Map2-ENSMUSE00000443205.3* (top) of microglia subcluster 8 and the lack of specificity of its corresponding gene *Map2* (bottom). Single-cell gene/exon expression was normalized first by library size, log-transformed, and then scaled to Z-scores. **d.** UMAP visualizations showing subclustering analysis for choroid plexus epithelial cells colored by subcluster

IDs, highlighting the rare subcluster shown in Fig. 2c. **e.** Dot plot showing the expression of selected marker genes for choroid plexus epithelial cells-7, including both normal genes (left five genes) and transcription factors (right five genes). **f.** UMAP visualizations showing subclustering analysis for vascular leptomenigeal cells colored by subcluster IDs, highlighting the rare subclusters shown in Fig. 2c. **g.** Dot plot showing the expression of selected marker genes for vascular leptomenigeal cells-2, including both normal genes (left five genes) and transcription factors (right five genes). **h-i.** Mouse brain sagittal sections showing spatial abundances of main cell types and related subclusters for cortical projection neurons 1 (**h**) and astrocytes (**i**) in anterior (top) and posterior (bottom) regions, estimated using the cell2location²⁷.



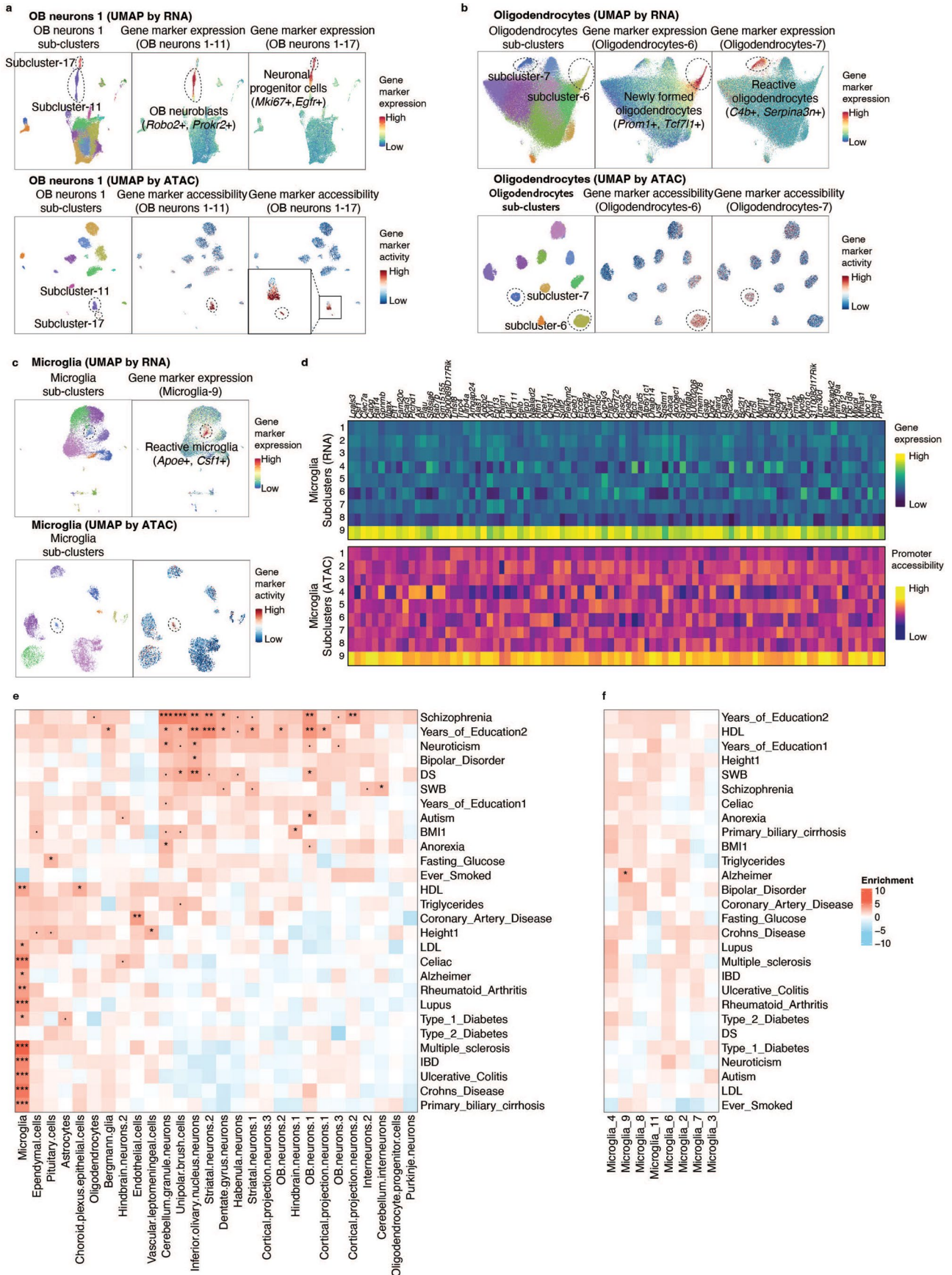
Extended Data Fig. 6 | Cis-regulatory elements regulation of gene expression at the subcluster level. **a.** Overview of analytical steps (left) and density plot showing the distribution of Pearson correlation coefficients between gene expression and the accessibility of distal linked site (colored in red) or all nearby accessible elements (within ± 500 kb of the promoter, colored in gray) across pseudo-cells. Dash line indicates the Pearson correlation coefficient cutoff with 0.01 FDR. **b.** Left, histogram showing the number of accessible sites per gene. Right, histogram shows the distance distribution of accessible sites within 500 kb of genes. Both plots include all nearby accessible sites (colored in

black) and the linked accessible sites (colored in red). **c.** Heatmaps showing the concordant expression of genes (left) and accessibility of putative distal linked sites (right) across all matched subclusters. Rows represent the aggregated gene expression or peak accessibility for a given subcluster. The raw aggregated RNA-seq and ATAC-seq data was normalized first by the total number of reads for each subcluster and then scaled to Z-score across all subtypes. **d.** Correlation plot showing examples of subcluster-specific genes and linked accessible sites for microglia-9 and unipolar brush cells-2. Gray area indicates 95% confidence intervals around the linear regression line.



Extended Data Fig. 7 | Identification of key molecular programs underlying cell type specificity in the mouse brain. **a.** UMAP visualizations of genes colored by identified gene module IDs. **b.** Scatter plots showing examples of gene modules and their expression levels across subclusters (ordered by the level of gene module expression): GM-11 is specific to ependymal cells; GM-9 is specific to pituitary cell-6 (corticotropic cells); GM-6 marks four proliferating subclusters from different main cell types. UMI counts for genes from each gene module are scaled for library size, log-transformed, mapped to Z-scores and then aggregated. **c.** UMAP visualization showing three proliferating subclusters

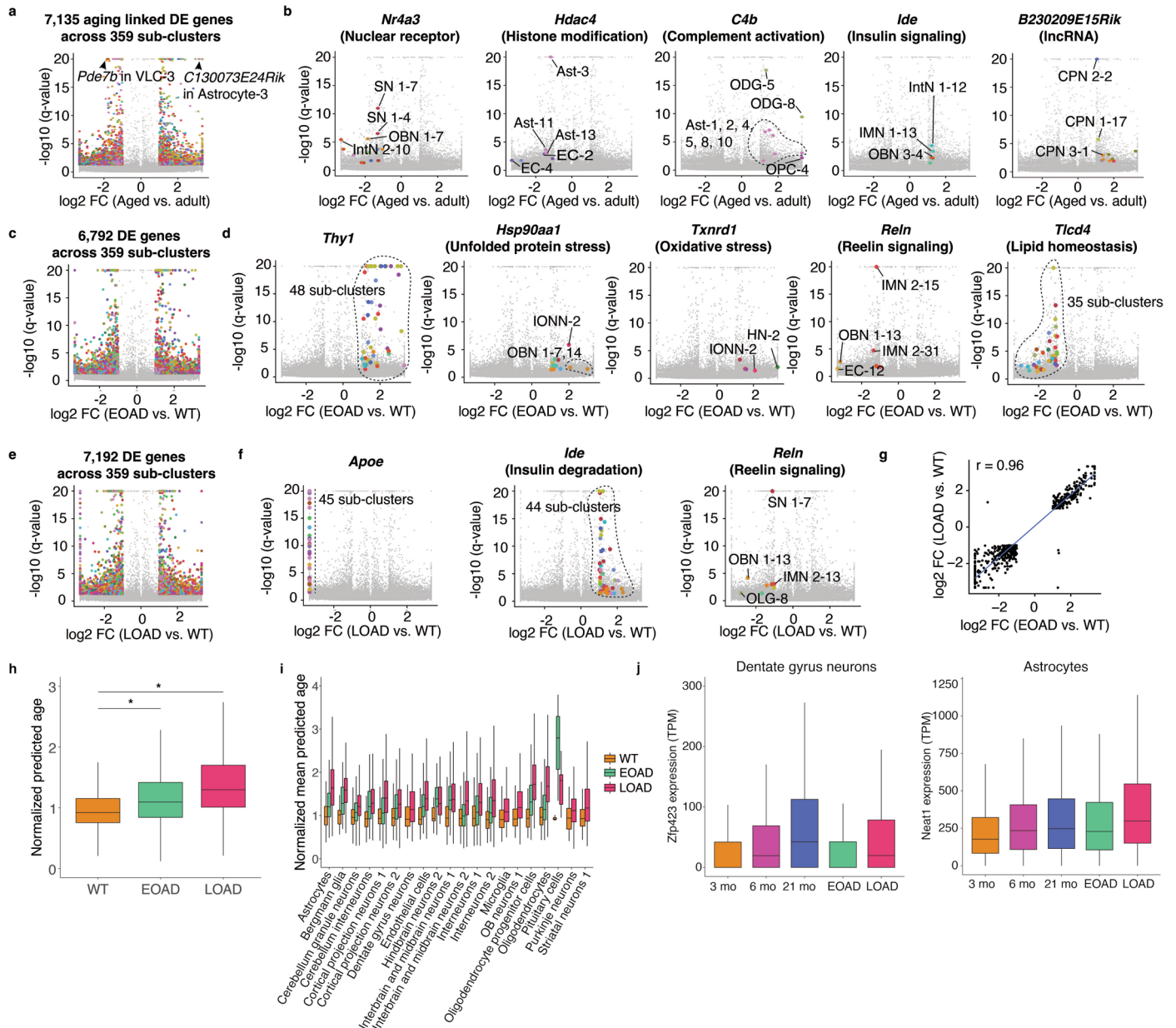
identified from OB neurons 1, oligodendrocyte progenitor cells, and microglia, colored by the normalized expression of canonical proliferating marker *Mki67* (top) and the aggregated expression of lncRNAs in GM-6 (bottom). UMI counts are first normalized by library size, log-transformed, and then mapped to Z-scores. OPCs, oligodendrocyte progenitor cells. **d.** Scatter plot showing the expression of each gene module across 359 subclusters. The associated cell types were annotated on the plot. UMI counts for genes from each gene module are scaled for library size, log-transformed, mapped to Z-scores and then aggregated.



Extended Data Fig. 8 | See next page for caption.

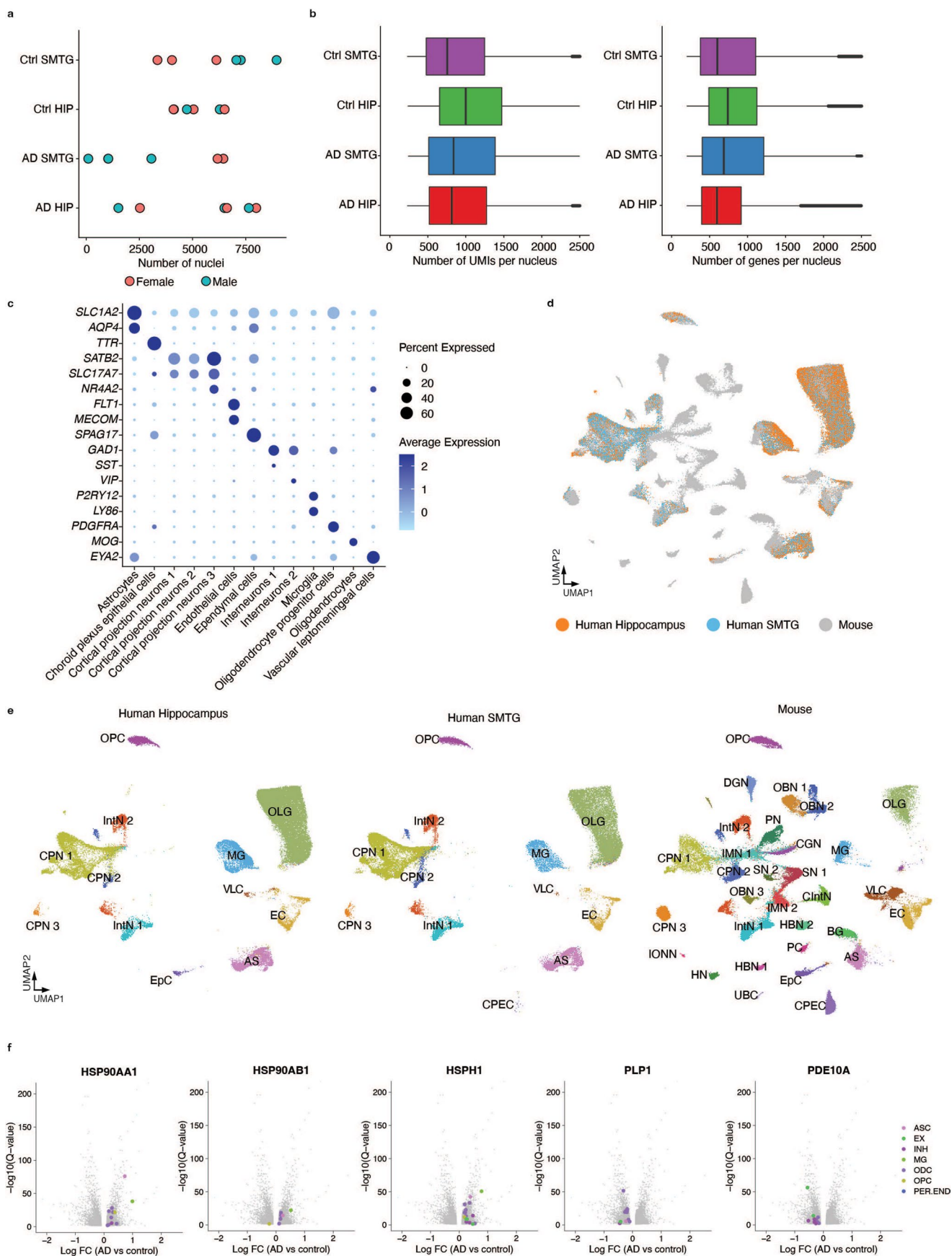
Extended Data Fig. 8 | Identifying aging-associated subclusters in *EasySci-ATAC* and cell-type-specific enrichment of GWAS heritability of human phenotypes. **a.** UMAP visualization showing OB neurons 1-11 and OB neurons 1-17 identified from *EasySci-RNA* (top) and *EasySci-ATAC* (bottom), colored by subcluster id (left), aggregated gene expression or gene activity of OB neurons 1-11 gene markers (middle) and OB neurons 1-17 gene markers (right). **b.** UMAP visualization showing oligodendrocytes-6 and oligodendrocytes-7 identified from *EasySci-RNA* (top) and *EasySci-ATAC* (bottom), colored by subcluster id (left), aggregated gene expression or gene activity of oligodendrocytes-6 gene markers (middle) and oligodendrocytes-7 markers (right). **c.** UMAP visualization showing microglia-9 identified from *EasySci-RNA* (top) and *EasySci-ATAC* (bottom), colored by subcluster id (left), aggregated gene expression or gene

activity of microglia-9 gene markers (right). Subcluster marker genes were identified by differential expression analysis using scRNA-seq data (Methods). **d.** Heatmap showing the gene expression (top) and the promoter accessibility (bottom) of microglia-9 enriched genes across subclusters. The *EasySci-RNA* data (UMI count matrix) and *EasySci-ATAC* data (read count matrix) were aggregated per subcluster, normalized by the total number of reads, column centered, and scaled. Of note, rare subclusters from RNA-seq data that were not detected in ATAC-seq data were not included in this analysis. **e-f.** Heatmap showing the results of LDSC analysis of the SNPs associated with the indicated phenotypes in DEpeaks across main cell types (**e**) or across microglia subtypes (**f**), colored by z score of regression coefficient. *FDR < 0.05, **FDR < 0.01, ***FDR < 0.001. Only cell types that contain DEpeaks from all autosomes are included in this analysis.



Extended Data Fig. 9 | Identifying aging and AD pathogenesis-associated gene expression signatures. a-f. Volcano plots showing the differentially expressed (DE) genes between adult (6 months) and aged (21 months) mice (a-b), between WT and EOAD (c-d) and between WT and LOAD (e-f). Significantly changed genes are colored by the main cell type identity. Examples with concordant changes across subclusters are labeled with related biological pathways. **g.** Scatter plot shows the correlation of the \log_2 -transformed fold changes of 559 DE genes significantly changed within the same subcluster in both AD models compared to WT. **h.** Box plot displays the predicted age of 3-month-old WT mice ($n = 4745$ pseudobulk cells) against AD models (EOAD: $n = 11959$; LOAD: $n = 25152$) across various cell types, normalized by the mean predicted age of the WT samples. Stars indicate $p\text{-value} < 0.05$ (WT vs EOAD: $p = 7.6 \times 10^{-142}$; WT vs LOAD: $p = 0$) using the two-sided Wilcoxon rank-sum test. **i.** Box plots displaying the predicted biological age of 3-month-old wild-type mice against AD models, showing cell types and conditions demonstrating significant

differences (FDR adjusted $p\text{-value} < 0.05$, two-sided Wilcoxon rank-sum test.). Values are normalized by the mean predicted age of the WT samples. Number of pseudobulk cells per box plot (left to right): 179, 615, 1190, 37, 85, 179, 1819, 3272, 7875, 87, 220, 498, 493, 1701, 3057, 85, 359, 673, 62, 338, 65, 194, 424, 11, 30, 56, 605, 1096, 2732, 427, 1389, 2403, 77, 261, 483, 49, 158, 298, 63, 276, 50, 302, 59, 216, 432, 293, 880, 1837, 19, 3, 8, 35, 226, 118, 1178. **j.** Box plots showing *Zfp423* expression in dentate gyrus neurons (3 months = 301, 6 months = 204, 21 months = 334, EOAD = 192 and LOAD = 338 pseudobulk cells) and *Neat1* expression in astrocytes (3 months = 968, 6 months = 969, 21 months = 700, EOAD = 615 and LOAD = 1190 pseudobulk cells) across various conditions. Both genes, ranking among the top predictors of age in their respective cell types, exhibited increased expression in the 3-month-old AD models (*Zfp423* solely in the LOAD model, *Neat1* in both AD models). For all box plots, boxes indicate the median and interquartile range (IQR) with whiskers indicating 1.5X IQR.



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

Extended Data Fig. 10 | Quality control of the human brain dataset and conserved gene expression changes across mouse AD models and human AD samples from the prefrontal cortex. **a.** Scatter plot showing the number of single-cell transcriptomes profiled in each human sample in two regions, colored by sexes. **b.** Box plots showing the number of unique transcripts (left) and genes (right) detected per nucleus profiled in the human dataset (Control SMTG: $n = 36,781$ cells, Control HIP (hippocampus): $n = 30,801$ cells; AD SMTG: $n = 17,864$ cells, AD HIP: $n = 32,794$ cells). **c.** Dotplot showing the markers for the main cell types identified in the human dataset. **d-e.** UMAP plot showing the integration between human and mouse cells, colored by the dataset (**d**) and main cell types (**e**). AS, astrocytes; BG, Bergmann glia; CGN, cerebellum granule neurons; ClntN, cerebellum interneurons; CPEC, choroid plexus epithelial cells; CPN1, cortical projection neurons 1; CPN2, cortical projection neurons 2; CPN3, cortical projection neurons 3; DGN, dentate gyrus neurons; EC, endothelial cells;

EpC, ependymal cells; HN, habenula neurons; HBN1, hindbrain neurons 1; HBN2, hindbrain neurons 2; IONN, inferior olivary nucleus neurons; IMN1, interbrain and midbrain neurons 1; IMN2, interbrain and midbrain neurons 2; IntN1, interneurons 1; IntN2, interneurons 2; MG, microglia; OBN1, OB neurons 1; OBN2, OB neurons 2; OBN3, OB neurons 3; OPC, oligodendrocyte progenitor cells; OLG, oligodendrocytes; PC, pituitary cells; PN, purkinje cells; SN1, striatal neurons 1; SN2, striatal neurons 2; UBC, unipolar brush cells; VLC, vascular leptomeningeal cells. **f.** Volcano plots showing AD-associated gene expression changes from human prefrontal cortex samples⁶ colored by main cell types. These genes show consistent changes in multiple cell subclusters between mouse AD models, human hippocampus and SMTG samples, and human prefrontal cortex samples from the above-mentioned publication. ASC, astrocytes; EX, excitatory neurons; INH, inhibitory neurons; MG, microglia; ODG, oligodendrocytes; OPC, oligodendrocyte progenitor cells; PER.END, pericyte/endothelial cells.

Reporting Summary

Nature Portfolio 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 Portfolio 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 except Illumina bcl2fastq/v2.19.0.316 for base calling.

Data analysis Common, freely available sequencing data analysis software was used to analyze data, as described in Methods: python/v2.7.13, trim_galore/v0.4.1, STAR/v2.5.2b, bedtools/v2.30.0, R/v4.1, Seurat/v4.1, monocle/v2.22.0, Signac/v1.7.0, SnapATAC2/v.1.99.99.3, scJoint/v1.0, umap/v0.5.3&v0.3.2, MACS2/v2.1.1, chromVar/v.1.16.0, chromVarmotifs/v0.2.0, motifmatchR/v.1.16.0, LDSC/v1.0.1. Scripts for processing EasySci data was uploaded to github (<https://github.com/JunyueCaoLab/EasySci>) and have been referred in the manuscript.

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 Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Processed and raw data can be downloaded from NCBI GEO (GSE212606) and have been referred in the data availability statement in the manuscript.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	We use the term sex throughout, and in our study this refers to genetic sex, determined by the composition of sex chromosomes. Balanced male and female samples were included in this study.
Reporting on race, ethnicity, or other socially relevant groupings	Socially relevant grouping were not taken into account to select the population of interest.
Population characteristics	Twelve individuals, including six controls and six alzheimer's disease patients ranging from 70 to 94 in age were used.
Recruitment	Samples were collected from the University of Kentucky AD Center Tissue Bank.
Ethics oversight	Each included participant who donated samples for this study signed a relevant consent form (including consent for unrestricted sharing of clinical, pathological, and genetic information for dementia research) that was approved by the UK Internal Review Board (UK IRB #44009).

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

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	In our previous single-cell genomic studies consisting of similar or lower numbers of replicates, we were able to detect significantly changed cell types with highly variable cell population sizes. For downstream analysis, we will in-silico sort and aggregate cells from the same cell type and individual, followed by differential expression analysis. We therefore anticipate that the sample size will provide sufficient power to detect significant change of cell types or DE genes in inter-condition analysis.
Data exclusions	No data were excluded from the study.
Replication	We confirm all attempts at replication were successful
Randomization	The experiments were carried out in a pooled manner.
Blinding	For collection brain tissues from different mice groups, blinding was not possible, since we need to record the genotype of each mice group and compare between conditions

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

Methods

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

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

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s) HEK293T, NIH/3T3 cells were gift from Dr. Jay Shendure, University of Washington

Authentication The cell lines were authenticated.

Mycoplasma contamination Cell lines were not tested for Mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register) No commonly misidentified cell lines were used.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals C57BL/6 wild-type mouse brains at three months (n=4), six months (n=4), and twenty-one months (n=4) were collected in this study together with 3 months old SxFAD (JAX stock #034840) (n=4) and APOE*4/Trem2*R47H (JAX stock #028709) (n=4) models. Two male and two female mice are included in each condition.

Wild animals No wild animals were used.

Reporting on sex Sex was determined by the composition of sex chromosomes. Both male and female mice were included in each condition. No sex-

Field-collected samples No field-collected samples were used.

Ethics oversight All animal procedures were in accordance with institutional, state, and government regulations and approved under the IACUC protocol 21049 and 20047.

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