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

Authors: Andras Sziraki[1,2,7], Ziyu Lu[1,2,7], Jasper Lee[1,7], Gabor Banyai[1], Sonya Anderson[3], Abdulraouf Abdulraouf[1,4], Eli Metzner[1,5], Andrew Liao[1,4], Alexander Epstein[1,2], Zihan Xu[1,2], Zehao Zhang[1,2], Li Gan[6], Peter T. Nelson[3], Wei Zhou[1,8]*, Junyue Cao[1,8,9]*

Affiliations:

[1]Laboratory of Single Cell Genomics and Population Dynamics, The Rockefeller University, New York, NY, USA

[2]The David Rockefeller Graduate Program in Bioscience, The Rockefeller University, New York, NY, USA

[3]Department of Pathology and Sanders-Brown Center on Aging, University of Kentucky, Lexington, KY

[4]The Tri-Institutional M.D-Ph.D Program, New York, NY, USA

[5]The Tri-Institutional Ph.D. Program in Computational Biology & Medicine, New York, NY, USA

[6]Helen & Robert Appel Alzheimer's Disease Research Institute, Weill Cornell Medicine, New York, NY, USA

[7]These authors contributed equally

[8]Senior author

[9]Lead Contact

*Correspondence: wzhou@rockefeller.edu (W.Z.), jcao@rockefeller.edu (J.C.)

Summary

Although several studies have applied single-cell approaches to explore gene expression changes in aged brains, they were limited by the relatively shallow sampling of brain cell populations, and thus may have failed to capture aspects of the molecular signatures and dynamics of rare cell types associated with aging and diseases. Here, we set out to investigate the age-dependent dynamics of transcription and chromatin accessibility across diverse brain cell types. With *EasySci*, an extensively improved single-cell combinatorial indexing strategy, we profiled ~1.5 million single-cell transcriptomes and ~400,000 single-cell chromatin accessibility profiles across mouse brains spanning different ages, genotypes, and both sexes. With a novel computational framework designed for characterizing cellular subtypes based on the expression of both genes and exons, we identified > 300 cell subtypes and deciphered the underlying molecular programs and spatial locations of rare cell types (*e.g.,* pinealocytes, tanycytes) and subtypes. Leveraging these data, we generate a global readout of age-dependent cell population dynamics with high cellular subtype resolution, providing insights into cell types that expand (e.g., rare astrocytes and vascular leptomeningeal cells in the olfactory bulb, reactive microglia and oligodendrocytes) or are depleted (*e.g.,* neuronal progenitors, neuroblasts, committed oligodendrocyte precursors) as age progresses. Furthermore, we explored cell-type-specific responses to genetic perturbations associated with Alzheimer's disease (AD) and identify rare cell types depleted (*e.g., mt-Cytb+, mt-Rnr2+* choroid plexus epithelial cells) or enriched (*e.g., Col25a1+*, *Ndrg1+* interbrain and midbrain neurons) in both AD models. Key findings are consistent between males and females, validated across the transcriptome, chromatin accessibility, and spatial analyses. Finally, we profiled a total of 118,240 single-nuclei transcriptomes from twenty-four human brain samples derived from control and AD patients, revealing highly cell-type-specific and region-specific gene expression changes associated with AD pathogenesis. Critical AD-associated gene signatures were validated in both human and mice. In summary, these data comprise a rich resource for exploring cell-type-specific dynamics and the underlying molecular mechanisms in both normal and pathological mammalian aging.

Introduction

The mammalian brain is a remarkably complex system made up of millions to billions of highly heterogeneous cells, comprising a myriad of different cell types and subtypes (Erö et al., 2018; Zeisel et al., 2018). Progressive changes in brain cell populations, which occur during the normal aging process, may contribute to functional decline of the entire organ and increased risks for neurodegenerative diseases such as Alzheimer's disease (AD) (Mathys et al., 2019; Xia et al., 2018). While the recent advances in single-cell genomics have created unprecedented opportunities to explore the cell-type-specific dynamics across the entire mammalian brain in aging and AD models (Morabito et al., 2021; Tabula Muris Consortium, 2020; Wang et al., 2022; Ximerakis et al., 2019), most prior studies relied on a relatively shallow sampling of the brain cell populations, possibly resulting in poor sensitivity to investigate the dynamics of cell types during aging, particularly with respect to rare aging or AD-associated cell types. While providing proof of key concepts, these prior studies were also technically limited in several ways, including failing to recover isoform-level gene expression patterns for rare cell types, providing few insights into how the chromatin landscape regulates cell-type-specific alterations across aging stages, and often lacked integrative analyses with spatial visualization to explore the anatomic region-specific changes.

We previously developed single-cell RNA sequencing by combinatorial indexing, a methodological framework involving split-pool barcoding of cells or nuclei for single-cell transcriptome profiling (Cao et al., 2017). While the method has been widely used to study embryonic and fetal tissues (Cao et al., 2019, 2020), it remains restricted to gene quantification proximal to the 3' end (*i.e.,* full-length transcript isoform information is lost) and is limited in terms of efficiency and cell recovery (up to 95% cell loss rate) (Cao et al., 2019), which pose a challenge when dealing with aged tissues. We have now performed over 350 optimization experiments to overcome the above limitations (representative examples are shown in **Figure S1 and S2; Methods**). Several test conditions were inspired by optimizations described in recently developed or optimized single-cell techniques (Ma et al., 2020; Martin et al., 2021). The major improvements of the resulting method, *EasySci-RNA* (**Figure 1A**), include: (i) *one million* single-cell transcriptomes prepared at a library preparation cost of around $700, less than 1/100 the cost of the commercial platforms (Ding et al., 2020) (**Figure 1B and 1C)**. Of note, this cost mainly includes the reagents cost for scRNA-seq library preparation and does not include the cost of personnel or sequencing; (ii) nuclei are deposited to different wells for reverse transcription with indexed oligo-dT and random hexamer primers (*i.e.,* different molecular barcodes to separate reads primed by two types of primers and across different wells), thus recovering cell-type-specific gene expression with full gene body coverage (**Figure 1D**); (iii) chemically modified oligos were included in the ligation reaction to prevent the formation of primer-dimers and increase the detection efficiency (**Figure S2**); (iv) cell recovery rate, as well as the number of transcripts detected per cell, were significantly improved through optimized nuclei storage and enzymatic reactions (**Figure S2**). The optimized technique yields significantly higher signals per nucleus compared with the original sci-RNA-seq3 protocols as well as the popular commercial platform (*i.e.,* 10x Genomics) (**Figure 1E; Figure S2N-O**); (v) An extensively improved single-cell data processing pipeline was developed for both gene counting and exonic counting utilizing paired-end single-cell RNA sequencing data (**Methods**).

Leveraging the technical innovations during the development of *EasySci-RNA*, we further optimized the recently published single-cell chromatin accessibility profiling method by combinatorial indexing (sci-ATAC-seq3) (Cusanovich et al., 2018; Domcke et al., 2020). Critical additional improvements include: (i)

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 reaction efficiency and signal specificity (**Figure S3**). The detailed protocols for the *EasySci* method (RNA and ATAC), as well as the data processing pipeline, are both included as supplementary files (**Supplementary file 1-6**) to facilitate the uptake of the techniques to further enable individual laboratories to cost-efficiently generate gene expression and chromatin accessibility profiles from millions of single cells.

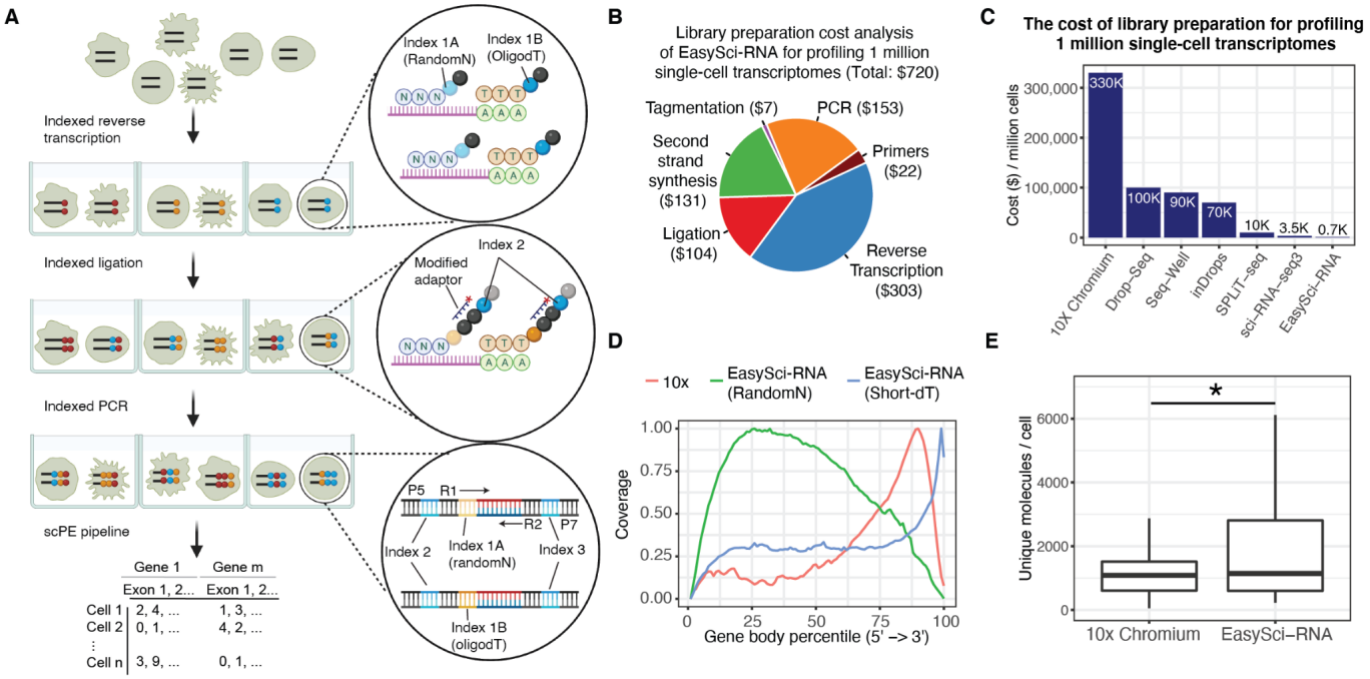


Figure 1. *EasySci-RNA* enables high-throughput and low-cost single-cell transcriptome profiling with full gene body coverage.

(A) *EasySci-RNA* workflow. Key steps are outlined in the texts.

(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 sci-RNA-seq3 and SPLiT-seq were calculated using data from (Martin et al., 2021; Rosenberg et al., 2018). The cost of other techniques was calculated using data from (Ding et al., 2020).

(D) Density plot showing the gene body coverage comparing single-cell transcriptome profiling using *10X genomics* and *EasySci-RNA*. Reads from indexed oligo-dT priming and random hexamers priming are plotted separately for *EasySci-RNA*.

(E) Box plot showing the number of unique transcripts detected per mouse brain nucleus comparing *10X genomics* and an *EasySci-RNA* library at similar sequencing depth (~ 4,450 raw reads/cell, Methods). For the box plot: middle lines, medians; upper and lower box edges, first and third quartiles, respectively; whiskers, 1.5 times the interquartile range. The star indicates p-value < 0.05 using a Wilcoxon rank-sum test.

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

We first applied the *EasySci* method to characterize cell-type-specific gene expression, and chromatin accessibility across the entire mouse brain sampling at different ages, sexes, and genotypes (**Figure 2A**). We collected C57BL/6 wild-type mouse brains at three months (n=4), six months (n=4), and twenty-one months (n=4). To gain insight into the early molecular changes associated with the pathophysiology of AD, we included two AD models from the same C57BL/6 background at three months. These include an early-onset AD (EOAD) model (5xFAD) that overexpresses mutant human amyloid-beta precursor protein (APP) and human presenilin 1 (PS1) harboring multiple AD-associated mutations (Oakley et al., 2006); 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 (Desimone et al., 2021; Xiang et al., 2018).

Nuclei were first extracted from the whole brain, then deposited to different wells for indexed reverse transcription (*EasySci-RNA*) or transposition (*EasySci-ATAC*), such that the first index identified the originating sample and assay type of any given well. The resulting *EasySci* libraries were sequenced in two Illumina NovaSeq runs, yielding a total of 20 billion reads (around 10 billion for each library). After filtering out low-quality cells and potential doublets, we recovered gene expression profiles in 1,469,111 single nuclei (a median of 70,589 nuclei per brain sample, **Figure S4A; Methods**) and chromatin accessibility profiles in 376,309 single nuclei (a median of 18,112 nuclei per brain sample, **Figure S4B; Methods**) across conditions. Despite shallow sequencing depth (~ 4500 and ~ 16,000 raw reads per cell for RNA and ATAC, respectively), we recovered an average of 1,788 UMIs (RNA, median of 935, 12.8% duplication rate) and 5,515 unique fragments (ATAC, median of 3,918, 9.3% duplication rate) per nucleus (**Figure S4C and S4D)**, comparable to the published datasets (Cao et al., 2019, 2020; Domcke et al., 2020). A median of 19% of ATAC-seq reads was mapped to locations near the transcription start site (±1 kb) (**Figure S4E)**, comparable to the published sci-ATAC-seq3 approach (Domcke et al., 2020).

With UMAP visualization (McInnes et al., 2018) and Louvain clustering (Blondel et al., 2008), we identified 31 main cell types by gene expression clusters (a median of 16,370 cells per cell type; **Figure 2C; Methods**), annotated based on cell-type-specific gene markers (Zeisel et al., 2018). Each cell type was observed in almost every individual, except the rare pituitary cells (0.09% of the cell population) that were missing in three out of twenty individuals (**Figure S5**). The cell-type-specific fractions in the global cell population are highly biased, ranging from 0.05% (Inferior olivary nucleus neurons) to 32.5% (Cerebellum granule neurons) (**Figure 2B**). An average of 74 marker genes were identified for each main cell type (defined as differentially expressed genes with at least a 2-fold difference between first and second-ranked cell types with respect to expression; FDR of 5%; and TPM > 50 in the target cell type; **Table S1**). In addition to the established marker genes, we identified many 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*) **(Figure S5B**).

We next sought to quantify isoform expression through a published computation pipeline (Booeshaghi et al., 2021). Briefly, we merged random hexamer reads from each cell type in every individual mouse brain, yielding 613 pseudocells. We then pseudo-aligned the merged reads to the mouse transcriptome using Kallisto (Bray et al., 2016), producing a raw isoform count matrix that was processed per the above pipeline. After filtering and normalization, we recovered abundance estimates for 33,361 isoforms

corresponding to 12,636 genes (**Methods**). As expected, the previously identified main clusters can be readily resolved through isoform expression (**Figure S6A**). Compared with single-cell RNA-seq libraries (Booeshaghi et al., 2021), a relatively lower fraction (~40%) of *EasySci-RNA* reads were mapped to transcriptome with the above pipeline, potentially due to the high fraction of intronic reads in single nucleus RNA sequencing. Nevertheless, we identified certain isoforms strongly expressed in a given cell type even though their corresponding genes are not cell-type-specific (**Table S2**). For example, *App-202*, an isoform of the amyloid precursor protein gene, is preferentially expressed in choroid plexus epithelial cells, while its corresponding gene is not (**Figure S6B**). Similarly, *Aplp2-209*, an isoform of the amyloid beta precursor-like protein 2 gene, is differentially expressed in oligodendrocytes, while its cell-type-specificity is not detected at the gene level (**Figure S6C**). The differential expression of *Aplp2-209* in oligodendrocytes is further validated using the Tabula Muris Senis mouse aging atlas dataset (Tabula Muris Consortium, 2020) (**Figure S6D**).

To reconstruct a brain cell atlas of both gene expression and chromatin accessibility, we applied a deep learning-based strategy (Lin et al., 2022) to integrate the 376,309 single-cell chromatin accessibility profiles with gene expression data (**Figure 2C; Methods**), yielding all 31 main cell types defined by chromatin accessibility. The gene body accessibility and expression of marker genes across cell types were highly correlated (**Figure 2D**). The fraction of each cell type was highly consistent between two molecular measurements as well (**Figure 2E**). To gain more insight into the epigenetic controls of the diverse cell types in the brain, we next identified peaks of accessibility within each cell type, yielding a master set of 339,951 peaks. There was a median of 34% of reads in peaks per nucleus. UMAP dimension reduction using the resulting peak count matrix readily separates main cell types, further validating the integration-based annotations **(Figure S7A**). Through differential accessibility (DA) analysis, we identified a median of 474 differential accessible peaks per cell type (FDR of 5%, TPM > 20 in the target cell type, **Figure S7B and S7C; Table S3**). Key cell-type-specific TF regulators were discovered by correlation analysis between motif accessibility and expression patterns across diverse cell types **(Figure S7D**), such as *Spi1* in microglia (Yeh and Ikezu, 2019), *Nr4a2* in cortical projection neurons 3 (Watakabe et al., 2007), and *Pou4f1* in inferior olivary nucleus neurons (McEvilly et al., 1996).

As a step toward a spatially resolved brain atlas, we integrated our dataset with a *10x Visium* spatial transcriptomics dataset (Genomics, 2019a, 2019b, 2019c) through a modified non-negative least squares (NNLS) approach (**Methods**). Aggregated cell-type-specific gene expression data were used as input to decompose mRNA counts at individual spatial locations of both sagittal and coronal sections of the entire mouse brain, thereby estimating the cell-type-specific abundance across locations. As expected, specific brain cell types were mapped to distinct anatomical locations (**Figure 2F and 2G**), especially for region-specific cell types such as cortical projection neurons (Clusters 6,7,8), cerebellum granule neurons (Cluster 3), and hippocampal dentate gyrus neurons (Cluster 9). The integration analysis further confirmed the annotations and spatial locations of main cell types in our single-cell datasets.

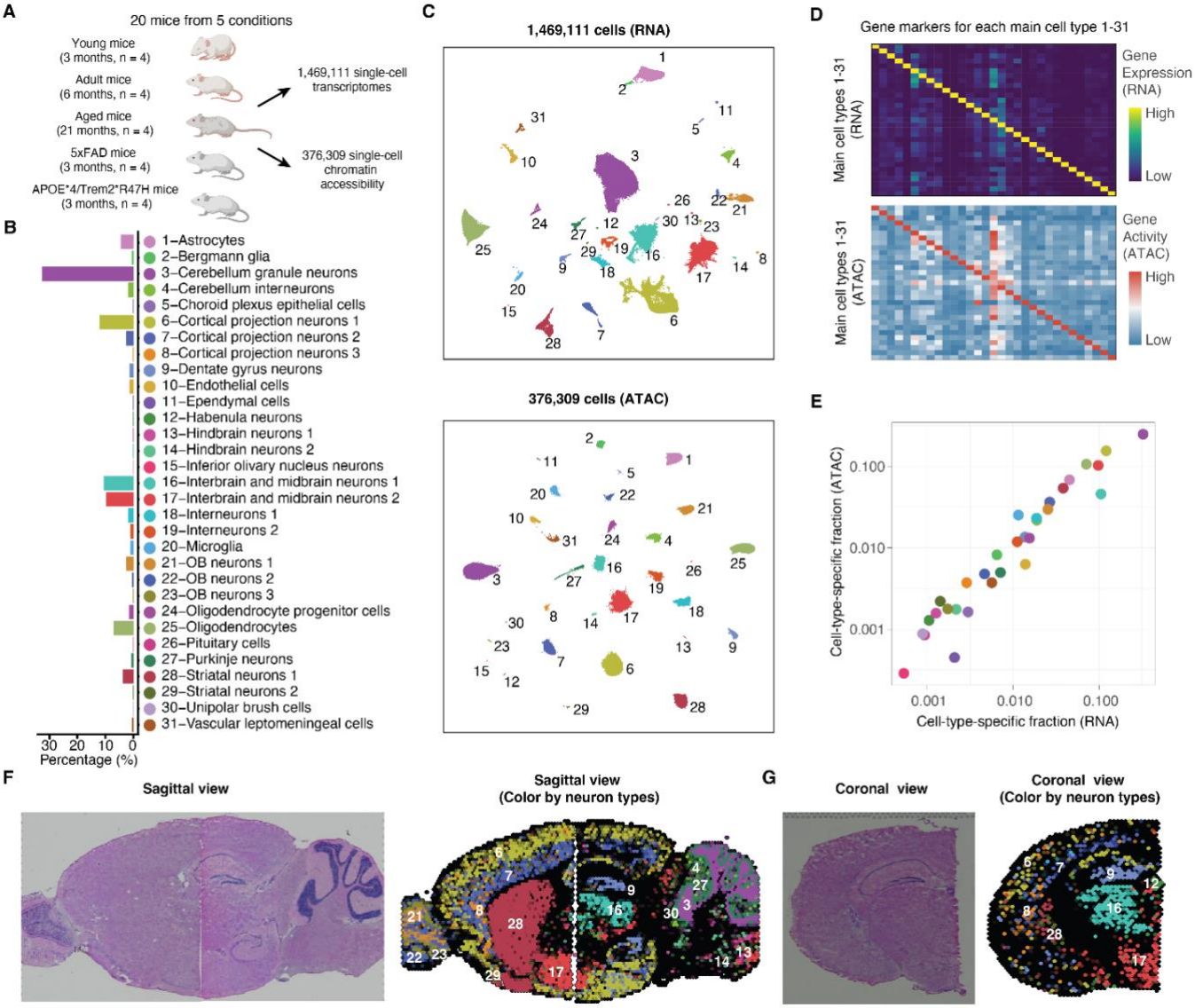


Figure 2. Single-cell transcriptome and chromatin accessibility profiling of mouse brains with *EasySci*.

(A) Experiment scheme to reconstruct a brain cell atlas of both gene expression and chromatin accessibility across different ages, sexes, and genotypes.

(B) Bar plot showing the cell-type-specific proportions of the brain cell population profiled by *EasySci-RNA*.

(C) UMAP visualization of mouse brain cells by single-cell transcriptome (top) and chromatin accessibility (bottom), colored by main cell types in (B).

(D) Heatmap showing the aggregated gene expression (top) and gene body accessibility (bottom) of the top ten marker genes (columns) in each main cell type (rows). For both RNA-seq and ATAC-seq, unique reads overlapping with the gene bodies of cell-type-specific markers were aggregated, normalized first by library size, and then scaled by the maximum expression or accessibility across all cell types.

(E) 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).

(F-G) Mouse brain sagittal (F) and coronal (G) sections showing the H&E staining (left) and the localizations of main neuron types through NNLS-based integration (right), colored by main cell types in (B). The numbers correspond to cell-type-specific cluster-ID in (B).

A computational framework tailored to characterize cellular subtypes in the mammalian brain

To investigate the molecular signatures and spatial distributions of cellular subtypes in the brain, we developed a computational framework tailored to sub-cluster level analysis. Key steps include: (i) sub-clustering analysis by expression of both genes and exons to increase the clustering resolution (**Figure 3**); (ii) Integration analysis with spatial datasets to map the distribution of cellular subtypes (**Figure 3**). (iii) gene module analysis to identify the molecular signatures of main and rare cell types (**Figure 4**);

Rather than performing sub-clustering analysis with the gene expression alone, we exploited the unique feature of *EasySci-RNA* (*i.e.,* full gene body coverage), by incorporating both gene counts and exonic counts for principal component analysis followed by unsupervised clustering (**Methods**). The combined information greatly increased the resolution of sub-clustering analysis. For example, we recovered several microglia subtypes that were not easily separated in clusters defined by gene expression alone (*e.g.,* sub-cluster 13 in microglia marked by an exonic marker *Ttr-ENSMUSE00000477272.5*, **Figure 3A and 3B**). Leveraging this sub-clustering strategy, we identified a total of 359 sub-clusters, with a median of 1,038 cells in each group (**Figure 3C**). All sub-clusters were contributed by at least two individuals (median of twenty), with a median of nine exonic markers enriched in each sub-cluster (At least a 2-fold difference between first and second-ranked cell types with respect to expression; FDR of 5%; and TPM > 50 in the target sub-cluster, **Figure S9**; **Table S4**). Some subtype-specific exonic markers were not detected by conventional differential gene analysis (*e.g., Map2-ENSMUSE00000443205.3* in microglia, **Figure S8)**. Notably, our strategy favors detecting extremely low-abundance cell types. For example, the smallest sub-cluster (choroid plexus epithelial cells-7) contained only 21 cells (0.001% of the brain population, **Figure 3D-E, top**), representing rare pinealocytes in the brain based on gene markers such as *Tph1* and *Ddc* (Mays et al., 2018). Another example of the rare sub-clusters (vascular leptomeningeal cells-2, 35 cells, **Figure 3D-E, bottom**) represents the tanycytes, validated by multiple gene markers (*e.g., Fndc3c1*, *Scn7a* (Campbell et al., 2017)).

To spatially map the diverse sub-clusters, we integrated the *EasySci-RNA* dataset with the *10x Visium* spatial transcriptomics datasets (Genomics, 2019a, 2019b, 2019c) using cell2location (Kleshchevnikov et al., 2022), a Bayesian model designed to map fine-grained cell types. For example, sub-clusters of cortical projection neuron 1 can be mapped to distinct regions in the cortex by integration analysis (**Figure 3F**). A similar approach enabled us to deconvolute the region-specificity for other broadly distributed cell types such as astrocytes (**Figure 3G**). For instance, we identified astrocyte subtypes that specifically mapped to the olfactory bulb (sub-cluster 7), cortex (sub-cluster 3, 12), hippocampus (sub-cluster 6), thalamus (sub-cluster 11), midbrain (sub-cluster 4), hindbrain (sub-cluster 10), consistent with the known region-specificity of astrocytes (Kleshchevnikov et al., 2022).

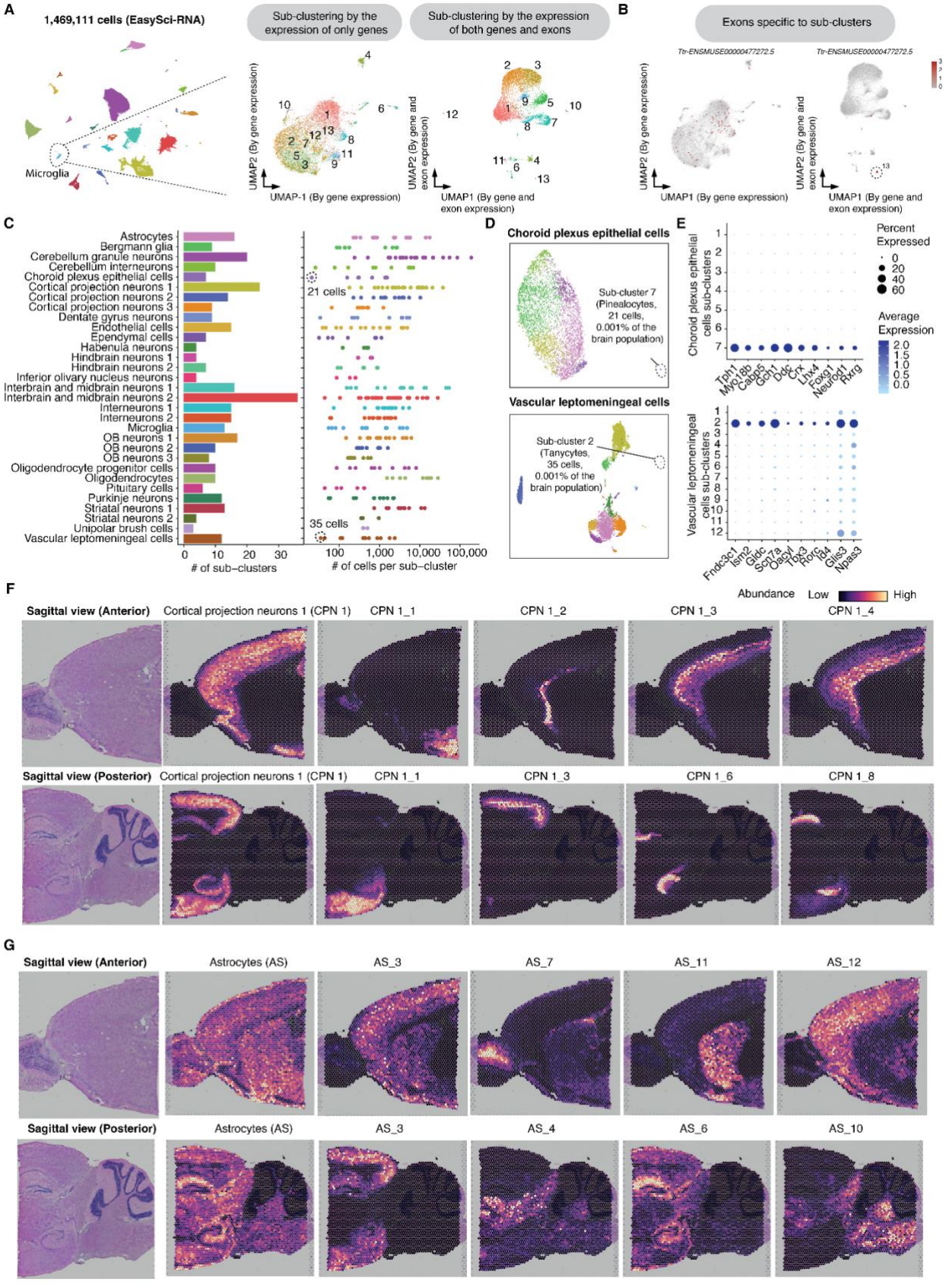


Figure 3. Identification and characterization of cellular subtypes in the mouse brain.

(A) Schematic plot showing the computational framework for identifying and characterizing cell sub-clusters. We subjected each main cell type to sub-clustering analysis based on both gene and exon expression. As an example, we performed UMAP analysis of microglia cells based on gene expression alone (left), or both gene and exon level expression (right). Cells are colored by sub-cluster ID from Louvain clustering analysis with combined gene and exon level information. Several sub-clusters cannot be separated from each other in the UMAP space by gene expression alone.

(B) UMAP plots same as (A), showing the expression of an exonic marker *Ttr-ENSMUSE00000477272.5* of microglia sub-cluster 13. Microglia-13 can be better separated when combining both gene and exon level information.

(C) By sub-clustering analysis, we identified a total of 359 sub-clusters across 31 main cell types. The barplot (left) shows the number of sub-clusters for each main cell type. The dot plot (right) shows the number of cells from each sub-cluster. Two rare sub-clusters (choroid plexus epithelial cells-7 and vascular leptomeningeal cells-2) are circled out.

(D) UMAP visualizations showing sub-clustering analysis for choroid plexus epithelial cells (top) and vascular leptomeningeal cells (bottom) colored by sub-cluster IDs, highlighting two rare sub-clusters shown in (C).

(E) Dot plot showing the expression of selected marker genes for choroid plexus epithelial cells-7 (top) and vascular leptomeningeal cells-2 (bottom), including both normal genes (left five genes) and transcription factors (right five genes).

(F-G) Mouse brain sagittal sections showing spatial abundances of main cell types and related sub-clusters for cortical projection neurons 1 (F) and astrocytes (G) in anterior (top) and posterior (bottom) regions, estimated using the cell2location (Kleshchevnikov et al., 2022).

Gene module analysis to determine cell-type-specific molecular programs

We next examined the key molecular programs underlying diverse cellular subtypes by gene module analysis. We clustered genes based on their expression variance across all 359 cell sub-clusters, revealing a total of 21 gene modules (GM) (**Figure 4A; Figure S10; Table S5**). The largest gene module (GM1) corresponds to a group of housekeeping genes (*e,g.,* ribosomal synthesis) universally expressed across all sub-clusters. Several gene modules were enriched in specific cell types, such as the ependymal cell-specific gene module (GM11, enriched biological process: cilium movement, adjusted p-value = 1.2e-26) (Kuleshov et al., 2016) (**Figure 4B**). Meanwhile, we detected gene modules that marked rare subtypes. For example, GM9, including genes in neuropeptide signaling (*e,g., Tbx19, Pomc* (Liu et al., 2001)), was highly enriched in a subtype of pituitary cells (PC-6) corresponding to corticotropic cells (**Figure 4B**). A similar analysis enabled us to characterize other rare cell subtypes, including myeloid cells (microglia-13, 0.005% of the cell population, marked by GM19), pars tuberalis cells (vascular leptomeningeal cells-12 (VLC-12), 0.003% of the cell population, marked by GM20), as well as aforementioned pinealocytes (choroid plexus epithelial cells-7 (CPEC-7), 0.001% of the cell population, marked by GM2) (**Figure S10**).

Remarkably, rare proliferating cell types were identified through a cell-cycle-related gene module (GM6, enriched biological process: microtubule cytoskeleton organization involved in mitosis, adjusted p-value = 1.2e-44) (Kuleshov et al., 2016), including proliferating cells of neurons (OB neurons 1-17, 0.03% of the cell population), astrocytes (astrocytes-7, 0.15% of the cell population), oligodendrocytes progenitor cells (oligodendrocytes progenitor cells-4, 0.04% of the cell population) and microglia (microglia-10, 82 cells, 0.006% of the cell population) (**Figure 4B**). These sub-clusters were marked by conventional proliferating markers such as *Mki67*, as well as a group of lncRNAs (*e.g., Gm29260, Gm37065*), most of which were not well-characterized in previous studies (**Figure 4C**).

To spatially map the rare cell types, we next investigated the expression patterns of cell-type-specific gene modules across spatial spots of the *10x Visium* spatial transcriptomic datasets (Genomics, 2019a, 2019b, 2019c), which resolved the anatomical locations of multiple main and rare cell types. For example, ependymal cells, a critical cell type regulating cerebrospinal fluid (CSF) homeostasis, were mapped along brain ventricles as expected (**Figure 4D**). Furthermore, rare proliferating cells were mapped to the subventricular zone area (**Figure 4E**). A similar analysis enables us to spatially map other rare cell types with high resolution, including pinealocytes (CPEC-7, GM2), corticotropic cells (PC-6, GM9), pars tuberalis cells (VLC-12, GM20), tanycytes (VLC-2, GM14) and a less-characterized endothelial cell type in the pituitary gland (*Igfbp3*+ *Sfn*+ endothelial cells, EC-10, GM7) (**Figure 4F**).

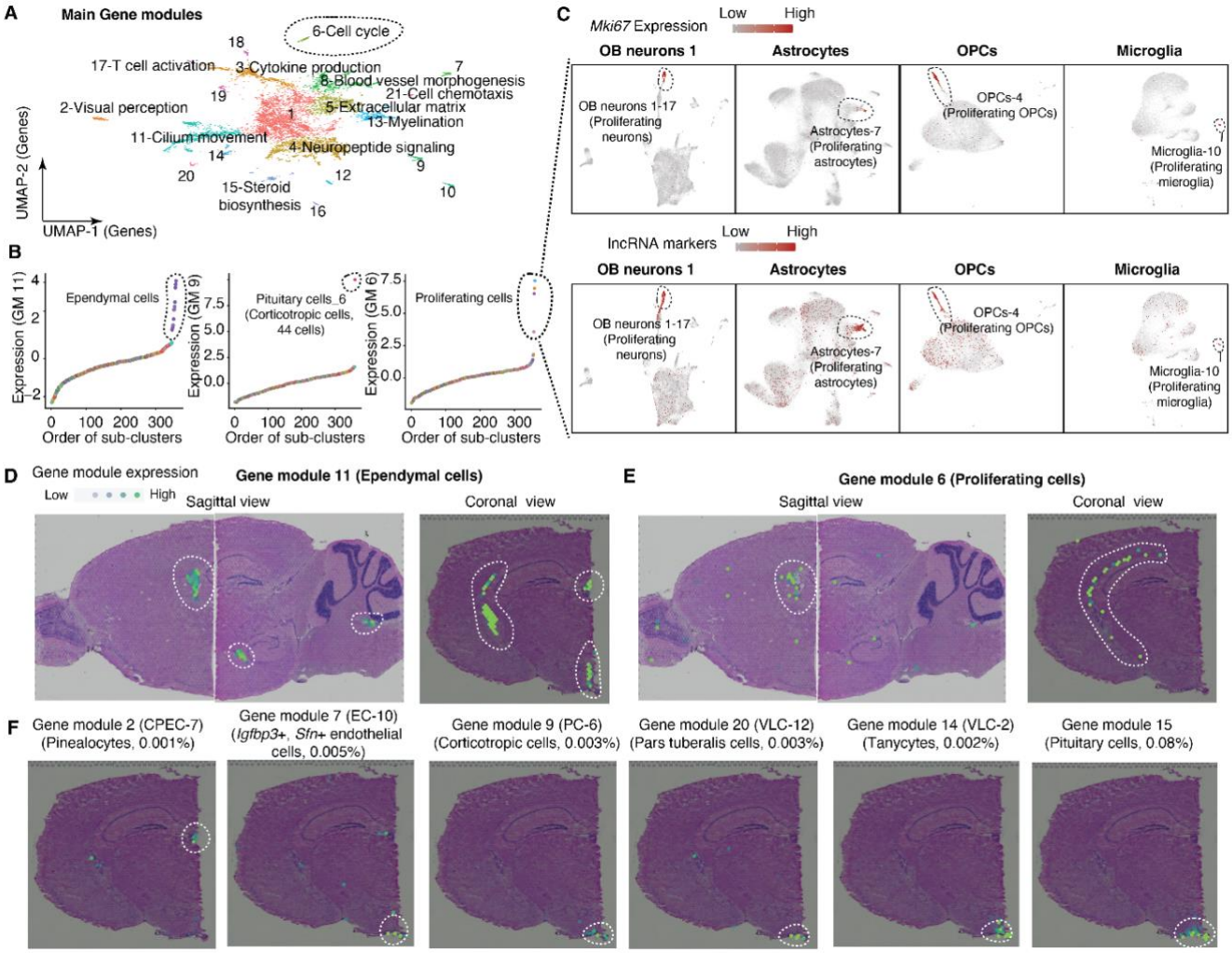


Figure 4. Identification of key molecular programs underlying cell type specificity in the mouse brain.

(A) UMAP visualizations of genes colored by identified gene module IDs.

(B) Scatter plots showing examples of gene modules and their expression levels across sub-clusters (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 sub-clusters from different main cell types.

(C) UMAP visualization showing four proliferating sub-clusters identified from OB neurons 1, astrocytes, 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, aggregated (for multiple genes), and then mapped to Z-scores. OPCs, oligodendrocyte progenitor cells.

(D-E) Plots showing the normalized expression of gene modules in spatial transcriptomic datasets profiling mouse sagittal (left) and coronal (right) sections: GM-11 is specific to ependymal cells (D); GM-6 is specific to proliferating cells (E).

(F) Similar to (D), plots showing the normalized expression of gene modules in spatial transcriptomic dataset profiling a mouse coronal section. UMI counts for genes from each gene module are scaled for

library size, log-transformed, aggregated, and then mapped to Z-scores. CPEC, choroid plexus epithelial cells; EC, endothelial cells; PC, pituitary cells; VLC, vascular leptomeningeal cells.

A global view of brain cell population dynamics across the adult lifespan 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 cell populations in each individual mouse. We next performed differential abundance analysis across all 359 sub-clusters (**Methods**), yielding 45 significantly changed sub-clusters during the early growth stage (between 3 and 6 months) and 29 significantly changed sub-clusters upon aging (between 6 and 21 months; FDR of 0.05, at least two-fold change of cellular fractions, **Figure 5A; Table S6 and S7**). Most significantly changed cell types were consistent between male and female mice (**Figure 5B**).

Both the main clusters and subtypes of olfactory bulb (OB) neurons showed a remarkable population expansion from young to adult mice (**Figure 5A, left),** consistent with the expansion of the OB region during the early growth stage (Tufo et al., 2022). Meanwhile, a rare astrocytes subtype (AS-14, *Lyn*+ *Adgrb1*+; 0.05% of the global population) and a vascular leptomeningeal cell subtype (VLC-14, *Sox10*+ *Mybpc1*+; 0.06% of the global population) also showed substantial expansion (over 4-fold) in the same period. We further characterized the chromatin accessibility of these two rare cell types, along with many OB neuron subtypes, by single-cell RNA-seq and ATAC-seq integration analysis through the deep-learning-based strategy (Lin et al., 2022) described above (**Figure S11A-C; Methods**). As expected, the observed cell population dynamics can be cross-validated by two molecular layers (*i.e.,* RNA and ATAC) (**Figure S11D**). Furthermore, both cell subtypes were spatially mapped to the OB region based on the expression of cell-type-specific gene markers in *10x Visium* spatial transcriptomic data (Genomics, 2019a, 2019b, 2019c) (**Figure 5C, left**), suggesting their potential roles in OB expansion during the early growth stage. As a further illustration of this point, the astrocytes subtype 14 is featured with the high expression of *BAI1*, a gene marker involved in the clean-up of apoptotic neuronal debris produced during fast growth of the brain (Sokolowski et al., 2011). The vascular leptomeningeal cell subtype 4 highly expresses gene markers of olfactory ensheathing cells (*e.g., Sox10* and *Mybpc1* (Rosenberg et al., 2018; Tepe et al., 2018)), a key cell type that supports the growth and regeneration of axons in the central nervous system (Barraud et al., 2010).

The aging-associated cell population changes between 6 and 21 months differed remarkably from the early growth stage. For example, all main cell types of OB neurons remain relatively stable during aging. Instead, we found aging-associated changes mostly in specific neuron subtypes. Key examples include the expansion of an OB neurons 3 subtype (OBN 3-3, marked by *Cpa6* and *Col23a1*) corresponding to a group of less-characterized excitatory neurons in the mitral cell layer of the OB region (Monavarfeshani et al., 2017), and the depletion of another OB neurons 1 subtype (OBN 1-11, marked by *Robo2* and *Prokr2*) corresponding to the OB neuroblasts (Puverel et al., 2009; Zeisel et al., 2018)). These subtypes were spatially mapped to different areas of the olfactory bulb (**Figure 5C, right**), which is in contrast with the early growth stage, where almost all subtypes of OB neurons expanded across all regions.

We identified a total of 21 subtypes showing a marked reduction across the adult lifespan of the mouse brain. For example, the most depleted populations in the aged brain include OB neuroblasts (OBN 1-11, marked by *Prokr2* and *Robo2* (Puverel et al., 2009; Zeisel et al., 2018)), OB neuronal progenitor cells (OBN 1-17, marked by *Mki67* and *Egfr* (Pastrana et al., 2009)), and dentate gyrus (DG) neuroblasts (DGN-8, marked by *Sema3c* and *Igfbpl1* (Kumar et al., 2020))(**Figure 5D, left**). Interestingly, the population of DG neuroblasts showed a substantial decrease even in the early growth stage, suggesting an earlier decline

of DG neurogenesis compared to OB neurogenesis. In contrast to the depleted pool of neurogenesis progenitors, the proliferating oligodendrocyte progenitor cells (cycling OPCs, OPC-4, marked by *Pdgfra* and *Mki67*) remain relatively stable during aging. Instead, we detected the aging-associated depletion of the newly formed oligodendrocytes (oligodendrocytes-6 (OLG-6), marked by *Prom1* and *Tcf7l1* (Marques et al., 2018; Pastrana et al., 2009)) and committed oligodendrocyte precursors (OPC-6, marked by *Bmp4* and *Enpp6* ((Marques et al., 2018; Pastrana et al., 2009; Zhang et al., 2014)), indicating that the oligodendrocyte differentiation is impaired upon aging.

While the aforementioned integrative approach successfully identified the chromatin landscape of all main cell types, there were several substantial challenges for the sub-clustering level analysis, including the relatively lower number of profiled cells and lower resolution of the single-cell chromatin accessibility dataset compared with the single-cell transcriptome analysis. Nevertheless, we were able to recover several cell subtypes with relatively high abundance and unique epigenetic signatures. For example, we identified OB neuroblasts (OBN 1-11), OB neuronal progenitors (OBN 1-17), and newly formed oligodendrocytes (OLG-6) (**Figure S12A and S12B**), all of which exhibited a sharply decreased cell proportions in the aged brain similar to the single-cell transcriptome analysis (**Figure 5D, right**). Moreover, cell-type-specific TF regulators were identified and validated by both gene expression and TF motif accessibility, such as known regulators of neurogenesis (*e.g., Sox2* and *E2f2* (Graham et al., 2003; Li et al., 2018)) (**Figure 5E**), which further validated this integration approach for characterizing key epigenetic signatures of cellular subtypes associated with aging.

We identified a total of 14 cellular sub-clusters that exhibited a remarkable expansion in the aged brain, such as a microglia sub-cluster (MG-9, *Apoe*+, *Csf1*+) corresponding to a previously reported disease-associated microglia subtype (Keren-Shaul et al., 2017). In addition, we identified a reactive oligodendrocyte subtype (OLG-7, *C4b*+, *Serpina3n*+ (Kenigsbuch et al., 2022; Zhou et al., 2020)) significantly enriched in the aged brain. With the chromatin accessibility dataset, we further confirmed the expansion of this cell type (**Figure 5F; Figure S12B and S12C**), and identified its associated transcription factors (**Figure 5E**), such as *Stat3*, a critical factor involved in the regulation of inflammation and immunity in the brain (See et al., 2012). To further characterize the spatial distribution of the reactive oligodendrocytes in the brain, we performed spatial transcriptomics analysis of both adult and aged mouse brains. A striking enrichment of the reactive oligodendrocyte-specific markers (*e.g., C4b*, *Serpina3n)* was detected around the subventricular zone (SVZ), a region critical for the continual production of new neurons in adulthood (**Figure 5G and 5H**), indicating an age-related activation of inflammation signaling around the adult neurogenesis niche.

We next explored the subtype-specific manifestation of key aging-related molecular signatures. Through differentially expressed gene analysis, we identified 7,135 aging-associated signatures across 359 sub-clusters (FDR of 5%, with at least a 2-fold change between aged and adult brains, **Table S8; Figure S13A**). Out of the 580 genes significantly altered in multiple (>= 3) subtypes, we detected 241 genes that were differentially expressed in concordant directions across subtypes (**Figure S13B**). For example, *Nr4a3*, a component of DNA repair machinery and a potential anti-aging target (Paillasse and de Medina, 2015), was significantly decreased in aged neurons, including striatal neurons, OB neurons, and interneurons. *Hdac4,* encoding a histone deacetylase and a recognized regulator of cellular senescence (Di Giorgio et al., 2021), was significantly reduced in aged astrocytes and ependymal cells. Meanwhile, the Insulin-degrading enzyme (IDE), a key factor involved in amyloid-beta clearance (Zhang and Wang, 2018),

showed increased expression mostly in subtypes of neurons, including interneurons, OB neurons, interbrain, and midbrain neurons. While many of these genes have been previously reported to be associated with aging, our analysis represents the first global view of their alterations across over 300 subtypes. In addition, we identified several non-coding RNAs that underwent age-associated changes in multiple cell subtypes, most of which showed high cell-type-specificity (e.g., *B230209E15Rik* in cortical projection neurons subtypes) but were not well-characterized previously (**Figure S13B**).

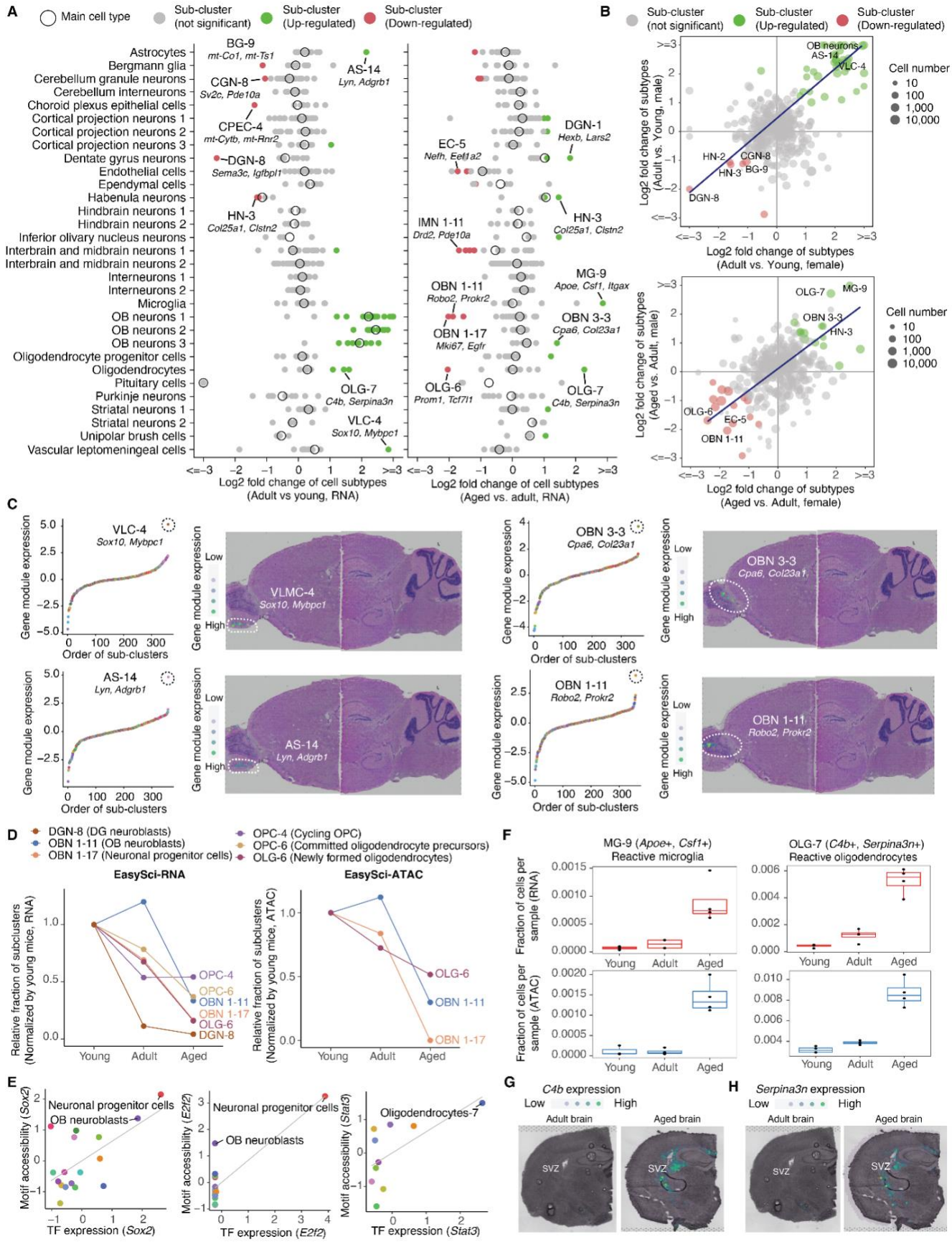


Figure 5. Identifying brain cell population changes across the adult lifespan at subtype resolution.
(A) Dot plots showing the cell-type-specific fraction changes (*i.e.,* log-transformed fold change) of main cell types (circles) and sub-clusters (dots) in the early growth stage (adult vs. young, left plot) and the aging process (aged vs. adult, right plot) from *EasySci*-RNA data. Differential abundant sub-clusters were colored by the direction of changes. Representative sub-clusters 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; VLC; vascular leptomeningeal cells.
(B) Scatter plots showing the correlation of the sub-cluster 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 sub-clusters are annotated on the plots.
(C) Examples of development- or aging-associated subclusters are highlighted in (a) and their spatial positions. Left: scatter plots showing the aggregated expression of sub-cluster-specific marker genes across all sub-clusters. Right: plots showing the aggregated expression of sub-cluster-specific marker genes across a brain sagittal section in 10x Visium spatial transcriptomics data(Genomics, 2019a, 2019b, 2019c). UMI counts for gene markers are scaled for library size, log-transformed, aggregated, and then mapped to Z-scores.
(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 transcription factors 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 profiled by *EasySci-RNA* (top) and *EasySci-ATAC* (bottom). For all box plots: middle lines, medians; upper and lower box edges, first and third quartiles, respectively; whiskers, 1.5 times the interquartile range; and all individual data points are shown.
(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.

A global view of AD pathogenesis-associated signatures and subtypes

Toward a global view of AD-associated cell population dynamics, we quantified the relative fraction of sub-clusters in the two AD models for comparison with their age-matched wild-type controls (3-month-old). We detected 16 and 14 significantly changed sub-clusters (FDR of 5%, at least two-fold change) in the EOAD (5xFAD) model and LOAD (APOE*4/Trem2*R47H) model, respectively (**Figure 6A, Table S9 and S10**). Most significantly altered subtypes showed consistent proportion changes in male and female mice (**Figure 6B**).

Interestingly, while these two AD mutants involved different genetic perturbations, the significantly altered cell subtypes were highly concordant (**Figure 6C**). For example, a rare choroid plexus epithelial cell subtype (CPEC-4, 0.02% of the total brain cell population) was strongly depleted (> two-fold decrease) in both AD models. This cell type is marked by significant enrichment of multiple mitochondrial genes, including *mt-Rnr1*, *mt-Rnr2*, *mt-Co1*, *mt-Cytb*, *mt-Nd1*, *mt-Nd2*, *mt-Nd5,* and *mt-Nd6*. Out of these gene markers, *mt-Rnr2* is involved in synthesizing neuroprotective factors against neurodegeneration by suppressing apoptotic cell death (Hashimoto et al., 2001). Other markers (*e,g., mt-Rnr1* and *mt-Nd5*) are associated with the phosphorylated Tau protein levels in cerebrospinal fluid (Cavalcante et al., 2022). While this subtype was rarely detected in our single-cell ATAC data, we were able to map the cell subtype to the area around the subventricular zone by the expression of its cell-type-specific markers in the spatial transcriptomics data (**Figure 6D and 6E, top**). Furthermore, the spatial transcriptomics analysis further validated the depletion of this cell type in the EOAD (5xFAD) model (**Figure 6E**), suggesting a potential interplay between cell-type-specific mitochondrial functions and neurodegenerative phenotypes.

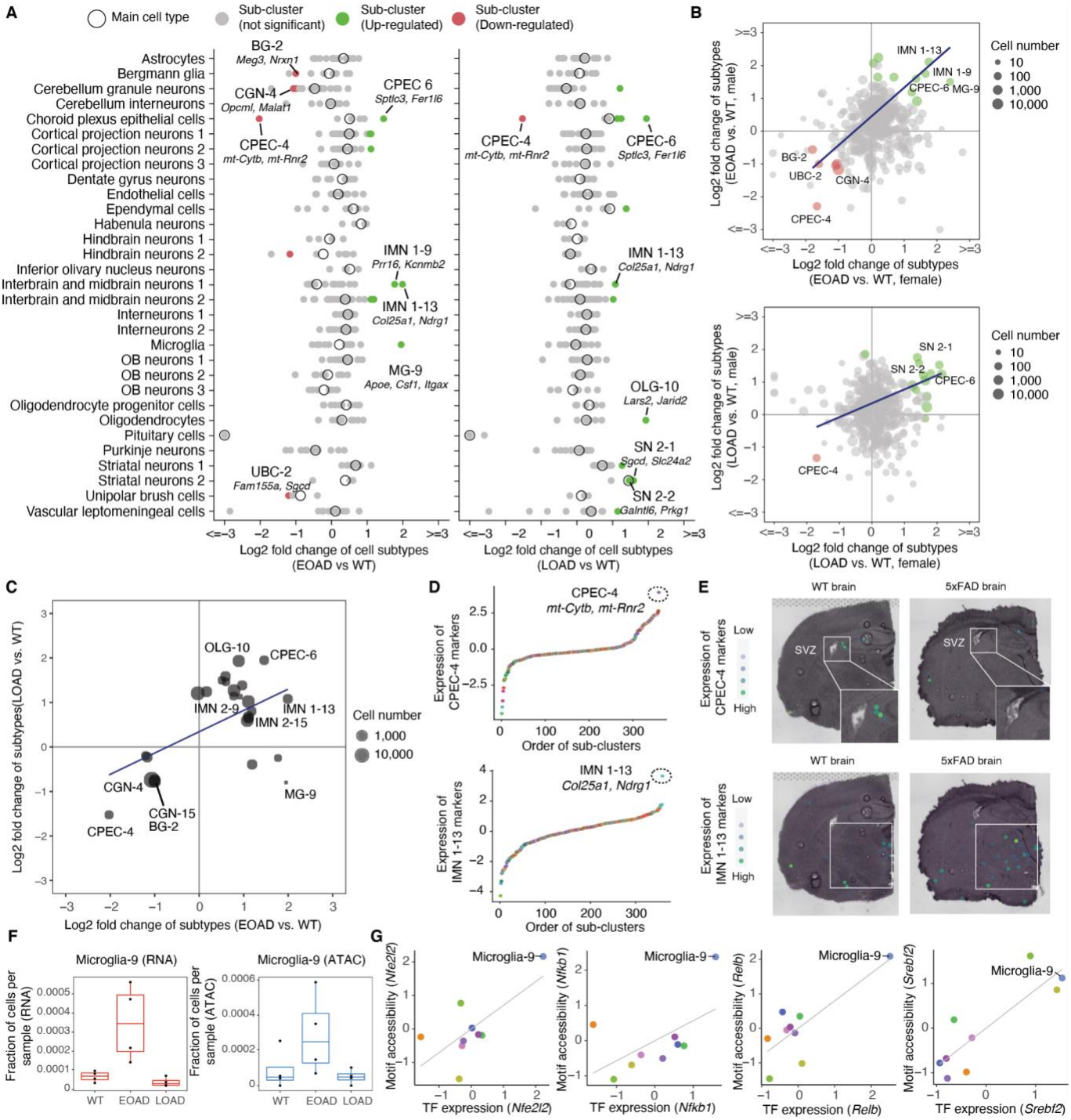
By contrast, another choroid plexus epithelial cell subtype (CPEC-6, 0.045% of the total brain cell population; marked by *Sptlc3*+, *Fer1l6*+) expanded in both AD models (over two-fold increase) (**Figure 6B**). It is marked by the gene *Sptlc3,* which encodes a subunit of a complex that catalyzes the synthesis of sphingolipids, a group of bioactive molecules contributing to amyloid-beta production and Alzheimer pathogenesis (Mielke and Lyketsos, 2010). Furthermore, we identified another rare interbrain and midbrain neuron subtype (IMN 1-13, 0.61% of the total brain population; marked by *Col25a1+, Ndrg1*+) that expanded considerably in both AD models (**Figure 6C**). This subtype is characterized by the expression of *Col25a1*, a membrane-associated collagen that has been reported to promote intracellular amyloid plaque formation in mouse models (Tong et al., 2010). Indeed, we identified an up-regulation of IMN 1-13 specific gene markers in the thalamus region of the 5xFAD mouse brain (**Figure 6D and 6E, bottom**), further validating the single-cell transcriptome analysis.

Finally, we detected a significant expansion of the microglia subtype 9 (MG-9, 0.026% of the total brain cell population, marked by *ApoE+, Csf1*+) in the early-onset 5xFAD mice, consistent with previous reports (Keren-Shaul et al., 2017). The reactive microglia subtype was also expanded in the aged group, but not in the late-onset APOE*4/Trem2*R47H model (3-month-old) (**Figure 6F, left**), potentially due to the different disease onset. Consistent proportion changes were detected with the chromatin accessibility dataset (**Figure 6F, right**). We next integrated both transcriptome and chromatin accessibility profiles to further delineate the molecular programs of this reactive microglia subtype., We first identified 199 genes differentially expressed in the reactive microglia subtype, many of which (44%) can be validated by the promoter accessibility (**Figure S12D**). We then revealed key transcription factors validated by both cell-type-specific gene expression and motif accessibility (**Figure 6G**), including TFs of the NF-kappa B

signaling pathway (e.g., *Nfkb1* and *Relb* (Oeckinghaus and Ghosh, 2009)) and TFs involved in oxidative stress protection (*e.g., Nfe2l2* (Liu et al., 2017)), and cholesterol homeostasis (*e.g., Srebf2* (Bommer and MacDougald, 2011)). These molecular pathways could play critical regulatory roles in microglia specification and expansion in aging and AD.

We next explored AD pathogenesis-associated signatures in AD mouse models. Through differentially expressed gene analysis (**Methods**), 6,792 and 7,192 sub-cluster-specific DE genes were detected in the 5xFAD (EOAD) model and the APOE*4/Trem2*R47H (LOAD) model, respectively (**Figure S13C and S13E, Table S11 and S12**). For example, we observed a global down-regulation of the mouse *Apoe* gene across many sub-clusters in the APOE*4/Trem2*R47H mice (**Figure S13D)**, potentially due to the fact that part of the *Apoe* gene is replaced with the human sequence that does not align to the mouse genome. Meanwhile, we detected a global change of *Thy1* across many neuron types in the 5xFAD mice, consistent with the fact that all transgenes introduced in the 5xFAD model were overexpressed under the *Thy1* promoter (**Figure S13F)**. Remarkably, many AD-associated gene signatures exhibited concordant changes across cellular subtypes (**Figure S13D and S13F**). For example, markers involved in unfolded protein stress (*e.g., Hsp90aa1*) and oxidative stress (*e.g., Txnrd1*) were significantly upregulated in an overlapped set of neuron subtypes in the early-onset 5xFAD mice (**Figure S13D**), indicating increased stress levels and cellular damages in neurons across the brain. Meanwhile, *Reln*, which encodes a large secreted extracellular matrix protease involved in the ApoE biochemical pathway (Seripa et al., 2008), significantly decreased in multiple cell types (*e.g.,* OB neurons, interbrain and midbrain neurons, vascular cells, oligodendrocytes) in both early- and late-onset models (**Figure S13D and S13F**). This is consistent with previous reports that the depletion of *Reln* is detectable even before the onset of amyloid-beta pathology in the human frontal cortex (Herring et al., 2012). Other interesting phenomena included the overall upregulation of *Ide*, a gene responsible for amyloid-beta degradation, in the late-onset model similar to the aged brain (**Figure S13B and S13F**). Less-characterized genes were identified as well. For example, *Tlcd4*, a gene involved in lipid trafficking and metabolism (Attwood and Schiöth, 2021), was significantly downregulated in thirty-five sub-clusters across broad cell types (*e.g.,* OB neurons, vascular cells, oligodendrocytes) in the EOAD mice (**Figure S13D**), indicating a potential interplay between the lipid homeostasis and cellular changes in the early stage of AD.

While the two AD mouse models are different in terms of genetic perturbations or disease onsets, their cell-type-specific molecular changes were surprisingly consistent. Illustrative of this, we detected 559 sub-cluster-specific DE genes shared between two AD mutants, such as genes involved in epilepsy (Adjusted p-value = 0.02, *e.g., Gria1, Med1, Plp1*) (Kuleshov et al., 2016) and oxidative stress protection pathway (Adjusted p-value = 0.05, *e.g., Arnt, Nfe2l2*) (Kuleshov et al., 2016). Intriguingly, 99% (555 of the 559) of the shared DE genes showed concordant changes in two AD mutants (Pearson correlation coefficient r = 0.96, p-value < 2.2e-16, **Figure S13G**), indicating shared molecular programs between early- and late-onset AD models. Of note, this analysis further validates that the APOE*4/Trem2*R47H mice mutant, a mouse model recently developed, can serve as an informative model to study Alzheimer's disease.



554 **Figure 6. Identifying AD pathogenesis-associated cell subtypes.**
555 (A) Dot plots showing the log-transformed fold changes of main cell types (circles) and sub-clusters (dots)
556 comparing EOAD vs. WT (left) and LOAD vs. WT (right). Differential abundant sub-clusters were colored
557 by the direction of changes. Representative sub-clusters were labeled along with top gene markers. BG,
558 Bergmann glia; CPEC, choroid plexus epithelial cells; IMN 1, interbrain and midbrain neurons 1; MG,
559 microglia; OLG, oligodendrocytes; SN 2, striatal neurons 2.

(B) Scatter plots showing the correlation of the log-transformed fold changes of sub-clusters (top: EOAD vs. WT, bottom: LOAD vs. WT) between male and female.
(C) Scatter plot showing the correlation of the log-transformed fold changes of sub-clusters in two AD models (both compared with the wild-type). Only sub-clusters showing significant changes in at least one AD model are included.
(D) Scatter plots showing the aggregated expression of gene markers of two cell subtypes (top: choroid plexus epithelial cells-4; bottom: the interbrain and midbrain neurons 1-13) across all sub-clusters from *EasySci-RNA* data.
(E) Brain coronal sections showing the spatial expression of subtype-specific gene markers of two subtypes (top: choroid plexus epithelial cells-4; bottom: the interbrain and midbrain neurons 1-13) in the WT and EOAD (5xFAD) brains in 10x Visium spatial transcriptomics data.
(F) Box plots showing the fraction of microglia-9 cells across different conditions profiled by *EasySci-RNA* (left) or *EasySci-ATAC* (right). For all box plots in this figure: middle lines, medians; upper and lower box edges, first and third quartiles, respectively; whiskers, 1.5 times the interquartile range; and all individual data points are shown.
(G) Scatter plot showing the correlated gene expression and motif accessibility of four transcription factors (*Nfe2l2*, *Nfkb1*, *Relb,* and *Srebf2*) enriched in microglia-9, together with a linear regression line.

Identification of dysregulated gene signatures in human AD brains

To examine the AD pathogenesis-associated gene signatures, we further sequenced a total of 118,240 single-nuclei transcriptomes (a median of 5,585 nuclei per sample, with a median of 1,109 UMIs per nucleus, **Figure S14A and S14B**) from twenty-four human brain samples across two brain regions (hippocampus, superior and middle temporal lobe (SMTG)), derived from six Alzheimer's disease patients and six age- and gender-matched controls (**Table S13**). Thirteen main cell types were identified through integration analysis with the mouse dataset and validated by the cluster-specific expression of known markers (**Figure 7A** and **Figure S14C-E**).

We next sought to investigate the region- and cell type-specific gene expression changes associated with human AD pathogenesis. By differential gene expression analysis, we identified a total of 4,171 and 2,149 cell-type-specific DE genes in the hippocampus and SMTG, respectively (**Figure 7B, Table S14**). 349 genes were significantly changed in the same cell type from two distinct regions, among which 332 were altered in concordant directions (**Figure 7C,** Pearson correlation coefficient r = 0.68, p-value < 2.2e-16). For example, oligodendrocytes from both regions exhibited down-regulated expression of an oligodendrocyte terminal differentiation factor *OPALIN* (de Faria et al., 2019) and an oxidation stress protector *OXR1* (Volkert and Crowley, 2020). Meanwhile, we detected an increased expression of genes related to programmed cell death (e.g., *FLCN* and *RASSF2*) (Cooper et al., 2009; Schmidt and Linehan, 2018), indicating an elevated stress level in oligodendrocytes from both regions. Other examples include the microglia-specific upregulation of *PTPRG*, a receptor protein tyrosine phosphatase that plays a key role in mediating AD-associated neuronal death (Luo et al., 2022). In astrocytes, we observed a decreased expression of several transmembrane transporters (e.g., *AQP4* and *SLCO1C1*) as well as neural transmitter metabolism enzymes (e.g., *GLUD1*), suggesting an impairment of blood-brain barrier (Silva et al., 2021) and altered metabolic state (Kulijewicz-Nawrot et al., 2013) in astrocytes from both regions of the AD brains.

Remarkably, some of the AD-associated gene signatures present with region-specific expression patterns. For example, *GPNMB*, a transmembrane glycoprotein associated with microglia activation in AD brains (Hüttenrauch et al., 2018), showed increased expression in the microglia from the hippocampus but not from SMTG (**Figure 7D, top**). On the other hand, *MMP24*, a gene of the matrix metalloproteinase family and extensively implicated in AD pathogenesis in previous studies (Zipfel et al., 2020), exhibited an increase in cortical projection neurons unique to SMTG (**Figure 7D, bottom**). In fact, inhibiting MMP24 has been shown to reduce the amyloid-beta levels and promoter cognitive functions in mouse models, serving as a novel target for AD therapy (Baranger et al., 2016).

Finally, we explored the human-mice relevance for certain AD-associated gene signatures and molecular pathways. Despite differences in the species and ages between the two datasets, several genes encoding heat shock proteins (*e.g., HSP90AA1*, *HSPH1*) were upregulated across multiple cell types in the EOAD mouse model and the human hippocampus (**Figure 7E**). The elevated expression of the chaperon system potentially serves as a compensatory mechanism to reduce the formation of toxic oligomeric assemblies in AD brains (Arawaka et al., 2010), further validating the dysfunction of proteostasis as a molecular marker of AD (Cornejo and Hetz, 2013). Meanwhile, we identified down-regulated genes in both human and mice (**Figure 7F**). One of the examples, *PLP1*, was reported as a subtype-specific driver gene contributing to AD pathogenesis (Neff et al., 2021). Another gene, *PDE10A*, plays a key role in the promotion of neuronal

survival, with its reduction detected in multiple neurodegenerative diseases (*e.g.,* Huntington's disease (Niccolini et al., 2015a), Parkinson's disease (Niccolini et al., 2015b)). Importantly, the above-mentioned trends were readily validated by another recently published single-cell dataset investigating Alzheimer's disease in the human prefrontal cortex (Morabito et al., 2021) (**Figure S15**). In summary, the human-mice relevance analysis further advances our current understanding of genetic programs associated with Alzheimer's pathogenesis.

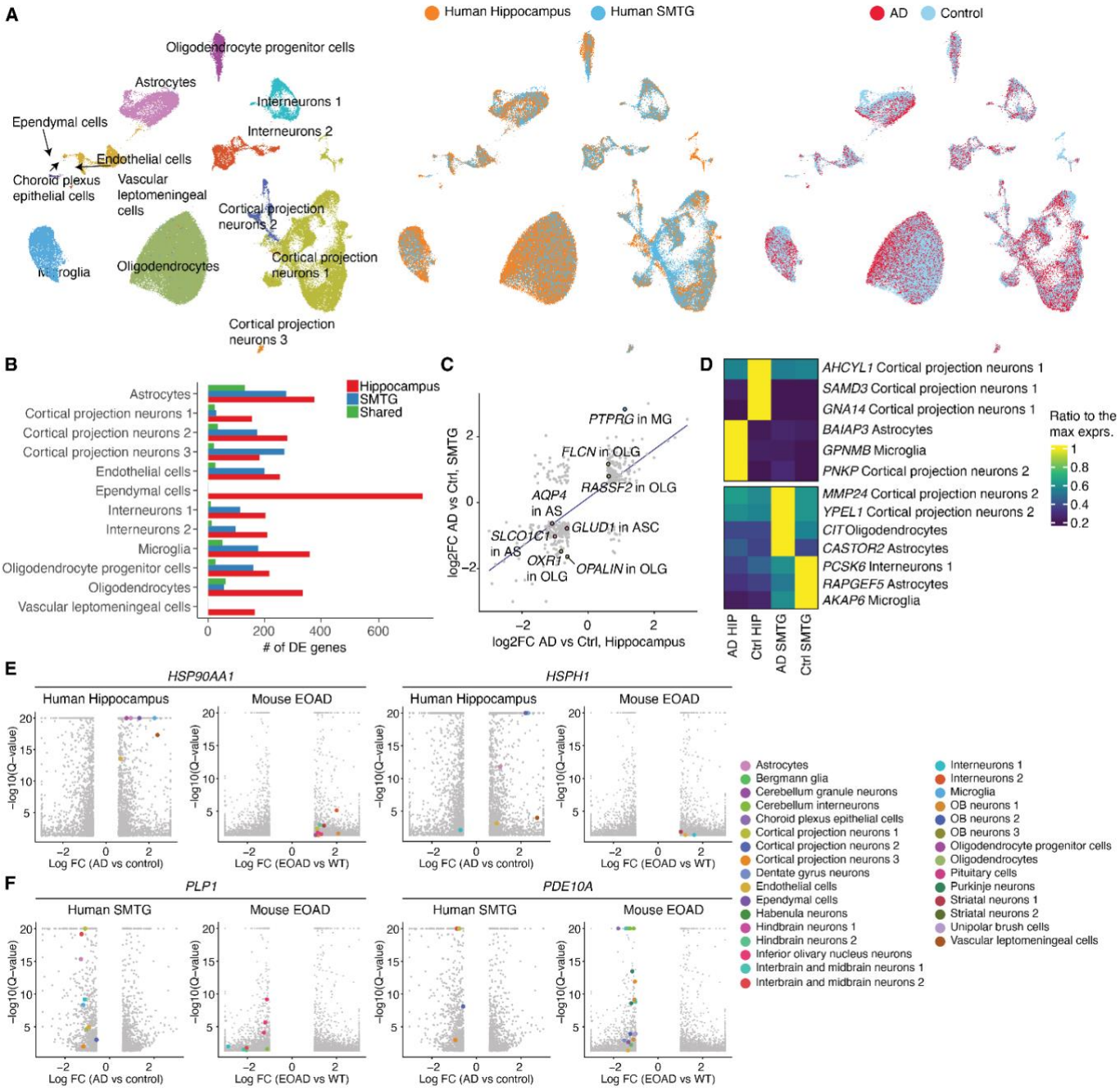


Figure 7. Identifying AD pathogenesis-associated gene expression signatures across regions and cell types in human brains.

(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 differentially expressed genes between AD and control samples in each cell type. DE genes are colored by whether they are unique to each region or shared between two regions. Of note, 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 scatterplot shows the correlation of the log2-transformed fold changes of these 394 shared DE genes in Hippocampus (x-axis) and in SMTG (y-axis). Key genes are annotated and colored by their corresponding main cell types. AS, astrocytes; MG, microglia; OLG, oligodendrocytes.
(D) Heatmaps showing examples of region-specific DE genes for the hippocampus (left) and SMTG (bottom). Gene expressions were quantified as transcripts per million 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 differentially expressed (DE) genes between the AD and control samples across main cell types in human brains or between EOAD and WT samples across cell subclusters in mouse brains. Highlighted genes are colored by the main cell type identity.

Discussion

In this study, we obtained a global view of aging and AD pathogenesis-associated cell population dynamics, by profiling ~1.5 million single-cell transcriptomes at full gene body coverage and ~380,000 single-cell chromatin accessibility profiles across the entire mammalian brains spanning various age and genotype groups. With the resulting datasets, we identified over 300 cellular subtypes across the brain, including extremely rare cell types (*e.g.,* pinealocytes, tanycytes) representing less than 0.01% of the brain cell population. In addition, we detected region-specific aging and AD effects with high-resolution spatial transcriptomic analysis and explored the cell-type-specific manifestation of aging and AD-associated molecular signatures. With the *EasySci* method, we introduced a technical framework for individual laboratories to generate gene expression and chromatin accessibility profiles from millions of single cells cost-effectively. We have made the *EasySci* pipeline, detailed experimental protocols, computation scripts, and datasets freely available to facilitate further exploration of the techniques and datasets.

As illustrated by our sub-cluster level analysis, the effects of aging and AD on the global brain cell population are highly cell-type-specific. While most brain cell types stay relatively stable under various conditions, we identified many cell subtypes that are significantly changed (over two-fold change) in aged and AD model brains, most of which were rare cell types and thus presumably missed in conventional "shallow" single-cell analysis. For example, the aged 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 subventricular zone (SVZ), suggesting a potential interplay between oligodendrocytes, local inflammatory signaling and the stem cell niche.

The lack of reliable mouse models remains one of the biggest challenges in studying late-onset Alzheimer's disease. The novel APOE*4/Trem2*R47H model aims to overcome this limitation by introducing two of the strongest late-onset Alzheimer's disease-associated mutations (Karch and Goate, 2015). However, limited validation is available to assess whether this novel model shows any characteristics of Alzheimer's disease. Here we observed overall concordant molecular and cell population dynamics between the well-established 5xFAD and the novel APOE*4/Trem2*R47H model, which emphasizes that the novel LOAD model indeed shows signs of Alzheimer's disease. For example, we observed shared subtypes that were depleted (*e.g., mt-Cytb*+ *mt-Rnr2*+ choroid plexus epithelial cell) or enriched (*e.g., Col25a1*+ *Ndrg1*+ interbrain and midbrain neuron) in both early- and late-onset AD mutant brains, validated by single-cell RNA-seq from both sexes as well as spatial transcriptomics analysis. On the other hand, differences were also observed between the two AD models, as expected by the different onset times. Most notably, the missing of the disease-associated microglia population increase in the LOAD model could be explained by the lack of amyloid deposition in the mouse model (Kotredes et al., 2021) or by genetic perturbations, as both Trem2 and Apoe play a role in the activation of this cell population(Keren-Shaul et al., 2017). To answer this question, further studies are needed to characterize APOE*4/Trem2*R47H mouse models at late stages.

In addition, we further explored AD-associated gene signatures in human brains by profiling over 100,000 single nucleus transcriptomes from twenty-four human brain samples from control and AD patients and two anatomical locations. While most AD-associated gene dynamics are highly cell-type- and region-specific, we identified dysregulated genetic signatures that are conserved between different anatomical 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 could advance our knowledge of biomarkers for Alzheimer's diagnosis.

In summary, this study demonstrated the potential of novel 'high-throughput' single-cell genomics for quantifying the dynamics of rare cell types and novel subtypes associated with development, aging, and disease. We anticipate that further development of high-throughput single-cell profiling strategies and computation approaches will enable a comprehensive view of cell-type-specific dynamics across all mammalian organs through "saturate sequencing", which may be especially critical for identifying rare cell types in human samples.

Endnotes

Acknowledgments: We thank all members of the Cao lab for helpful discussions and feedback. We thank Dr. Jay Shendure (University of Washington) for insightful feedback on this work. We also thank members of the Rockefeller University Genomics Resource Center and Comparative Bioscience Center for their exceptional assistance with library sequencing and animal maintenance.

Funding: This work was funded by grants from NIH (DP2HG012522, R01AG076932 and RM1HG011014 to J.C.; P30AG072946 and P01AG078116 to P.T.N.; R01AG072758 to L.G.) and the Sagol Network GerOmic Award for J.C.

Author contributions: J.C. and W.Z.. conceptualized and supervised the project. J.L. and A.S. developed the experimental and computation pipeline for *EasySci-RNA* profiling of all samples. G.B. and Z.L. developed the experimental and computation 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 experiment. A.S. and Z.L. performed the downstream analysis with assistance from E.M., A.L., A.E., Z.X., and Z.Z. J.C., W.Z., Z.L., and A.S. wrote the manuscript with input and biological insight from P.N., L.G. and other co-authors.

Supplementary Figures

Key steps	sci-RNA-seq3	EasySci-RNA optimizations	Improvements
Nuclei extraction	Nuclei lysis buffer (1% SUPERase)	EZ lysis buffer (1% DEPC + 0.1% SUPERase)	Lower mRNA degradation and reduced cost
Nuclei fixation	4% Paraformaldehyde (15min)	0.1% formaldehyde (10min)	Higher efficiency, less nuclei clumping
Nuclei storage	Flash freezing	Slow freezing (in 10% DMSO)	Higher cell recovery rate
Indexed reverse transcription	Enzyme: SuperScript IV Primer: Oligo-dT primer	Enzyme: Maxima RT Primer: Oligo-dT + indexed random primer	Higher efficiency, reduced cost and full gene body coverage
Indexed ligation	Enzyme: quick ligase Primer: sci-RNA-seq3 primer	Enzyme: T4 ligase, Primer: re-designed primer with longer annealing length and additional chemical modifications	Higher cell recovery rate, reduced primer dimers in the final library, and shorter sequencing length requirement
Second strand synthesis	NEB Second strand synthesis module	NEBnext Ultra II Non-directional RNA second strand synthesis module; Ampure beads purification after Second strand synthesis	Shorter reaction time; higher efficiency and reduced primer dimers in the final library
Tagmentation	Loaded Tn5 from Illumina (not commercially available); Ampure beads purification after tagmentation	Custom loaded Tn5 enzyme; no Ampure beads purification after tagmentation	Easy access to the reagent; higher efficiency
Library amplification and purification	Two rounds of Ampure beads purification	One round of Ampure beads purification	Reduced cost
Sequencing data processing	A custom pipeline for processing single-read RNA-seq data for gene counting	A new pipeline for processing paired-end RNA-seq data for gene counting at exon resolution	Exon level expression quantification

Figure S1. Summary of key optimizations of *EasySci-RNA* compared to published single-cell RNA-seq by combinatorial indexing (sci-RNA-seq3 (Cao et al., 2019)**).**

733

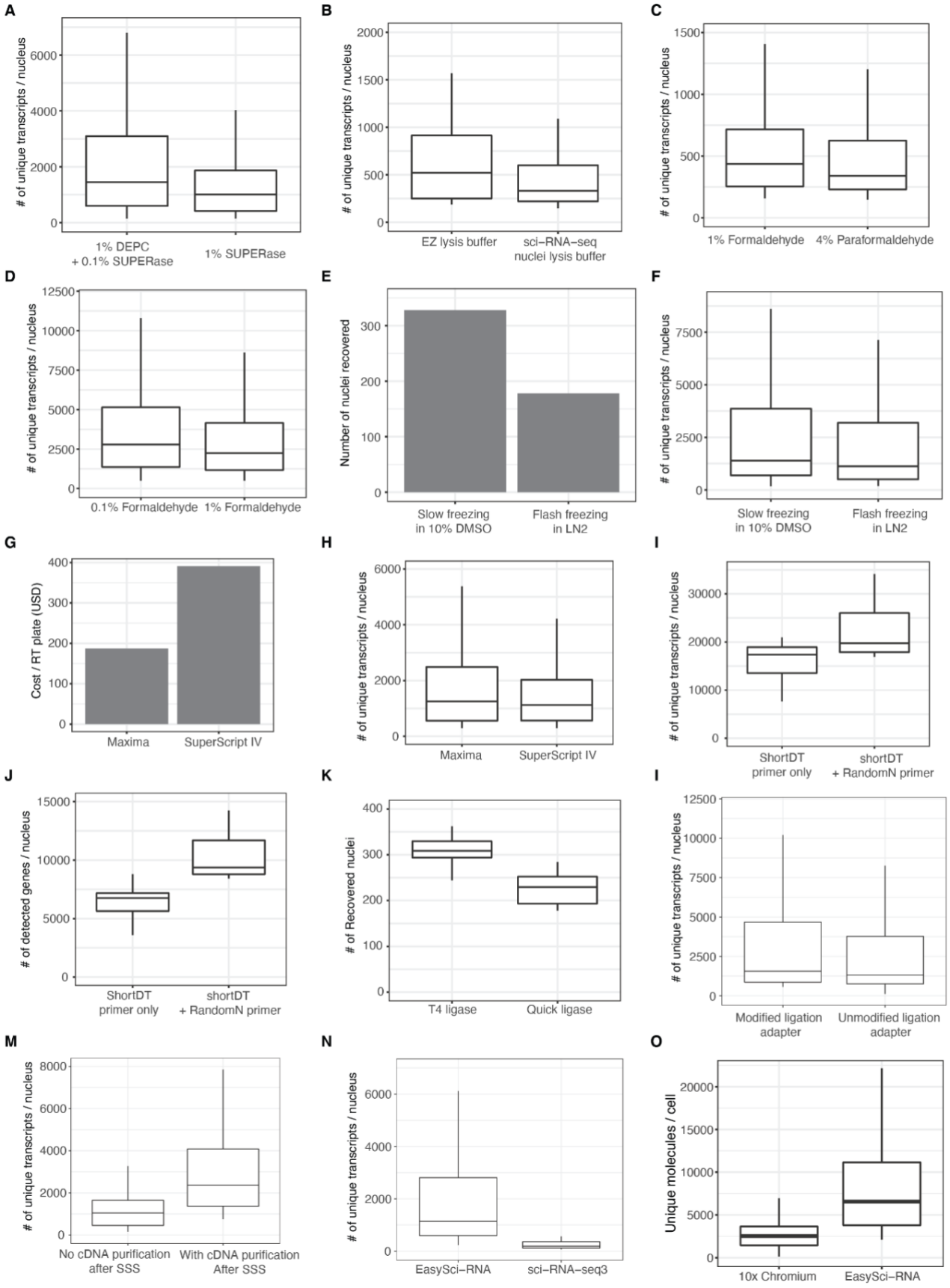


Figure S2. Representative examples showing the performance of optimized conditions of *EasySci-RNA*.

(A-B) Box plots showing the number of unique transcripts detected per nucleus in different lysis conditions: 1% DEPC vs. no DEPC in lysis buffer (A); EZ lysis buffer vs. nuclei lysis buffer used in the published sci-RNA-seq3 (Cao et al., 2019) (B). For all box plots in this figure: middle lines, medians; upper and lower box edges, first and third quartiles, respectively; whiskers, 1.5 times the interquartile range.

(C-D) Box plot showing the number of unique transcripts detected per nucleus across different fixation conditions: 1% formaldehyde vs 4% paraformaldehyde (C); 0.1% formaldehyde vs. 1% formaldehyde (D).

(E-F) We compared two conditions for preserving the fixed nuclei. The slow freezing condition (in 10% DMSO) outperformed the flash freezing condition in sci-RNA-seq3 (Cao et al., 2019) by increasing the number of nuclei recovered in the experiment (E) and the number of unique transcripts detected per nucleus (F).

(G-H) Maxima reverse transcriptase greatly reduces the enzyme cost (G) without affecting the number of transcripts detected per nucleus (H).

(I-J) We included both short oligo-dT and random primers in reverse transcription to increase the number of unique transcripts (I) and genes (J) detected per nucleus.

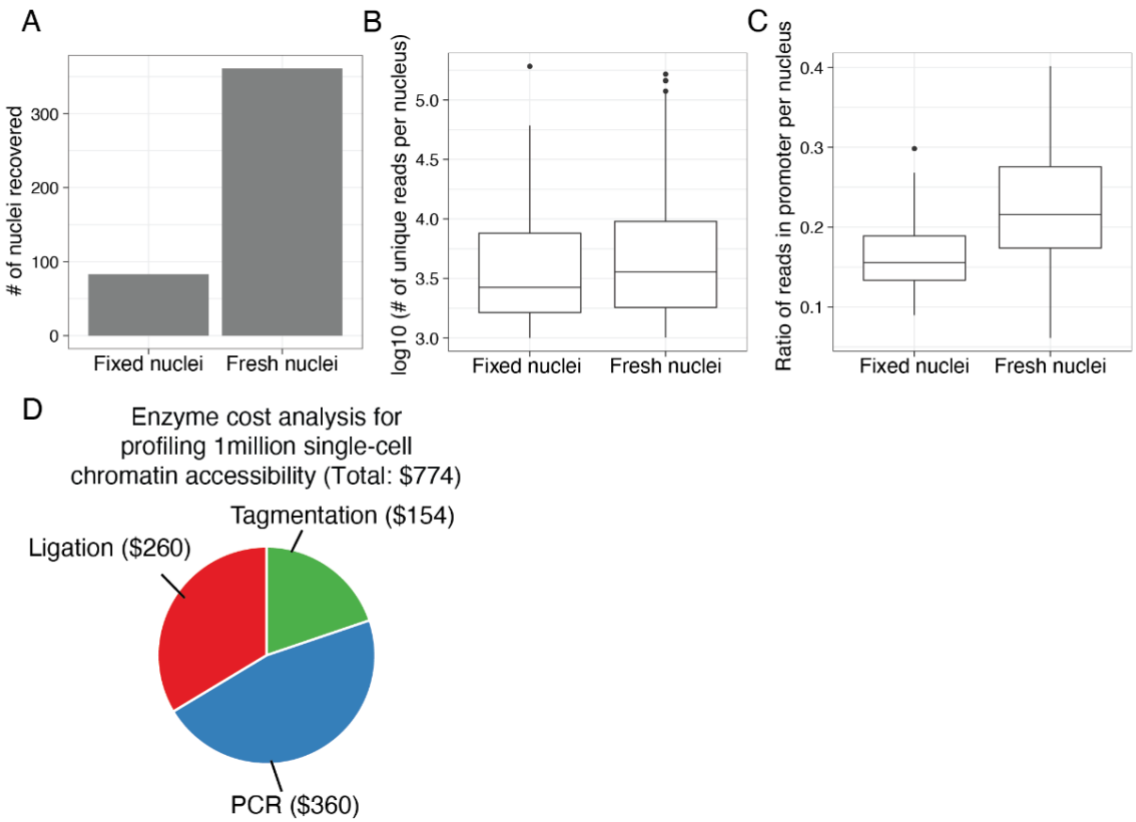
(K) *EasySci-RNA* used T4 ligase instead of quick ligase for a higher recovery rate of nuclei.

(I) We used chemically modified ligation primers in *EasySci* (Method), which greatly reduced primer dimers in the following PCR reaction and slightly increased the number of unique transcripts detected per nucleus.

(M) Additional cDNA purification step after second strand synthesis increased the number of unique transcripts per nucleus.

(N) We compared the efficiency of the novel *EasySci-RNA* method with the sci-RNA-seq3 using mouse brain nuclei. The raw data was subset to 4,448 reads/cell to remove any potential bias from sequencing depth.

(O) Box plot showing the number of unique transcripts detected per mouse brain nucleus in a deep sequenced dataset comparing *10X genomics* and a small-scale *EasySci-RNA* library at similar sequencing depth (~ 20,000 raw reads/cell, Methods).



762
763 **Figure S3. Representative examples showing the performance of optimized conditions of *EasySci-***
764 ***ATAC*.**
765 (A-C) We compared two fixation conditions: nuclei were either fixed with 1% formaldehyde for 10 minutes
766 at room temperature or directly used for tagmentation without fixation. The unfixed condition outperformed
767 the fixed condition by increasing cell recovery (A), the number of reads (B), and the ratio of reads in
768 promoters (C) per nucleus. For all box plots in this figure: middle lines, medians; upper and lower box
769 edges, first and third quartiles, respectively; whiskers, 1.5 times the interquartile range; circles, outliers.
770 (D) Pie chart showing the estimated enzyme cost compositions of library preparation for profiling 1 million
771 single-cell chromatin accessibility profiles using *EasySci-ATAC*.

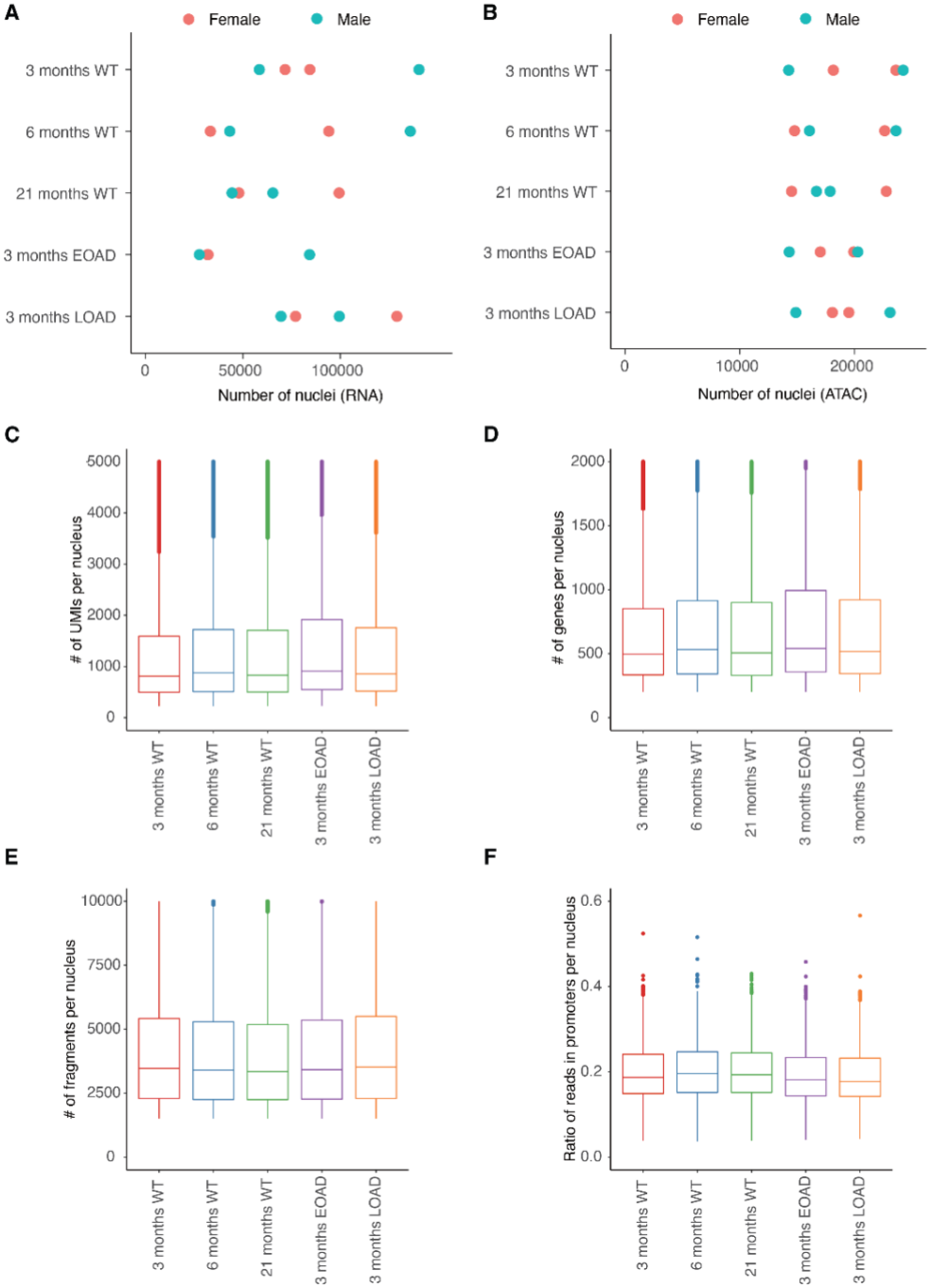


Figure S4. Performance of *EasySci-RNA* and *EasySci-ATAC* profiling of mouse brain samples.
(A-B) Scatter plots showing the number of single-cell transcriptomes (A) and single-cell chromatin accessibility (B) profiled in each mouse individual across five conditions, colored by sex. Of note, the number of cells recovered from two mouse individuals in the EOAD model (RNA) are very close and can not be separated in the plot.
(C-D) Box plots showing the number of unique transcripts (C) and genes (D) detected per nucleus in each condition profiled by *EasySci-RNA*. For all box plots: middle lines, medians; upper and lower box edges,

780 first and third quartiles, respectively; whiskers, 1.5 times the interquartile range; and circles are outliers.
781 (E-F) Box plots showing the number of unique fragments (E) and the ratio of reads in promoters (F) per
782 cell in each condition profiled by *EasySci-ATAC*.
783

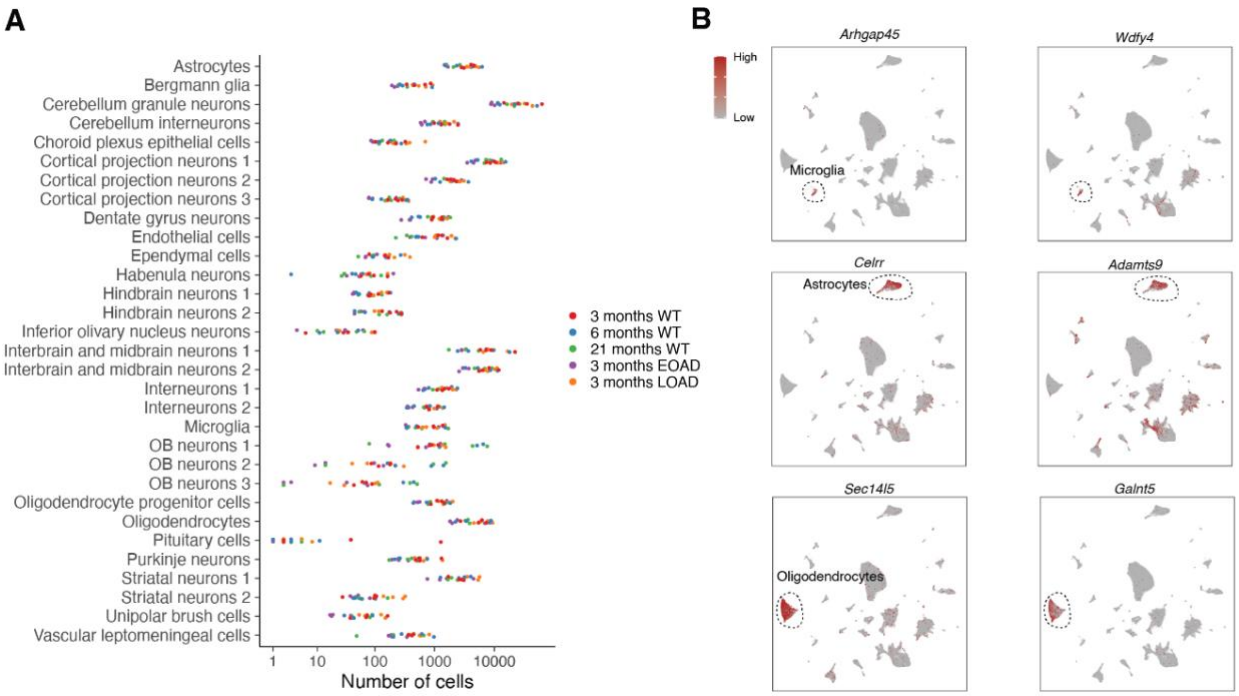
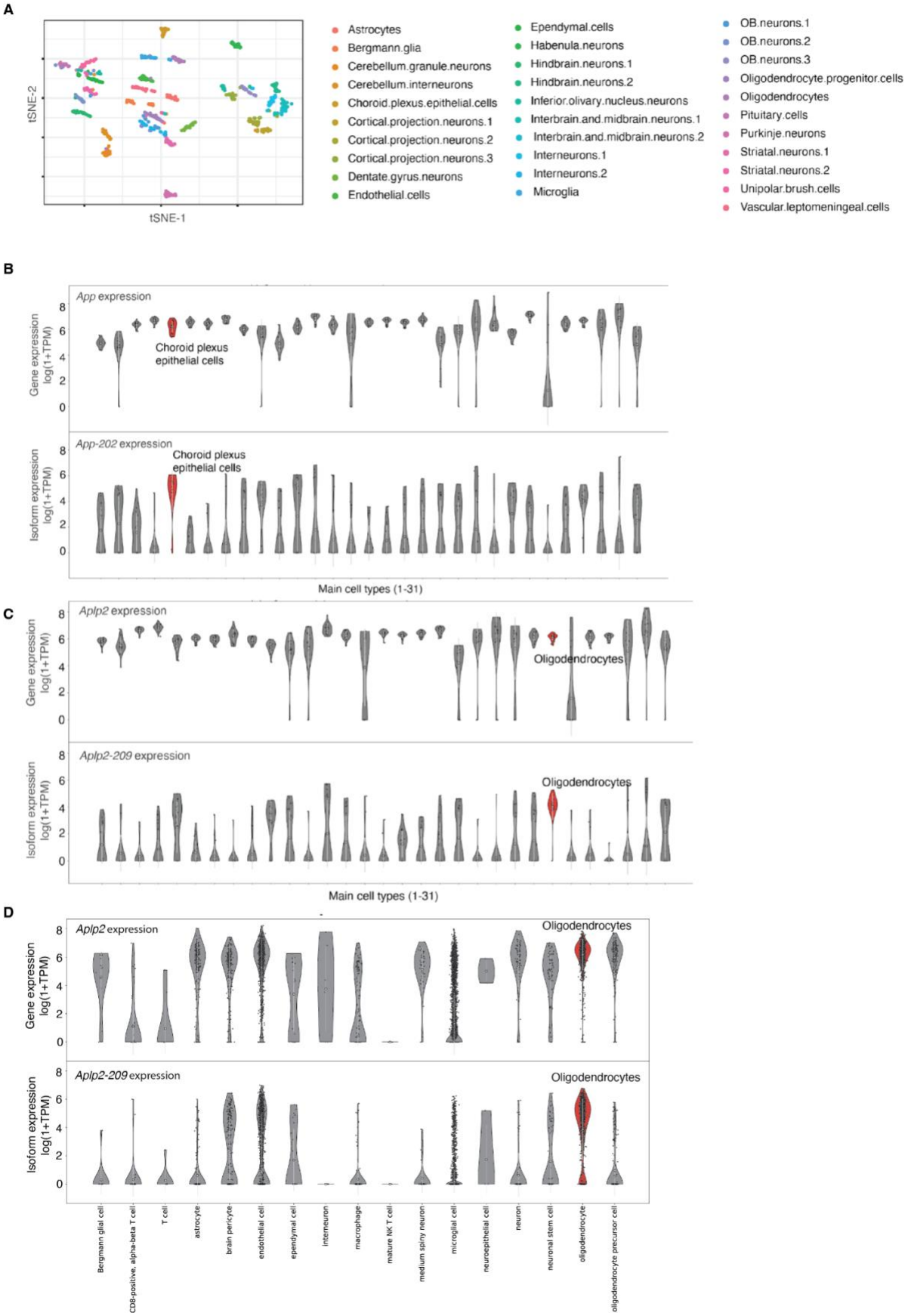


Figure S5. Identification of main brain cell types and cell-type-specific markers by *EasySci-RNA*. (A) Dot plot showing the number of single-cell transcriptomes recovered from each individual, colored by conditions.

(B) 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.



792

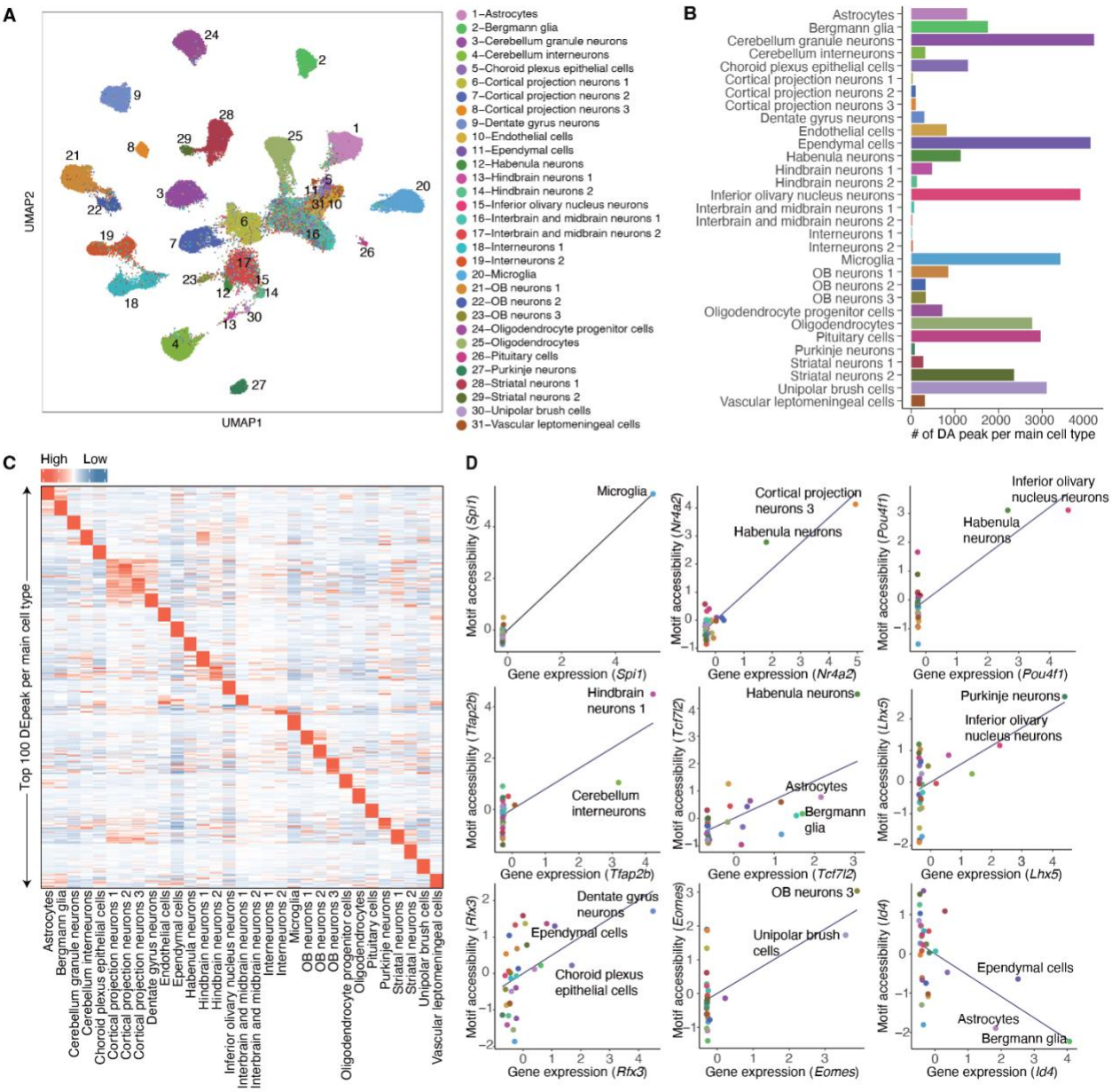
Figure S6. Identification of cell-type-specific isoforms in the mouse brain.

(A) We aggregated randomN primed *EasySci-RNA* reads from each main cell type in every mouse individual, yielding 613 pseudocells. The t-SNE plot showed the separation of main cell types by isoform expression.

(B) Violin plots showing the expression of gene *App* and isoform *App-202* across main cell types.

(C-D) Violin plots showing the expression of gene *Aplp2* and isoform *Aplp2-209* across main cell types in the *EasySci* dataset (C) and the Tabula Muris Senis mouse aging atlas dataset (Tabula Muris Consortium, 2020) (D). White circles represent the normalized expression of genes and isoforms (log(1+TPM)). White bars represent standard deviation.

804



805 **Figure S7. Characterization of cell-type-specific chromatin accessibility and key TF regulators**
806 **using *EasySci-ATAC*.**
807 (A) UMAP plot of the *EasySci-ATAC* dataset subsampled to 5,000 cells per cell type (or all cells if the
808 number of cells is less than 5,000), colored by main cell types in Figure 2C. The analysis was performed
809 using the peak-count matrix without integration with the *EasySci-RNA* dataset.
810 (B) Bar plot showing the number of cell-type-specific peaks for each main cell type (defined as differential
811 accessible (DA) peaks across main cell types with q-value < 0.05 and TPM > 20 in the target cell type).
812 (C) Heatmap showing the aggregated accessibility of top 100 DA peaks per cell type (ranked by fold
813 change between the maximum and the second accessible cell type). Unique counts for cell-type-specific
814 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 chromVar (Schep et al., 2017), then aggregated per main cell type and mapped to Z-scores (**Methods**).

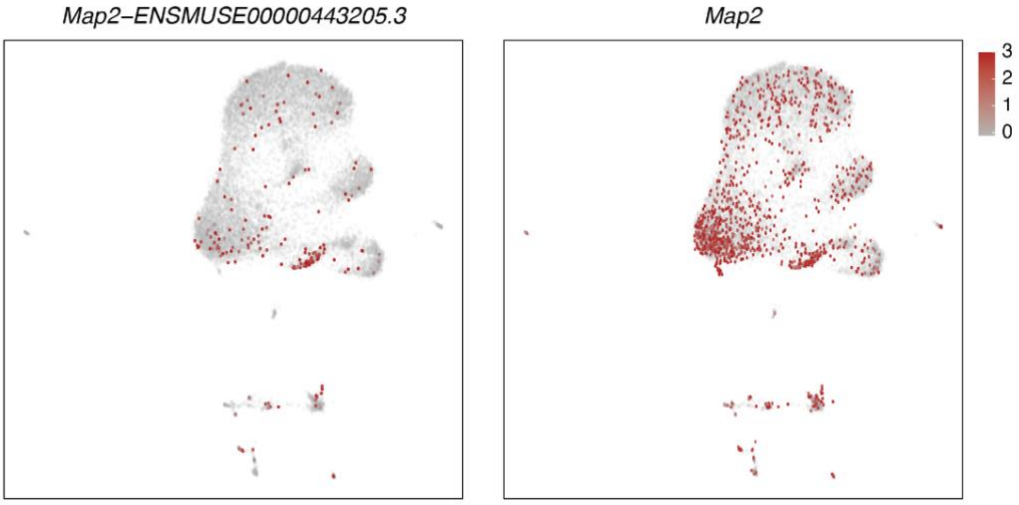


Figure S8. Characterizing microglia subtypes incorporating both gene and exon level expression. UMAP plots same as Figure 3A based on both gene and exon-level expression, showing the specific expression of an example exon marker *Map2-ENSMUSE00000443205.3* (left) of microglia sub-cluster 8 and the lack of specificity of its corresponding gene *Map2* (right). Single-cell gene/exon expression was normalized first by library size, log-transformed, and then scaled to Z-scores.

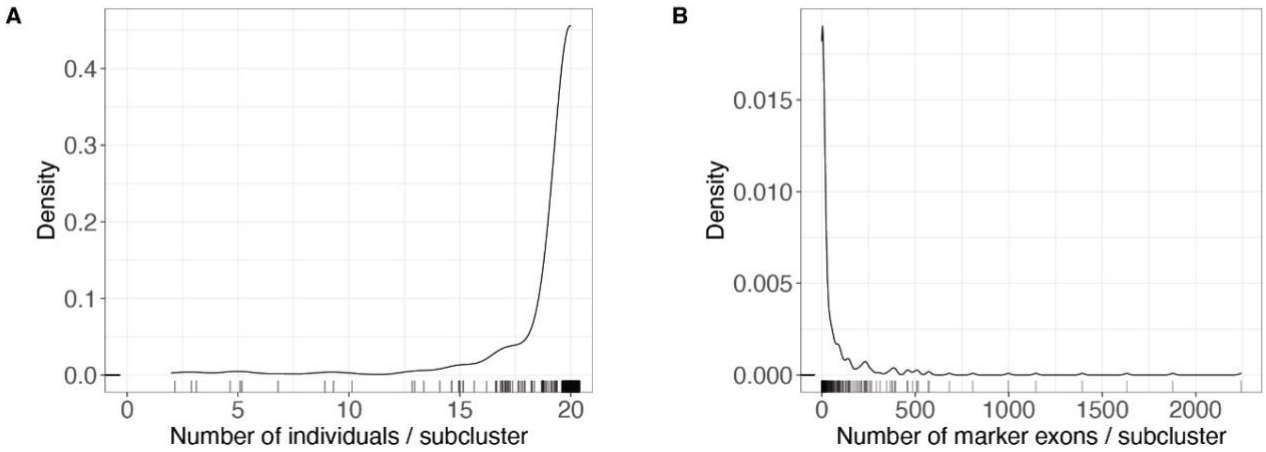


Figure S9. Characteristics of subclusters.

(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.

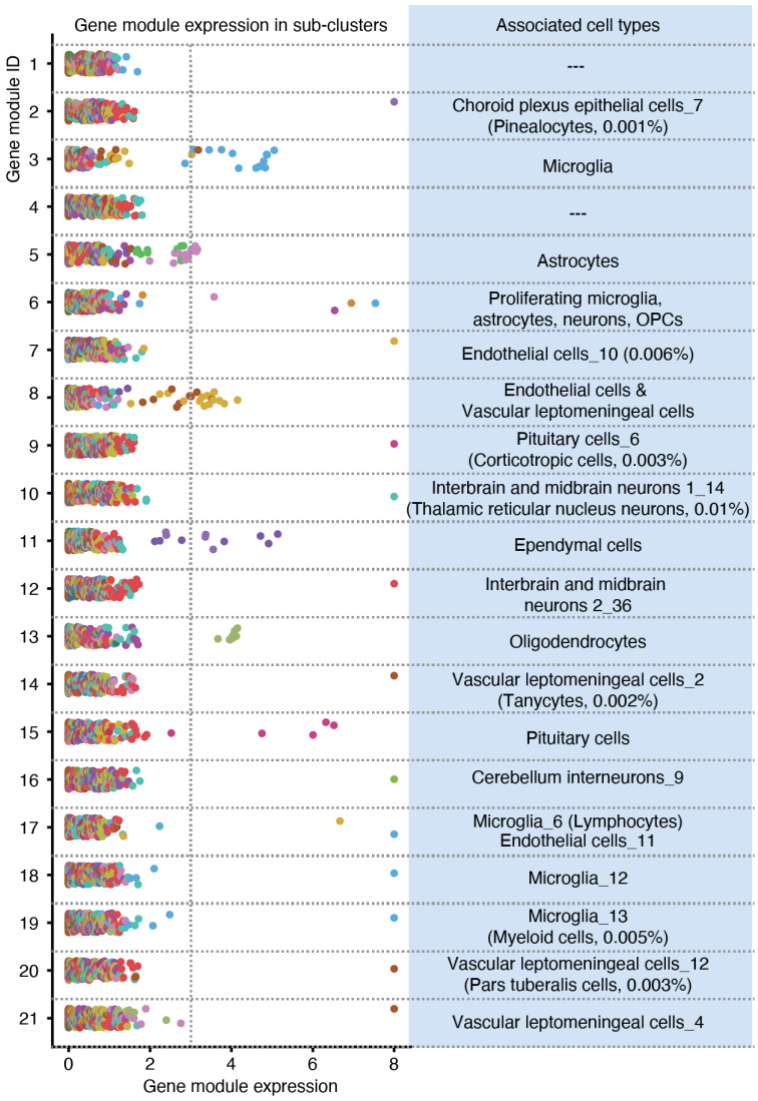


Figure S10. Characterization of cell types/subtypes by gene module expression. Scatter plot showing the expression of each gene module across 359 sub-clusters. The associated cell types were annotated on the plot. UMI counts for genes from each gene module are scaled for library size, log-transformed, aggregated, and then mapped to Z-scores.

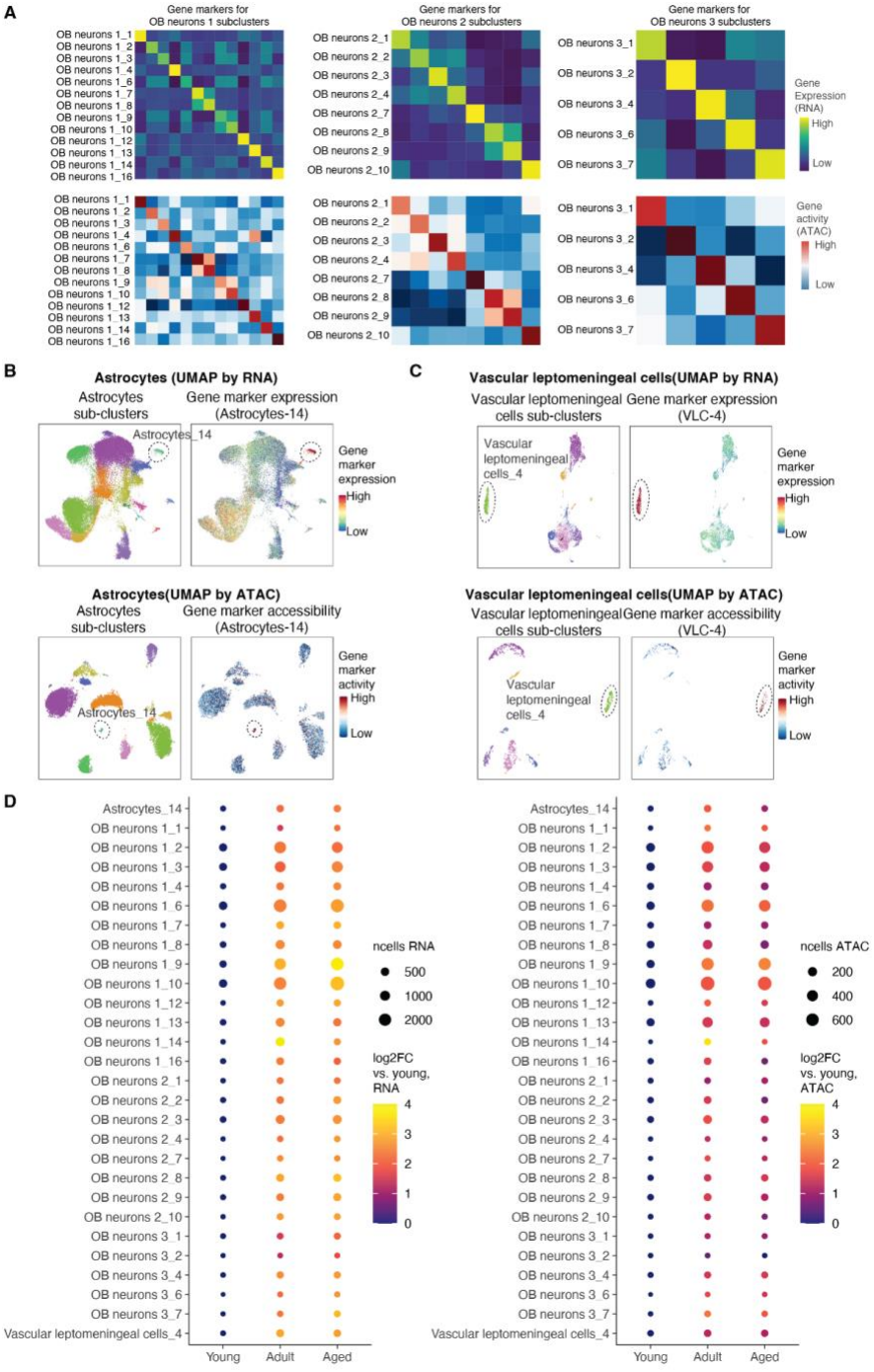


Figure S11. Identification of cell subtypes underlying olfactory bulb expansion from the young to adult stage in *EasySci-RNA* and *EasySci-ATAC*.

(A) Heatmaps showing the aggregated gene expression (top) and gene body accessibility (bottom) of sub-cluster specific gene markers (columns) in OB expansion-associated sub-clusters (rows) from OB neurons 1 (left), OB neurons 2 (middle), and OB neurons 3 (right). UMI counts for genes or reads overlapping with gene bodies were aggregated for each sub-cluster, normalized first by the total number of reads, column centered, and scaled across all cell sub-clusters.

(B-C) UMAP visualization showing astrocytes subtype 14 (B) and vascular leptomeningeal cells subtype 4 (VLC-4, C), colored by subcluster ID in *EasySci-RNA* (top left) and *EasySci-ATAC* (bottom left), the

aggregated gene expression (top right) and gene body accessibility (bottom right) of sub-cluster specific gene markers.
(D) For the OB expansion-related sub-clusters, we plotted their log2-transformed fold changes between each age group and the young mice, profiled by *EasySci-RNA* (left) and *EasySci-ATAC* (right).

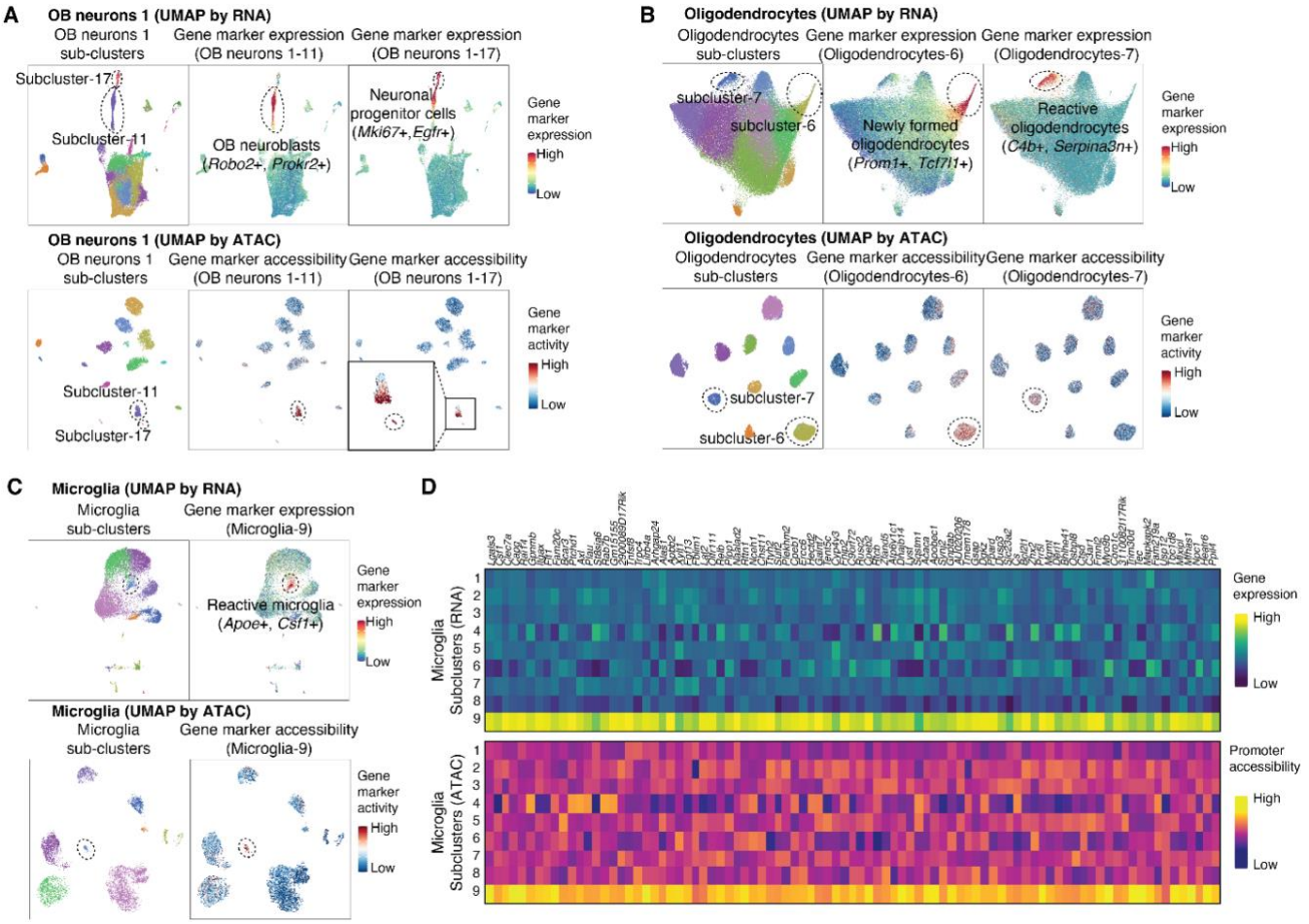


Figure S12. Identifying aging-associated sub-clusters related to neurogenesis, oligodendrogenesis, and inflammation in *EasySci-ATAC*.

(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 sub-cluster, 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.

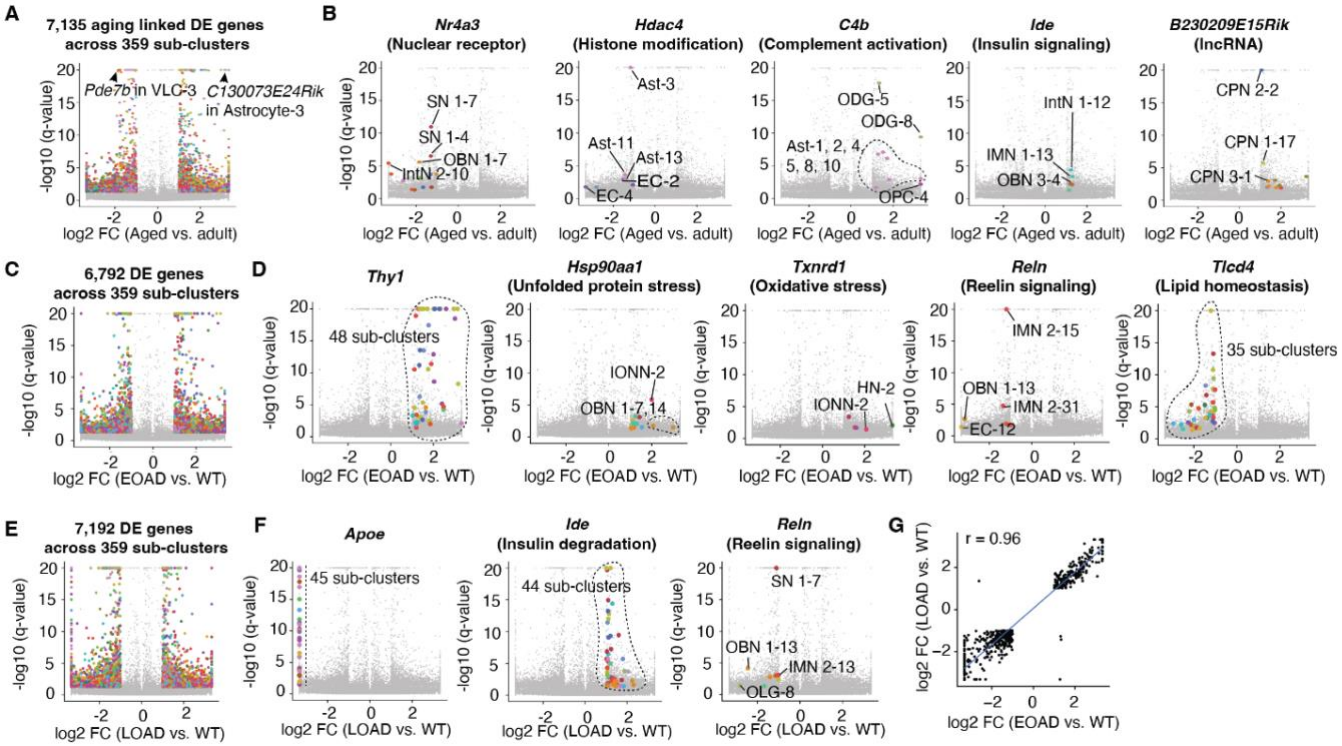


Figure S13. Identifying aging and AD pathogenesis-associated gene expression signatures.

(A) Volcano plots showing the differentially expressed (DE) genes between adult (6 months) and aged (21 months) mice across all sub-clusters. Significantly changed genes are colored by the main cell type identity for the corresponding sub-cluster.

(B) Volcano plot same as (A), highlighting example DE genes with concordant changes across multiple sub-clusters comparing adult and aged models, labeled with related biological pathways.

(C) Volcano plots showing the differentially expressed (DE) genes between WT and EOAD models across all sub-clusters. Significantly changed genes are colored by the main cell type identity for the corresponding sub-cluster.

(D) Volcano plot same as (C), highlighting example DE genes with concordant changes across multiple sub-clusters comparing WT and EOAD models, labeled with related biological pathways.

(E) Volcano plots showing the differentially expressed (DE) genes between WT and LOAD models across all sub-clusters. Significantly changed genes are colored by the main cell type identity for the corresponding sub-cluster.

(F) Volcano plot same as (E), highlighting example DE genes with concordant changes across multiple sub-clusters comparing WT and LOAD models, labeled with related biological pathways.

(G) We detected 559 DE genes significantly changed within the same sub-cluster in both AD models (both compared with the wild-type). The scatter plot shows the correlation of the log2-transformed fold changes of these 559 shared DE genes in the EOAD model (x-axis) and the LOAD model (y-axis).

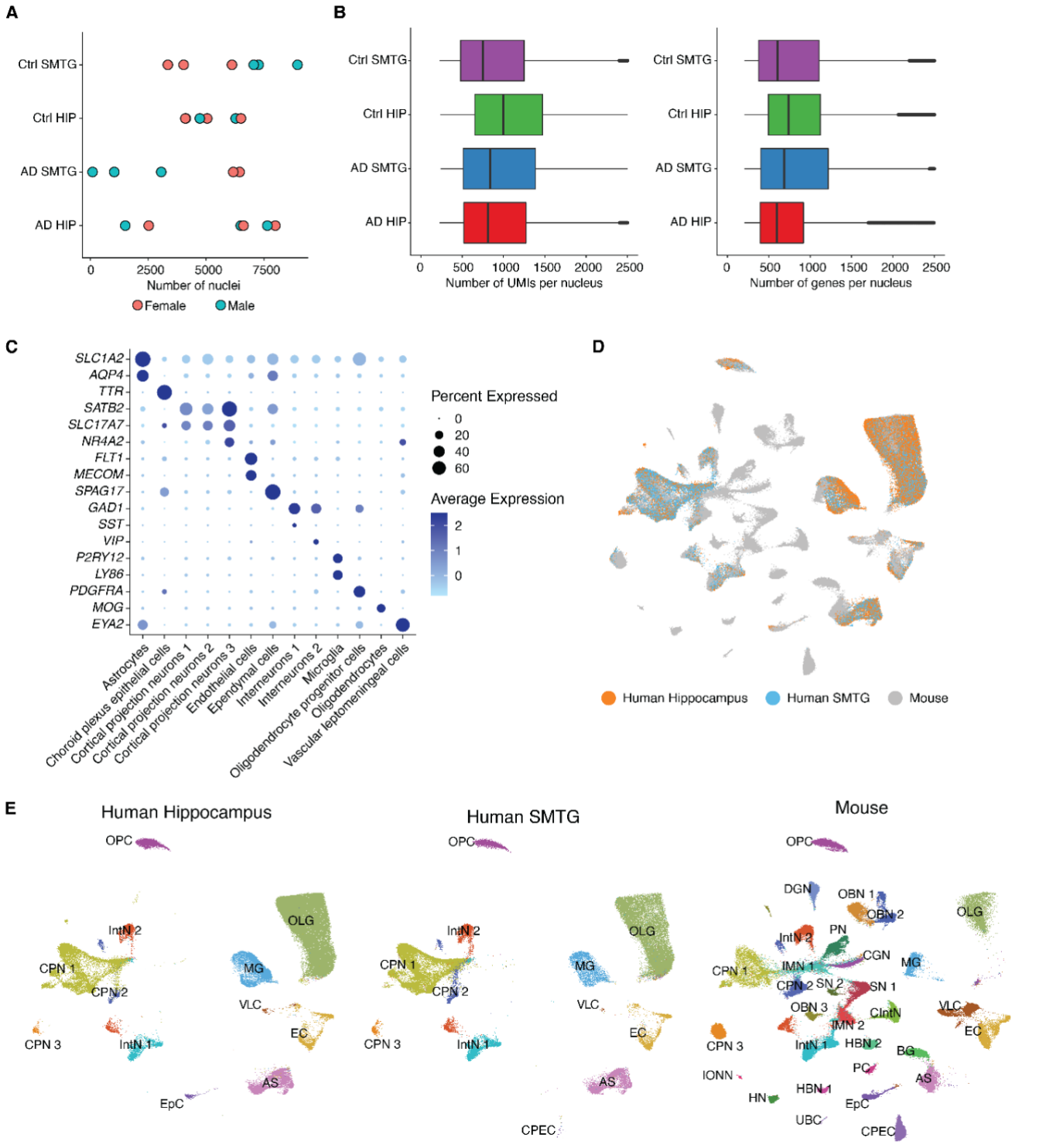
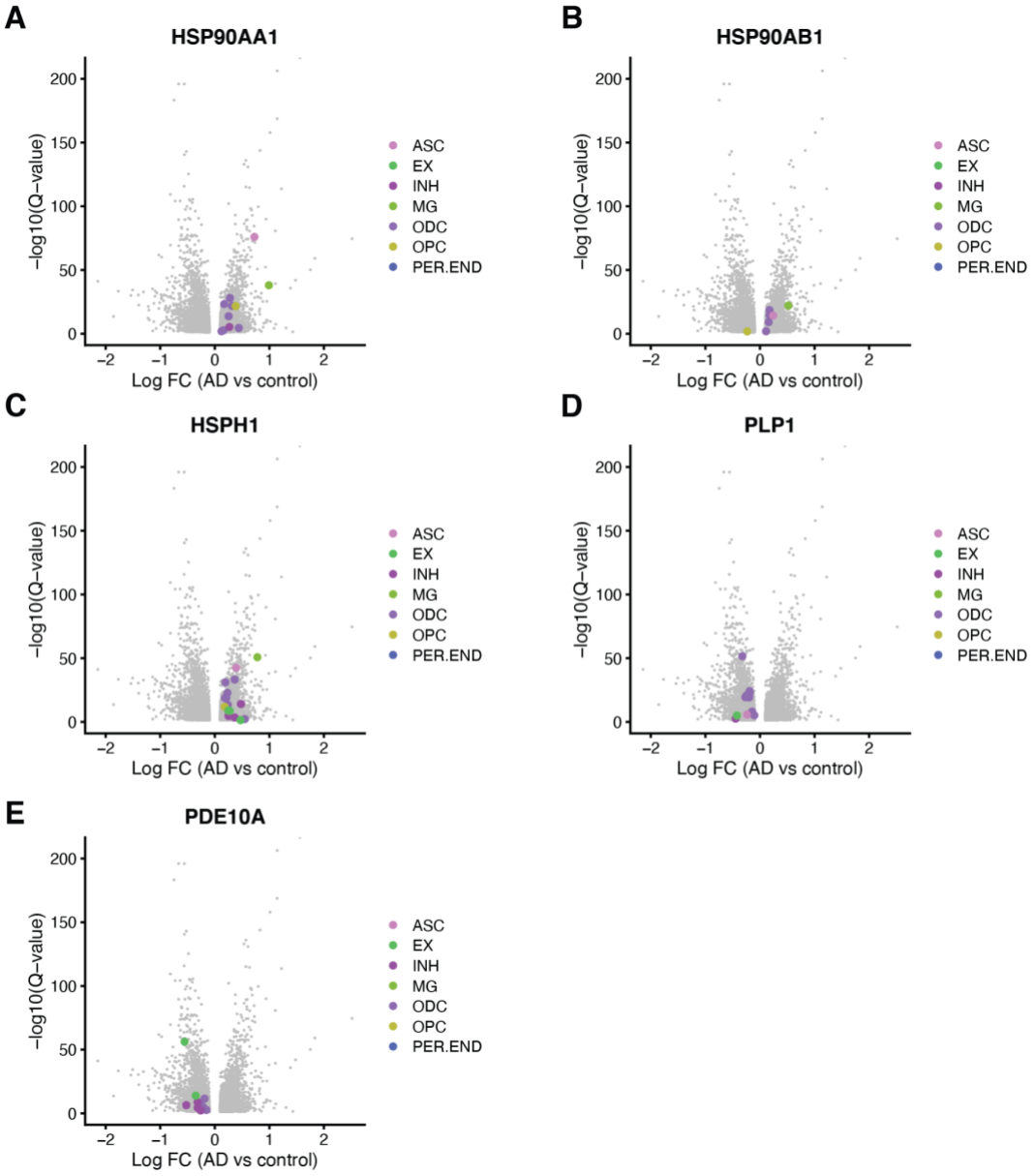


Figure S14. Performance and quality control of the human brain dataset.

(A) Scatter plot showing the number of single-cell transcriptomes profiled in each human sample in two regions, colored by sexes. Of note, the number of cells recovered from two AD individuals in the SMTG are very close and cannot be separated in the plot.

(B) Box plots showing the number of unique transcripts (left) and genes (right) detected per nucleus profiled in the human dataset. For all box plots: middle lines, medians; upper and lower box edges, first and third quartiles, respectively; whiskers, 1.5 times the interquartile range; and circles are outliers.
(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; CIntN, cerebellum interneurons; CPEC, choroid plexus epithelial cells; CPN 1, cortical projection neurons 1; CPN 2, cortical projection neurons 2; CPN 3, cortical projection neurons 3; DGN, dentate gyrus neurons; EC, endothelial cells; EpC, ependymal cells; HN, habenula neurons; HBN 1, hindbrain neurons 1; HBN 2, hindbrain neurons 2; IONN, Inferior olivary nucleus neurons; IMN 1, interbrain and midbrain neurons 1; IMN 2, interbrain and midbrain neurons 2; IntN1, interneurons 1; IntN2, interneurons 2; MG, microglia; OBN 1, OB neurons 1; OBN 2, OB neurons 2; OBN 3, OB neurons 3; OPC, oligodendrocyte progenitor cells; OLG, oligodendrocytes; PC, pituitary cells; PN, purkinje cells; SN 1, striatal neurons 1; SN 2, striatal neurons 2; UBC, unipolar brush cells; VLC, vascular leptomeningeal cells.



925
926

927 **Figure S15. Identifying conserved gene expression changes across mouse AD models and human**
928 **AD samples from the prefrontal cortex.** (A-E) Volcano plots of genes from human prefrontal cortex
929 samples (Morabito et al., 2021). These genes show consistent changes in multiple cell subclusters
930 between mouse AD models, human hippocampus and SMTG samples, and human prefrontal cortex
931 samples from the above-mentioned publication. Dots on the figures represent genes across cell
932 subclusters; abbreviations correspond to the following cell types: ASC, astrocytes; EX, excitatory neurons;
933 INH, inhibitory neurons; MG, microglia; ODG, oligodendrocytes; OPC, oligodendrocyte progenitor cells;
934 and PER.END, pericyte/endothelial cells.

935

Materials and Methods:

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. These age points correspond to approximately 20, 30, and 62 years in humans. Furthermore, to gain insight into the early cellular state changes underlying the pathophysiology of Alzheimer's disease, we added two AD models at 3-month-old from the same C57BL/6 background. These include an early-onset AD model (5xFAD, JAX stock #034840) that overexpresses mutant human amyloid-beta precursor protein (APP) with the Swedish (K670N, M671L), Florida (I716V), and London (V717I) Familial Alzheimer's Disease (FAD) mutations and human presenilin 1 (PS1) harboring two FAD mutations, M146L and L286V. Brain-specific overexpression is achieved by neural-specific elements of the mouse *Thy1* promoter (Oakley et al., 2006). 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 LOAD (Karch and Goate, 2015), including a humanized *APOE* knock-in allele, where exons 2, 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.

By studying 3-month-old animals, our goal was to gain insight into the early changes underlying the pathophysiology of the AD models. Mature adult mice start at the age of 3 months, but multiple AD hallmarks, including amyloid-beta plaques and gliosis, can be observed in the early-onset 5xFAD model (Oakley et al., 2006). Therefore, we decided that this age might be the most appropriate for our goal to study early contributors of Alzheimer's disease pathomechanism.

EasySci-RNA library preparation and sequencing

Extracted mouse brains were snap-frozen in liquid nitrogen and stored at -80°C. Detailed step-by-step *EasySci-RNA* protocol is included as a supplementary file (**Supplementary file 1**).

Human brain sample

Twenty-four post-mortem human brain samples across two regions (hippocampus and superior and middle temporal gyrus) and twelve individuals, including six controls and six Alzheimer's disease patients ranging from 70 to 94 in age, were collected from the University of Kentucky Alzheimer's Disease Center Tissue Bank. Each surveyed sample underwent rigorous quality control, including short PMI, and was subjected to *EasySci-RNA* profiling. The libraries were sequenced across four Illumina NextSeq™ 1000 sequencer runs.

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 (Cao et al., 2019, 2020), 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 Galore (Krueger et al., 2021) 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 aligner (Dobin et al., 2013), and the PCR duplicates were removed based on the UMI sequence and the alignment location. Finally, the reads are split into SAM files per cell, and the gene expression is counted using a custom script. At this level, 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 1000 bp upstream of the read or genes on the opposite strand. Reads without any overlapped genes were discarded.

We used a similar strategy to generate an exon count matrix across cells. Specifically, we counted the number of expressed exons based on the number of reads overlapping each exon. If one read overlapped with multiple exons, this read was split between the exons. Read overlapped with multiple genes were discarded, except if we can determine the exact gene based on the other paired-end read. For reads without overlapped genes, we checked if there are any overlapped exons on the opposite strand. Reads without any overlapped exons were discarded.

To compare the performance of *EasySci-RNA* with the commercial 10x Chromium system on mouse brain samples, we subsampled ~4,450 (from one randomly selected PCR batch of our large-scale mouse brain experiment and from the following 10x Chromium dataset (Ding et al., 2020)) or ~20,000 (from a separate deep-sequenced dataset we generated and the previously mentioned publicly available 10x Chromium dataset) raw reads/cells to account for the different sequencing depths. The experiment, with the deeply sequenced *EasySci* dataset (~20,000 reads/cells), contained cells from human and mouse cell lines as well. Because these cell lines are known to have higher recovered signals per cell, we pre-selected the cell barcodes corresponding to brain samples from the raw data before subsampling. This pre-selection of expected barcodes removes sequences with non-matching barcodes in the target location of the reads, like primer dimers. To adjust for the enrichment of the signal of interest, we performed the same barcode pre-selection on the 10x Chromium data. After the subsampling, the *EasySci* data was processed with the custom computational pipeline, while the 10x Chromium data was processed with 10x Genomics' Cell Ranger software (Zheng et al., 2017).

Cell clustering and cell type annotation of single-cell RNA-seq data

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 Scrublet (Wolock et al., 2019) computational pipeline to identify and remove potential doublets, similar to our previous study (Cao et al., 2020). At the end of these filtering steps, we had around 1.5 million brain cells in the dataset.

To identify distinct clusters of cells corresponding to different cell types, we subjected the 1,469,111 single-cell gene expression profiles to UMAP visualization and Louvain clustering, similar to our previous study

(Cao et al., 2020). We then co-embedded our data with the published datasets (Kozareva et al., 2021; Yao et al., 2021; Zeisel et al., 2018) through Seurat (Stuart et al., 2019), and clusters were annotated based on overlapped cell types. The annotations were manually verified and refined based on marker genes. Differentially expressed genes across cell types were identified with the differentialGeneTest() function of Monocle 2 (Qiu et al., 2017). To identify cell type-specific gene markers, we selected genes that were differentially expressed across different cell types (FDR of 5%, likelihood) and also with a > 2-fold expression difference between first and second-ranked cell types and TPM > 50 in the first-ranked cell types.

Isoform expression analysis

Isoform expression was quantified in *EasySci* data using an adapted version of the pipeline built by Booeshaghi et al. (Booeshaghi et al., 2021) RandomN-primed reads for ~1.5M single cells were merged into 613 pseudocells, grouping by individual mouse and cell types (31 cell types). The pseudocells were aligned to the mouse transcriptome with *kallisto* (Melsted et al., 2021), generating a raw isoform count matrix. To filter and pre-process the raw data, isoform counts were normalized by length, and genes and isoforms with a dispersion of less than 0.001 were removed. The gene count matrix was produced by aggregating counts of all isoforms of a given gene. Both isoform and gene count matrices were normalized by dividing the counts in each cell by the sum of the counts for that cell, then multiplying by 1,000,000 and transforming with numpy's log1p() function. The filtered data contained 33,361 isoforms corresponding to 12,636 genes. Highly variable isoforms and genes were identified using scanpy, by binning into 20 bins and scaling the dispersion for each feature to zero mean and unit variance within each bin. The top 5,000 gene and isoforms in each matrix were retained based on normalized dispersion. Neighborhood components analysis was performed on the filtered and normalized isoform matrix after scaling the log(1+TPM) expression to zero mean and unit variance, training on cell type labels from each pseudocell with random state 42, and visualized using t-SNE with perplexity 10, 5,000 iterations and random state 42. Differentially expressed isoforms were identified by looking for isoforms that were upregulated across a given cell type, while the genes containing those isoforms were not significantly expressed more among that cell type than its complement (the rest of the dataset). Isoforms expressed in less than 90% of pseudocells within a cell type were discarded. T-tests used a significance level of 0.01 with Bonferroni correction for multiple comparisons.

Sub-cluster analysis of the single-cell RNA-seq data

To identify cell subtypes, we selected each main cell type and applied PCA, UMAP and Louvain clustering similarly to the major cluster analysis, based on a 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. We then merged sub-clusters that were not readily distinguishable in the UMAP space through an intra-dataset cross-validation procedure described before (Cao et al., 2020). A total of 359 cell subtypes were identified, with a median of 1,038 cells in each group. All subtypes were contributed by at least two individuals (median of twenty). Differentially expressed genes and exons across cell types were identified with the differentialGeneTest() function of Monocle 2 (Qiu et al., 2017). To identify sub-cluster-specific differentially expressed 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 package (Qiu et al., 2017). The sex of the animals was included as a covariate to reduce gender-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 sub-cluster in each reverse transcription well profiled by *EasySci-RNA*. Each RT well was regarded as a replicate comprising cells from a specific mouse individual. We then applied the likelihood-ratio test to identify significantly changed sub-clusters between different conditions, with the differentialGeneTest() function of Monocle 2 (Qiu et al., 2017). Sub-clusters were removed if they had less than 20 cells in either the male or female samples. The fold change was calculated manually by first 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 at least a two-fold change between two groups and the q-value was less than 0.05.

Spatial transcriptomic analysis to estimate spatial abundances of cell subtypes

To spatially map *EasySci* cell subtypes, we integrated the *EasySci-RNA* data with a publicly available 10x Visium spatial transcriptomics dataset (Genomics, 2019a, 2019b, 2019c) using cell2location, a Bayesian model designed to map fine-grained cell types. We first aggregated ~50 single-cell transcriptomes identified by k-means clustering (k = 50) of cells in the UMAP space of sub-clustering analysis. The cell2location model first used negative binomial regression to estimate reference cell type signatures from the *EasySci-RNA* data. In a second step, cell2location decomposed spatial mRNA counts from 10x Visium data into the reference signatures to estimate cell type spatial abundances. Training of the model utilized durations of 25 and 15,000 epochs for the negative binomial regression and spatial mapping steps, respectively.

Gene module analysis

We performed gene module analysis to identify the molecular programs underlying different cell types in the brain. First, we aggregated the gene expression across all sub-clusters. The aggregated gene count matrix was then normalized by the library size and then log-transformed (log10(TPM / 10 + 1)). Genes were removed if they exhibited low expression (less than 1 in all sub-clusters) or low variance of expression (*i.e.,* the gene expression fold change between the maximum expressed sub-cluster and the median expression across sub-clusters is less than 5). The filtered matrix was used as input for UMAP/0.3.2 visualization (McInnes et al., 2018) (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) (Rodriguez and Laio, 2014).

***EasySci-ATAC* library preparation and sequencing**

Mouse brain samples were snap-frozen in liquid nitrogen and stored at -80°C. For nuclei extraction, thawed brain samples were minced in PBS using a blade, re-frozen, stored at -80°C, and processed in multiple batches. The detailed step-by-step protocol is included as a supplementary file (**Supplementary file 4**).

Data processing for *EasySci-ATAC*

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 (ED) < 2). Downstream sequence processing was similar to sci-ATAC-seq (Cao et al., 2018). Indexed Tn5 barcodes and ligation barcodes were extracted, corrected to its nearest barcode (edit distance (ED) < 2) and reads with uncorrected barcodes (ED >= 2) were removed. Tn5 adaptors were removed from 5'-end and clipped from 3'-end using trim_galore/0.4.1 (Krueger et al., 2021). Trimmed reads were mapped to the mouse genome (mm39) using STAR/v2.5.2b (Dobin et al., 2013) with default settings. Aligned reads were filtered using samtools/v1.4.1 (Li et al., 2009) to retain reads mapped in proper pairs with quality score MAPQ > 30 and to keep only the primary aligment. Duplicates were removed by picard MarkDuplicates/v2.25.2 (Broad Institute, 2019) per PCR sample. Deduplicated bam files were converted to bedpe format using bedtools/v2.30.0 (Quinlan and Hall, 2010), which were further converted to offset-adjusted (+4 bp for plus strand and -5 bp for minus) fragment files (.bed). Deduplicated reads were further split into constituent cellular indices by further demultiplexing reads using the Tn5 and ligation indexes. For each cell, we also created sparse matrices counting reads falling into promoter regions (±1 kb around TSS) for downstream analysis.

Cell filtering, clustering and annotation for EasySci-ATAC

We used SnapATAC2/v1.99.99.3(Fang et al., 2021; Zhang, 2022) to perform preprocessing steps for the *EasySci-ATAC* dataset. Cells with less than 1500 fragments and less than 2 TSS Enrichment were discarded. Potential doublet cells and doublet-derived subclusters were detected using an iterative clustering strategy (Cao et al., 2020) modified to suit for scATAC-seq data. Briefly, cells were splitted by individual animals to overcome the large memory use when simulating doublets for the full dataset, and doublet scores were calculated using snap.pp.scrublet() (Wolock et al., 2019). Then, all cells were combined, followed by clustering and sub-clustering analysis with spectral embedding and graph-based clustering implemented in SnapATAC2. Cells labeled as doublets (defined by a doublet score cutoff of 0.2) or from doublet-derived sub-clusters (defined by a doublet ratio cutoff of 0.4) were filtered out. In addition, cells with high fragment numbers in each main cluster (defined as cells with fragments number higher than the 95th quantile within the main cluster) were also filtered out. We then generated a gene activity matrix using snap.pp.make_gene_matrix() for the following integration analysis.

We used a deep-learning-based framework scJoint (Lin et al., 2022) to annotate main ATAC-seq cell types 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 (McInnes, 2018). 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 (*i.e.,* less than 80% cells were from the highest abundant cell type) were removed upon inspections. Finally, to validate the integration-based annotations, we selected differentially expressed 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 (transcripts per million) > 20 in the maximum RNA group and RPM (reads per million) > 50 in the maximum ATAC group. Top 10

genes ranked by fold change between the maximum and the second maximum expressed group 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 (gene activity) for each cell type.

Subcluster level integrations were similar to the main cluster level integrations with mild modifications. For astrocytes, microglia, OB neurons 1, OB neurons 2, OB neurons 3 and vascular leptomeningeal cells, we used all cells from the *EasySci-RNA* dataset as input for the integrations. For oligodendrocytes, we subsampled 2,000 cells from each subcluster from the *EasySci-RNA* data for integration analysis. Similarly, we validated the subcluster level integrations by inspecting the aggregated gene activity of subcluster-specific gene markers in the predicted ATAC subclusters. Subcluster marker genes were identified by differential expression analysis using scRNA-seq data and selected by the following criteria: fold change between the maximum expressed sub-cluster and the mean of all the other subclusters within the same main cell type > 2, FDR < 0.05, TPM (transcripts per million) > 50 in the maximum expressed RNA group and RPM (reads per million) > 50 in the maximum accessible ATAC group.

Peak calling, peak-based dimension reduction and identifications of differential accessible peaks

To define peaks of accessibility, we used MACS2/v2.1.1 (Zhang et al., 2008). 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. 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' (Corces et al., 2018) and filtered peaks by choosing a score-per-million cut-off of 1.3. Peak summits were extended by 250bp 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 package (Anders et al., 2015).

We used R package Signac/v1.7.0 (Stuart et al., 2021) 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 (Qiu et al., 2017) 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 (Schep et al., 2017) to access the TF motif accessibility using a collection of the cisBP motif sets curated by chromVARmotifs/v0.2.0 (Weirauch et al., 2014; Schep et al., 2017). To investigate TF regulators at the main cluster level, 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 subcluster 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 (Schep, 2017). 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. ATAC-seq subclusters with less than 20 cells were excluded from the correlation analysis.

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). Briefly, mice were sacrificed, and brains were extracted and frozen with liquid nitrogen. Frozen brain was embedded in OCT (Tissue TEK O.C.T compound) and cryosectioned at -15C (Leica cryostat). Coronally placed brains were cut halfway, to place half coronally sectioned brains at 10um on Visium tissue optimization, or gene expression analysis slides capture areas. User guide CG000160 from 10x Genomics was followed for methanol fixation and H&E stain. After fixation and staining, imaging was performed using Leica DMI8, and images were stitched using Leica Application Suite X and saved into .tiff format. After tissue fixation and staining, Visium Spatial Tissue Optimization User Guide (catalog no. CG000238 Rev A, 10x Genomics) or Visium Spatial Gene Expression User Guide (catalog no. CG000239 Rev A, 10x Genomics) were followed for either protocol optimization, or gene expression analysis, respectively. Tissue optimization was performed according to CG000238 , and according to optimization experiments, 18 min permeabilization provided the most optimal signal, and was followed for gene expression library preparation as well. Libraries were prepared according to Visium Spatial Gene Expression User Guide (CG000239, 10x Genomics)

Library preparation and data processing of spatial transcriptomics

Libraries were sequenced using a NextSeq1000 system. BCL files were converted to FASTQ, and raw FASTQ files and .tiff histology images were processed with spaceranger-1.2.2 software. Spaceranger-1.2.2 uses STAR for RNA reads genome alignment, and utilized the GRCm38 (mouse mm10) as the reference genome provided from 10X Genomics. We performed the downstream visualization and clustering analysis of the spatial transcriptomic data following the tutorial of Seurat (Stuart et al., 2019) (https://satijalab.org/seurat/articles/spatial_vignette.html) with default parameters.

Spatial transcriptomic analysis 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 dataset (https://satijalab.org/seurat/articles/spatial_vignette.html (Genomics, 2019a, 2019b, 2019c) through a non-negative least squares (NNLS) approach modified from our previous study (Cao et al., 2020). We first aggregated cell-type-specific UMI counts, normalized by the library size, multiplied by 100,000, and log-transformed after adding a pseudo-count. A similar procedure was applied to calculate the normalized gene expression in each spatial spot captured in the 10x Visium dataset. We then applied non-negative least squares (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:

$$T_a = \beta_{0a} + \beta_{1a} M_b$$

where $T_a$and M_b represent filtered gene expression for target spatial spot from 10x Visium dataset A and all cell types from *EasySci-RNA* dataset B, respectively. To improve accuracy and specificity, we selected cell type-specific genes for each target cell type by: 1) ranking genes based on the expression fold-change between the target cell type vs. the median expression across all cell types, and then selecting the top 200 genes. 2) ranking genes based on the expression fold-change between the target cell type vs. the cell type with maximum expression among all other cell types, and then selecting the top 200 genes. 3) merging the gene lists from step (1) and (2). $\beta_{1a}$is the correlation coefficient computed by NNLS regression.

Similarly, we then switch the order of datasets A and B, and predict the gene expression of target cell type (T_b) in dataset B with the gene expression of all spatial spots (M_a) in dataset A:

$$T_b = \beta_{0b} + \beta_{1b} M_a$$

Thus, each spatial spot a in 10x Visium dataset A and each cell type b in *EasySci* dataset B are linked by two correlation coefficients from the above analysis: β_{ab} for predicting the gene expression in each spatial spot a using b, and $\beta_{ba}$for predicting gene expression in each cell type b using a. We combine the two values by:

$$\beta = (\beta_{ab} + 0.01) * (\beta_{ba} + 0.01)$$

The β is then capped to [1, 3]. We find β reflects the cell-type-specific abundance across different spatial spots in 10x Visium datasets with high specificity. We thus use β as the alpha value (*i.e.,* the opacity of a geom) to plot the spatial distribution of different cell types.

To characterize the expression of sub-cluster specific gene markers, we first normalized the gene expression in each spatial spot of 10x Visium data by the library size, multiplied by 100,000, and log-transformed after adding a pseudo-count. The expression of genes from sub-cluster specific gene markers was aggregated, scaled to z-score and capped to [3, 6]. Of note, the sub-cluster specific gene markers were selected by differentiation expression analysis described above and only DE genes (FDR of 5%, with a >2-fold expression difference between first and second ranked sub-clusters, expression TPM > 50 in at least one sub-cluster) were selected as gene markers. In addition, we examined the aggregated

expression of the selected gene markers across all 359 sub-clusters to further validate the specificity of gene markers for labeling target sub-clusters.

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.

Differentially expressed genes between AD and control samples for each cell type in each region were identified using Monocle 2 (Qiu et al., 2017) with the differentialGeneTest() function. Main cell types with less than 50 cells were excluded from the analysis (i.e, 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 FC > 1.5 between the maximum and second expressed condition, and with transcripts per million (TPM) > 50 in the highest expressed condition. To further validate human-mouse shared gene expression changes, we used a recently published Alzheimer's disease single-cell dataset from the human prefrontal cortex (Morabito et al., 2021).

Code Availability

The detailed experimental protocol and computation scripts of *EasySci* were included as supplementary files.

Supplementary Tables (provided as Microsoft Excel files)

Supplementary Table 1: Differentially expressed genes across main cell types. For each gene, the "Cell type" is the cell type with the highest expression, with the expression level quantified by transcripts per million in "TPM in cell type". The "Q-value" is the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the differential expression test across different cell clusters. The "Fold change" is the fold change between the max expressed cell type and the second expressed cell type.

Supplementary Table 2: Differentially expressed isoforms across main cell types. For each isoform ("Isoform"), the "Cell type" is the cell type with the highest expression. The "P-value" is the raw p-value for the differential expression test across different cell types; and the "Q-value" is the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons). The "Effect size" is the effect size between the max expressed cell type and the second expressed cell type.

Supplementary Table 3: Differentially accessible sites for main cell types. For each peak ("Peak"), the "Max cell type" is the cell type with the highest accessibility ("Peak accessibility in max cell type"). The "Second cell type" is the cell type with the second highest accessibility ("Peak accessibility in second cell type"). The "Fold change" is the fold change between the max accessibility and the second max accessibility. The "P-value" is the raw p-value for the differential accessibility test across different cell types,

and the "Q-value" is the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons).

Supplementary Table 4: Differentially expressed exons across sub-clusters within each main cell type. For each sub-cluster ("Cell sub-cluster ID"), the following features of marker genes are listed: gene symbol ("Gene name"); Ensembl ID of the gene and the exon ("Exon ID"); false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the differential expression test across different cell sub-clusters within each main cell type ("Q-value"); fold change of the marker exon expression between the max and second expressed cell sub-cluster ("Fold change"); expression level of the marker exon quantified by transcripts per million in max sub-cluster ("TPM in max sub-cluster"). Marker exons are defined by Q-value < 0.05, Fold change > 2 and TPM in max subcluster > 50.

Supplementary Table 5: Gene module analysis results. For each gene module ("Gene module ID"), the following information about the genes belonging to that gene module is listed: Ensembl ID ("Gene ID"); type of gene ("Gene type"); gene symbol ("Gene name"); UMAP visualization coordinates of the genes based on their expression variance across all 359 cell sub-clusters ("UMAP 1", "UMAP 2").

Supplementary Table 6: Differentially abundant sub-clusters between adult and young samples. Sub-clusters ("Cell sub-clusters") abundances were compared between the 3 vs 6 months old groups ("Condition"), and the following statistical values are listed: false detection rate (likelihood ratio test with adjustment for multiple comparisons) for the differential abundance test across age groups ("Q-value"); log2 fold change of the cell sub-cluster abundance between the age groups ("Log2(Fold change)"); the number of cells compared in the sub-cluster ("Number of cells"); whether the sub-cluster is upregulated, downregulated or there is no significant change ("Final change", significance was determined by Fold change > 2, Q-value < 0.05 and more than 20 cells in both male and female samples).

Supplementary Table 7: Differentially abundant sub-clusters between aged and adult samples. Sub-clusters ("Cell sub-clusters") abundances were compared between the 6 vs 21 months old groups ("Condition"), and the following statistical values are listed: false detection rate (likelihood ratio test with adjustment for multiple comparisons) for the differential abundance test across age groups ("Q-value"); log2 fold change of the cell sub-cluster abundance between the age groups ("Log2(Fold change)"); the number of cells compared in the sub-cluster ("Number of cells"); whether the sub-cluster is upregulated, downregulated or there is no significant change ("Final change", significance was determined by Fold change > 2, Q-value < 0.05 and more than 20 cells in both male and female samples).

Supplementary Table 8: Differentially expressed genes between aged and adult for all sub-clusters. For each subcluster in the dataset ("Cell subcluster"), the following information is listed: gene symbol ("Gene name"); gene Ensembl ID ("Gene ID"); the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the differential expression test across the age groups ("Q-value"); fold change between the max expressed age group and second expressed age group ("Fold change"); expression level quantified by transcripts per million in the max age group ("TPM in max condition"); the age group where the max expression was detected ("Max condition"). Only significant genes are listed in the table, which is defined by Q-value < 0.05, Fold change > 2 and TPM in max condition > 50.

Supplementary Table 9: Differentially abundant sub-clusters between wild-type and EOAD model. Sub-clusters ("Cell sub-clusters") abundances were compared between the wild-type and EOAD model (5xFAD) groups ("Condition"), and the following statistical values are listed: false detection rate (likelihood ratio test with adjustment for multiple comparisons) for the differential abundance test between conditions ("Q-value"); log2 fold change of the cell sub-cluster abundance between conditions ("Log2(Fold change)"); the number of cells compared in the sub-cluster ("Number of cells"); whether the sub-cluster is upregulated, downregulated or there is no significant change ("Final change", significance was determined by Fold change > 2, Q-value < 0.05 and more than 20 cells in both male and female samples).

Supplementary Table 10: Differentially abundant sub-clusters between wild-type and LOAD model. Sub-clusters ("Cell sub-clusters") abundances were compared between the wild-type and LOAD model (APOE*4/Trem2*R47H) groups ("Condition"), and the following statistical values are listed: false detection rate (likelihood ratio test with adjustment for multiple comparisons) for the differential abundance test between conditions ("Q-value"); log2 fold change of the cell sub-cluster abundance between conditions ("Log2(Fold change)"); the number of cells compared in the sub-cluster ("Number of cells"); whether the sub-cluster is upregulated, downregulated or there is no significant change ("Final change", significance was determined by Fold change > 2, Q-value < 0.05 and more than 20 cells in both male and female samples).

Supplementary Table 11: Differentially expressed genes between wild-type and EOAD model (5xFAD) for all sub-clusters. For each subcluster in the dataset ("Cell subcluster"), the following information is listed: gene symbol ("Gene name"); gene Ensembl ID ("Gene ID"); the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the differential expression test across conditions ("Q-value"); fold change between the max expressed group and second expressed group ("Fold change"); expression level quantified by transcripts per million in the max group ("TPM in max condition"); the condition where the max expression was detected ("Max condition"). Only significant genes are listed in the table, which is defined by Q-value < 0.05, Fold change > 2 and TPM in max condition > 50.

Supplementary Table 12: Differentially expressed genes between wild-type and LOAD model (APOE*4/Trem2*R47H) for all sub-clusters. For each subcluster in the dataset ("Cell subcluster"), the following information is listed: gene symbol ("Gene name"); gene Ensembl ID ("Gene ID"); the false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the differential expression test across conditions ("Q-value"); fold change between the max expressed group and second expressed group ("Fold change"); expression level quantified by transcripts per million in the max group ("TPM in max condition"); the condition where the max expression was detected ("Max condition"). Only significant genes are listed in the table, which is defined by Q-value < 0.05, Fold change > 2 and TPM in max condition > 50.

Supplementary Table 13: Metadata of human brain samples included in this study.

Supplementary Table 14: Differentially expressed genes between control and AD human brain samples for each main cell type in each region. For each main cell type ("Main cluster name") in each region ("Region"), the following information is listed: gene symbol ("Gene name"); the max and the second expressed group ("Max condition", "Second condition") along with expression level quantified by transcripts per million ("TPM in max condition", "TPM in second condition") and the fold change ("Fold change"); the

false detection rate (one-sided likelihood ratio test with adjustment for multiple comparisons) for the differential expression test across the two conditions ("Q-value"). Only significant genes are listed in the table, which is defined by Q-value < 0.05, Fold change > 1.5 and TPM in max condition > 50.

Supplementary files

Supplementary file 1: Detailed experiment protocols for *EasySci-RNA*, including all materials and equipment needed, step-by-step descriptions, and representative gel images.

Supplementary file 2: Primer sequences used in the *EasySci-RNA* experiment, including multiple plates of short dT RT primers, random hexamer RT primers, ligation primers and P7 PCR primers. The columns indicate the positions on the 96-well plate (Well position), an identifier of the sequence (Name), the full primer sequence (Sequence) and the barcode sequence (Barcode).

Supplementary file 3: Computational pipeline scripts and notes for processing *EasySci-RNA* data, from sequencer-generated files to single-cell gene count matrix.

Supplementary file 4: Detailed experiment protocols for *EasySci-ATAC*, including all materials and equipment needed, step-by-step descriptions, and representative gel images.

Supplementary file 5: Primer sequences used in the *EasySci-ATAC* experiment, including N5/N7 oligos used in indexed Tn5 assembly, ligation primers and P7 PCR primers. The columns indicate the positions on the 96-well plate (Well position), an identifier of the sequence (Name), the full primer sequence (Sequence) and the barcode sequence (Barcode).

Supplementary file 6: Computational pipeline scripts and notes for processing *EasySci-ATAC* data, from sequencer-generated files to single-cell read files.

Reference:

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Arawaka, S., Machiya, Y., and Kato, T. (2010). Heat shock proteins as suppressors of accumulation of toxic prefibrillar intermediates and misfolded proteins in neurodegenerative diseases. Curr. Pharm. Biotechnol. *11*, 158–166.

Attwood, M.M., and Schiöth, H.B. (2021). Characterization of Five Transmembrane Proteins: With Focus on the Tweety, Sideroflexin, and YIP1 Domain Families. Front Cell Dev Biol *9*, 708754.

Baranger, K., Marchalant, Y., Bonnet, A.E., Crouzin, N., Carrete, A., Paumier, J.-M., Py, N.A., Bernard, A., Bauer, C., Charrat, E., et al. (2016). 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.

Barraud, P., Seferiadis, A.A., Tyson, L.D., Zwart, M.F., Szabo-Rogers, H.L., Ruhrberg, C., Liu, K.J., and Baker, C.V.H. (2010). Neural crest origin of olfactory ensheathing glia. Proc. Natl. Acad. Sci. U. S. A. *107*, 21040–21045.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment *2008*, P10008. https://doi.org/10.1088/1742-5468/2008/10/p10008.

Bommer, G.T., and MacDougald, O.A. (2011). Regulation of lipid homeostasis by the bifunctional SREBF2-miR33a locus. Cell Metab. *13*, 241–247.

Booeshaghi, A.S., Yao, Z., van Velthoven, C., Smith, K., Tasic, B., Zeng, H., and Pachter, L. (2021). Isoform cell-type specificity in the mouse primary motor cortex. Nature *598*, 195–199. .

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Broad Institute (2019). Picard toolkit. Broad Institute, GitHub Repository.

Campbell, J.N., Macosko, E.Z., Fenselau, H., Pers, T.H., Lyubetskaya, A., Tenen, D., Goldman, M., Verstegen, A.M.J., Resch, J.M., McCarroll, S.A., et al. (2017). A molecular census of arcuate hypothalamus and median eminence cell types. Nat. Neurosci. *20*, 484–496.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science *357*, 661–667.

Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., Daza, R.M., McFaline-Figueroa, J.L., Packer, J.S., Christiansen, L., et al. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science *361*, 1380–1385.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502.

Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A., Aldinger, K.A., Blecher-Gonen, R., Zhang, F., et al. (2020). A human cell atlas of fetal gene expression. Science *370*.

https://doi.org/10.1126/science.aba7721.

Cavalcante, G.C., Brito, L.M., Schaan, A.P., Ribeiro-Dos-Santos, Â., de Araújo, G.S., and On Behalf Of Alzheimer's Disease Neuroimaging Initiative (2022). Mitochondrial Genetics Reinforces Multiple Layers of Interaction in Alzheimer's Disease. Biomedicines *10*. https://doi.org/10.3390/biomedicines10040880.

Cooper, W.N., Hesson, L.B., Matallanas, D., Dallol, A., von Kriegsheim, A., Ward, R., Kolch, W., and Latif, F. (2009). RASSF2 associates with and stabilizes the proapoptotic kinase MST2. Oncogene *28*, 2988–2998.

Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science *362*. https://doi.org/10.1126/science.aav1898.

Cornejo, V.H., and Hetz, C. (2013). The unfolded protein response in Alzheimer's disease. Semin. Immunopathol. *35*, 277–292.

Cusanovich, D.A., Hill, A.J., Aghamirzaie, D., Daza, R.M., Pliner, H.A., Berletch, J.B., Filippova, G.N., Huang, X., Christiansen, L., DeWitt, W.S., et al. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. Cell *174*, 1309–1324.e18.

Desimone, A., Hong, J., Brockie, S.T., Yu, W., Laliberte, A.M., and Fehlings, M.G. (2021). The influence of ApoE4 on the clinical outcomes and pathophysiology of degenerative cervical myelopathy. JCI Insight *6*. https://doi.org/10.1172/jci.insight.149227.

Di Giorgio, E., Paluvai, H., Dalla, E., Ranzino, L., Renzini, A., Moresi, V., Minisini, M., Picco, R., and Brancolini, C. (2021). HDAC4 degradation during senescence unleashes an epigenetic program driven by AP-1/p300 at selected enhancers and super-enhancers. Genome Biol. *22*, 129.

Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. *38*, 737–746.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Domcke, S., Hill, A.J., Daza, R.M., Cao, J., O'Day, D.R., Pliner, H.A., Aldinger, K.A., Pokholok, D., Zhang, F., Milbank, J.H., et al. (2020). A human cell atlas of fetal chromatin accessibility. Science *370*. https://doi.org/10.1126/science.aba7612.

Erö, C., Gewaltig, M.-O., Keller, D., and Markram, H. (2018). A Cell Atlas for the Mouse Brain. Front. Neuroinform. *12*, 84.

Fang, R., Preissl, S., Li, Y., Hou, X., Lucero, J., Wang, X., Motamedi, A., Shiau, A.K., Zhou, X., Xie, F., et al. (2021). Comprehensive analysis of single cell ATAC-seq data with SnapATAC. Nat. Commun. *12*, 1337.

de Faria, O., Jr, Dhaunchak, A.S., Kamen, Y., Roth, A.D., Kuhlmann, T., Colman, D.R., and Kennedy, T.E. (2019). TMEM10 Promotes Oligodendrocyte Differentiation and is Expressed by Oligodendrocytes in Human Remyelinating Multiple Sclerosis Plaques. Sci. Rep. *9*, 3606.

Genomics, 10x (2019a). Mouse Brain Section (Coronal).

Genomics, 10x (2019b). Mouse Brain Serial Section 1 (Sagittal-Posterior).

Genomics, 10x (2019c). Mouse Brain Serial Section 1 (Sagittal-Anterior).

Graham, V., Khudyakov, J., Ellis, P., and Pevny, L. (2003). SOX2 functions to maintain neural progenitor identity. Neuron *39*, 749–765.

Hashimoto, Y., Niikura, T., Tajima, H., Yasukawa, T., Sudo, H., Ito, Y., Kita, Y., Kawasumi, M., Kouyama, K., Doyu, M., et al. (2001). A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and Abeta. Proc. Natl. Acad. Sci. U. S. A. *98*, 6336–6341.

Herring, A., Donath, A., Steiner, K.M., Widera, M.P., Hamzehian, S., Kanakis, D., Kölble, K., ElAli, A., Hermann, D.M., Paulus, W., et al. (2012). Reelin depletion is an early phenomenon of Alzheimer's pathology. J. Alzheimers. Dis. *30*, 963–979.

Hüttenrauch, M., Ogorek, I., Klafki, H., Otto, M., Stadelmann, C., Weggen, S., Wiltfang, J., and Wirths, O. (2018). Glycoprotein NMB: a novel Alzheimer's disease associated marker expressed in a subset of activated microglia. Acta Neuropathol Commun *6*, 108.

Karch, C.M., and Goate, A.M. (2015). Alzheimer's disease risk genes and mechanisms of disease pathogenesis. Biol. Psychiatry *77*, 43–51.

Kenigsbuch, M., Bost, P., Halevi, S., Chang, Y., Chen, S., Ma, Q., Hajbi, R., Schwikowski, B., Bodenmiller, B., Fu, H., et al. (2022). A shared disease-associated oligodendrocyte signature among multiple CNS pathologies. Nat. Neurosci. *25*, 876–886.

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T.K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., et al. (2017). A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. Cell *169*, 1276–1290.e17. https://doi.org/10.1016/j.cell.2017.05.018.

Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H.W., Li, T., Elmentaite, R., Lomakin, A., Kedlian, V., Gayoso, A., et al. (2022). Cell2location maps fine-grained cell types in spatial transcriptomics. Nat. Biotechnol. 1–11.

Kotredes, K.P., Oblak, A., Pandey, R.S., Lin, P.B.-C., Garceau, D., Williams, H., Uyar, A., O'Rourke, R., O'Rourke, S., Ingraham, C., et al. (2021). Uncovering Disease Mechanisms in a Novel Mouse Model Expressing Humanized APOEε4 and Trem2*R47H. Front. Aging Neurosci. *13*, 735524.

Kozareva, V., Martin, C., Osorno, T., Rudolph, S., Guo, C., Vanderburg, C., Nadaf, N., Regev, A., Regehr, W.G., and Macosko, E. (2021). A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. Nature *598*, 214–219.

Krueger, F., James, F., Ewels, P., Afyounian, E., and Schuster-Boeckler, B. (2021). GitHub - FelixKrueger/TrimGalore: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data. https://doi.org/10.5281/zenodo.5127899.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

Kulijewicz-Nawrot, M., Syková, E., Chvátal, A., Verkhratsky, A., and Rodríguez, J.J. (2013). Astrocytes and glutamate homoeostasis in Alzheimer's disease: a decrease in glutamine synthetase, but not in glutamate transporter-1, in the prefrontal cortex. ASN Neuro *5*, 273–282.

Kumar, A., Pareek, V., Faiq, M.A., Kumar, P., Kumari, C., Singh, H.N., and Ghosh, S.K. (2020). Transcriptomic analysis of the signature of neurogenesis in human hippocampus suggests restricted progenitor cell progression post-childhood. IBRO Rep *9*, 224–232.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, J., Wang, C., Zhang, Z., Wen, Y., An, L., Liang, Q., Xu, Z., Wei, S., Li, W., Guo, T., et al. (2018). Transcription Factors Sp8 and Sp9 Coordinately Regulate Olfactory Bulb Interneuron Development. Cereb. Cortex *28*, 3278–3294.

Lin, Y., Wu, T.-Y., Wan, S., Yang, J.Y.H., Wong, W.H., and Wang, Y.X.R. (2022). scJoint integrates atlas-scale single-cell RNA-seq and ATAC-seq data with transfer learning. Nat. Biotechnol. *40*, 703–710.

Liu, J., Lin, C., Gleiberman, A., Ohgi, K.A., Herman, T., Huang, H.P., Tsai, M.J., and Rosenfeld, M.G. (2001). Tbx19, a tissue-selective regulator of POMC gene expression. Proc. Natl. Acad. Sci. U. S. A. *98*, 8674–8679.

Liu, X.-F., Hao, J.-L., Xie, T., Malik, T.H., Lu, C.-B., Liu, C., Shu, C., Lu, C.-W., and Zhou, D.-D. (2017). Nrf2 as a target for prevention of age-related and diabetic cataracts by against oxidative stress. Aging Cell *16*, 934–942.

Luo, J., Huang, X., Li, R., Xie, J., Chen, L., Zou, C., Pei, Z., Mao, Y., and Zou, D. (2022). PTPRG activates m6A methyltransferase VIRMA to block mitochondrial autophagy mediated neuronal death in Alzheimer's disease.

Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T., et al. (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell https://doi.org/10.1016/j.cell.2020.09.056.

Marques, S., van Bruggen, D., Vanichkina, D.P., Floriddia, E.M., Munguba, H., Väremo, L., Giacomello, S., Falcão, A.M., Meijer, M., Björklund, Å.K., et al. (2018). Transcriptional Convergence of Oligodendrocyte Lineage Progenitors during Development. Dev. Cell *46*, 504–517.e7.

Martin, B.K., Qiu, C., Nichols, E., Phung, M., Green-Gladden, R., Srivatsan, S., Blecher-Gonen, R., Beliveau, B.J., Trapnell, C., Cao, J., et al. (2021). An optimized protocol for single cell transcriptional profiling by combinatorial indexing. https://doi.org/10.48550/arXiv.2110.15400.

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. Nature *570*, 332–337.

Mays, J.C., Kelly, M.C., Coon, S.L., Holtzclaw, L., Rath, M.F., Kelley, M.W., and Klein, D.C. (2018). 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.

McEvilly, R.J., Erkman, L., Luo, L., Sawchenko, P.E., Ryan, A.F., and Rosenfeld, M.G. (1996). Requirement for Brn-3.0 in differentiation and survival of sensory and motor neurons. Nature *384*, 574–577.

McInnes, L. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction — umap 0.5 documentation, https://umap-learn.readthedocs.io/en/latest/.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software *3*, 861. https://doi.org/10.21105/joss.00861.

Melsted, P., Sina Booeshaghi, A., Liu, L., Gao, F., Lu, L., Min, K.H. (joseph), da Veiga Beltrame, E., Hjörleifsson, K.E., Gehring, J., and Pachter, L. (2021). Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat. Biotechnol. 1–6.

Mielke, M.M., and Lyketsos, C.G. (2010). Alterations of the sphingolipid pathway in Alzheimer's disease: new biomarkers and treatment targets? Neuromolecular Med. *12*, 331–340.

Monavarfeshani, A., Knill, C.N., Sabbagh, U., Su, J., and Fox, M.A. (2017). Region- and Cell-Specific Expression of Transmembrane Collagens in Mouse Brain. Front. Integr. Neurosci. *11*, 20.

Morabito, S., Miyoshi, E., Michael, N., Shahin, S., Martini, A.C., Head, E., Silva, J., Leavy, K., Perez-Rosendahl, M., and Swarup, V. (2021). Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. Nature Genetics *53*, 1143–1155. https://doi.org/10.1038/s41588-021-00894-z.

Neff, R.A., Wang, M., Vatansever, S., Guo, L., Ming, C., Wang, Q., Wang, E., Horgusluoglu-Moloch, E., Song, W.-M., Li, A., et al. (2021). Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets. Sci Adv *7*. https://doi.org/10.1126/sciadv.abb5398.

Niccolini, F., Haider, S., Reis Marques, T., Muhlert, N., Tziortzi, A.C., Searle, G.E., Natesan, S., Piccini, P., Kapur, S., Rabiner, E.A., et al. (2015a). Altered PDE10A expression detectable early before symptomatic onset in Huntington's disease. Brain *138*, 3016–3029.

Niccolini, F., Foltynie, T., Reis Marques, T., Muhlert, N., Tziortzi, A.C., Searle, G.E., Natesan, S., Kapur, S., Rabiner, E.A., Gunn, R.N., et al. (2015b). Loss of phosphodiesterase 10A expression is associated with progression and severity in Parkinson's disease. Brain *138*, 3003–3015.

Oakley, H., Cole, S.L., Logan, S., Maus, E., Shao, P., Craft, J., Guillozet-Bongaarts, A., Ohno, M., Disterhoft, J., Van Eldik, L., et al. (2006). 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.

Oeckinghaus, A., and Ghosh, S. (2009). The NF-kappaB family of transcription factors and its regulation. Cold Spring Harb. Perspect. Biol. *1*, a000034.

Paillasse, M.R., and de Medina, P. (2015). The NR4A nuclear receptors as potential targets for anti-aging interventions. Med. Hypotheses *84*, 135–140.

Pastrana, E., Cheng, L.-C., and Doetsch, F. (2009). Simultaneous prospective purification of adult subventricular zone neural stem cells and their progeny. Proc. Natl. Acad. Sci. U. S. A. *106*, 6387–6392.

Puverel, S., Nakatani, H., Parras, C., and Soussi-Yanicostas, N. (2009). Prokineticin receptor 2 expression identifies migrating neuroblasts and their subventricular zone transient-amplifying progenitors in adult mice. J. Comp. Neurol. *512*, 232–242.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rodriguez, A., and Laio, A. (2014). Machine learning. Clustering by fast search and find of density peaks. Science *344*, 1492–1496.

Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Graybuck, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science *360*, 176–182.

Schep, A. (2017). motifmatchr: Fast Motif Matching in R, https://github.com/GreenleafLab/motifmatchr/.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978.

Schmidt, L.S., and Linehan, W.M. (2018). FLCN: The causative gene for Birt-Hogg-Dubé syndrome. Gene *640*, 28–42.

See, A.P., Han, J.E., Phallen, J., Binder, Z., Gallia, G., Pan, F., Jinasena, D., Jackson, C., Belcaid, Z., Jeong, S.J., et al. (2012). The role of STAT3 activation in modulating the immune microenvironment of GBM. J. Neurooncol. *110*, 359–368.

Seripa, D., Matera, M.G., Franceschi, M., Daniele, A., Bizzarro, A., Rinaldi, M., Panza, F., Fazio, V.M., Gravina, C., D'Onofrio, G., et al. (2008). The RELN locus in Alzheimer's disease. J. Alzheimers. Dis. *14*, 335–344.

Silva, I., Silva, J., Ferreira, R., and Trigo, D. (2021). Glymphatic system, AQP4, and their implications in Alzheimer's disease. Neurol Res Pract *3*, 5.

Sokolowski, J.D., Nobles, S.L., Heffron, D.S., Park, D., Ravichandran, K.S., and Mandell, J.W. (2011). 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.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat. Methods *18*, 1333–1341.

Tabula Muris Consortium (2020). A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. Nature *583*, 590–595.

Tepe, B., Hill, M.C., Pekarek, B.T., Hunt, P.J., Martin, T.J., Martin, J.F., and Arenkiel, B.R. (2018). 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.e3.

Tong, Y., Xu, Y., Scearce-Levie, K., Ptácek, L.J., and Fu, Y.-H. (2010). COL25A1 triggers and promotes Alzheimer's disease-like pathology in vivo. Neurogenetics *11*, 41–52.

Tufo, C., Poopalasundaram, S., Dorrego-Rivas, A., Ford, M.C., Graham, A., and Grubb, M.S. (2022). Development of the mammalian main olfactory bulb. Development *149*. https://doi.org/10.1242/dev.200210.

Volkert, M.R., and Crowley, D.J. (2020). Preventing Neurodegeneration by Controlling Oxidative Stress: The Role of OXR1. Front. Neurosci. *14*, 611904.

1664 Wang, R., Zhang, P., Wang, J., Ma, L., E, W., Suo, S., Jiang, M., Li, J., Chen, H., Sun, H., et al. (2022).
1665 Construction of a cross-species cell landscape at single-cell level. Nucleic Acids Res.
1666 https://doi.org/10.1093/nar/gkac633.

1667 Watakabe, A., Ichinohe, N., Ohsawa, S., Hashikawa, T., Komatsu, Y., Rockland, K.S., and Yamamori, T.
1668 (2007). Comparative analysis of layer-specific genes in Mammalian neocortex. Cereb. Cortex *17*, 1918–
1669 1933.

1670 Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S.,
1671 Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription
1672 factor sequence specificity. Cell *158*, 1431–1443.

1673 Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in
1674 Single-Cell Transcriptomic Data. Cell Syst *8*, 281–291.e9.

1675 Xia, X., Jiang, Q., McDermott, J., and Han, J.-D.J. (2018). Aging and Alzheimer's disease: Comparison
1676 and associations from molecular to system level. Aging Cell *17*, e12802.

1677 Xiang, X., Piers, T.M., Wefers, B., Zhu, K., Mallach, A., Brunner, B., Kleinberger, G., Song, W., Colonna,
1678 M., Herms, J., et al. (2018). The Trem2 R47H Alzheimer's risk variant impairs splicing and reduces
1679 Trem2 mRNA and protein in mice but not in humans. Mol. Neurodegener. *13*, 49.

1680 Ximerakis, M., Lipnick, S.L., Innes, B.T., Simmons, S.K., Adiconis, X., Dionne, D., Mayweather, B.A.,
1681 Nguyen, L., Niziolek, Z., Ozek, C., et al. (2019). Single-cell transcriptomic profiling of the aging mouse
1682 brain. Nat. Neurosci. *22*, 1696–1708.

1683 Yao, Z., Liu, H., Xie, F., Fischer, S., Adkins, R.S., Aldridge, A.I., Ament, S.A., Bartlett, A., Behrens, M.M.,
1684 Van den Berge, K., et al. (2021). A transcriptomic and epigenomic cell atlas of the mouse primary motor
1685 cortex. Nature *598*, 103–110.

1686 Yeh, H., and Ikezu, T. (2019). Transcriptional and Epigenetic Regulation of Microglia in Health and
1687 Disease. Trends Mol. Med. *25*, 96–111.

1688 Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M.,
1689 Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse Nervous System.
1690 Cell *174*, 999–1014.e22.

1691 Zhang, K. (2022). SnapATAC2: Single Nucleus Analysis Pipeline for ATAC-seq — SnapATAC2 2.0.0
1692 documentation, https://kzhang.org/SnapATAC2/index.html.

1693 Zhang, Y., and Wang, P. (2018). Age-Related Increase of Insulin-Degrading Enzyme Is Inversely
1694 Correlated with Cognitive Function in APPswe/PS1dE9 Mice. Med. Sci. Monit. *24*, 2446–2455.

1695 Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers,
1696 R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*,
1697 R137.

1698 Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri,
1699 P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of
1700 glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. *34*, 11929–11947.

1701 Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler,
1702 T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single
1703 cells. Nat. Commun. *8*, 14049.

1704 Zhou, Y., Song, W.M., Andhey, P.S., Swain, A., Levy, T., Miller, K.R., Poliani, P.L., Cominelli, M., Grover,
1705 S., Gilfillan, S., et al. (2020). Human and mouse single-nucleus transcriptomics reveal TREM2-
1706 dependent and TREM2-independent cellular responses in Alzheimer's disease. Nat. Med. *26*, 131–142.

1707 Zipfel, P., Rochais, C., Baranger, K., Rivera, S., and Dallemagne, P. (2020). Matrix Metalloproteinases
1708 as New Targets in Alzheimer's Disease: Opportunities and Challenges. J. Med. Chem. *63*, 10705–
1709 10725.