scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Human neuronal differentiation under Aβ exposure: a single-cell transcriptomic and epigenomic dataset

Idoia Blanco-Luquin[1,15] ✉, Xabier Martínez-de-Morentin[2,15] ✉, Amaia Vilas-Zornoza[3,4,5], Daniel Mouzo[6], Alejandro Lumbreras Lopez[1], Mónica Macías[1], Diego Alignani[4,5], Alberto Maillo[2], Ana Cecilia Gonzalez Alvarez[2], Leena Ali Ibrahim[2], Vincenzo Lagani[2,7], Natalia Ramírez[8], Jesper Tegner[2,9,10,11], Felipe Prósper[3,4,5,12,13], Maite Mendioroz[1,14,16] ✉ & David Gomez-Cabrero[2,16] ✉

Human induced pluripotent stem cell–derived neural progenitor cells (NPCs) provide a controlled *in-vitro* model for studying early stages of neuronal differentiation. Here, we present a multimodal single-cell dataset generated from NPCs differentiated over four time points (days 0, 7, 13 and 20) under baseline conditions and following exposure to amyloid-β (Aβ) 1–42 peptide. The dataset comprises paired single-cell RNA sequencing (scRNA-seq) and single-cell ATAC sequencing (scATAC-seq) profiles, enabling joint characterization of transcriptional and chromatin accessibility dynamics during differentiation. Following stringent quality control, we analyzed 42575 and 30192 high-quality cells for RNA and ATAC respectively, providing an in-depth characterization of cell composition, molecular signaling pathways, functional properties, and gene regulatory network annotations. To facilitate reuse and benchmarking, transcriptional signatures derived from this dataset were additionally compared with a publicly available human hippocampal bulk RNA-seq dataset. This resource provides a reference multimodal dataset for human NPC-derived neuronal differentiation and supports a broad range of single-cell, multimodal integration, and regulatory network analyses.

[1]Neuroepigenetics Unit-Navarrabiomed, Hospital Universitario de Navarra (HUN), Universidad Pública de Navarra (UPNA), IdiSNA (Navarra Institute for Health Research), C/ Irunlarrea, 3, Pamplona, Navarra, 31008, Spain. [2]Division of Biomedical Sciences, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955, Saudi Arabia. [3]Advanced Genomics Laboratory, Program of Hemato-Oncology, Center for Applied Medical Research (CIMA), University of Navarra, Pamplona, Spain. [4]Hemato-Oncology Program, Center for Applied Medical Research (CIMA), IDISNA, University of Navarra, Pamplona, Spain. [5]Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Pamplona, Spain. [6]Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, UK. [7]SDAIA-KAUST Center of Excellence in Data Science and Artificial Intelligence, 4700 Thuwal, Jeddah, 23952, Saudi Arabia. [8]Oncohematology Research Group, Navarrabiomed, University Hospital of Navarra, Public University of Navarra, Navarra Medical Research Institute (IdiSNA), Pamplona, Spain. [9]Unit of Computational Medicine, Department of Medicine, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital, L8:05, SE-171 76, Stockholm, Sweden. [10]Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia. [11]Science for Life Laboratory, Tomtebodavagen 23A, SE-17165, Solna, Sweden. [12]Hematology Department, Clínica Universidad de Navarra, University of Navarra, Pamplona, Spain. [13]Program of Regenerative Medicine, Centre for Applied Medical Research (CIMA), University of Navarra, Pamplona, 31008, Spain; Instituto de Investigación Sanitaria de Navarra (IdiSNA), Pamplona, 31008, Spain. [14]Department of Neurology, Hospital Universitario de Navarra (HUN) - IdiSNA (Navarra Institute for Health Research), C/ Irunlarrea, 3, Pamplona, Navarra, 31008, Spain. [15]These authors contributed equally: Idoia Blanco-Luquin, Xabier Martínez-de-Morentin. [16]These authors jointly supervised this work: Maite Mendioroz, David Gomez-Cabrero. ✉e-mail: iblancol@navarra.es; xabier.iribarren@kaust.edu.sa; tmendioi@navarra.es; david.gomezcabrero@kaust.edu.sa

Background & Summary

Alzheimer's disease (AD) is the most common cause of age-related dementia and is characterized pathologically by amyloid β (Aβ) accumulation, tau pathology, and progressive neurodegeneration[1]. The hippocampus is among the regions most severely affected during disease progression, and both human and animal studies have investigated how neurogenic processes in this region may be altered in disease contexts[2–5]. In the adult mammalian brain, new neurons continue to arise in defined neurogenic niches, including the subgranular zone (SGZ) of the dentate gyrus and the subventricular zone (SVZ), through a multistep process involving neural stem cells, proliferative progenitors, neuroblasts, and maturing neurons[6–8].

Human induced pluripotent stem cell (hiPSC)–derived neural progenitor cells (NPCs) provide an experimentally accessible system for studying early stages of neuronal differentiation *in vitro*[9]. These systems offer controlled conditions, reproducibility, and compatibility with high-throughput single-cell profiling approaches, enabling the systematic characterization of transcriptional and regulatory programs across defined differentiation stages[10,11].

Here we present a multi-modal single-cell dataset generated from an NPC-based neuronal differentiation model. Cells were collected and profiled at multiple time points under baseline conditions and following exposure to Aβ 1–42 peptide. The dataset includes matched scRNA-seq and scATAC-seq measurements, together with clustering annotations, trajectory groupings, and processed matrices suitable for downstream single-cell analysis. The inclusion of both modalities enables users to explore gene expression changes, chromatin accessibility patterns, and regulatory annotations using standard computational frameworks. All raw and processed data, along with metadata and analysis code, are publicly available to support reuse in studies of human neuronal differentiation and related methodological and biological applications.

Methods

Single-cell experimental design. To generate a time-resolved single-cell dataset of human neuronal differentiation under baseline and Aβ-exposed conditions, we profiled neural progenitor cells (NPCs) derived from the XCL1-DCXpGFP line using single-cell RNA sequencing (scRNA-seq; 10x Genomics) and single-cell ATAC sequencing (scATAC-seq; Chromium Next GEM). his experimental system provides an *in vitro* model for examining early stages of neuronal differentiation[12]. NPCs were differentiated by adding differentiation medium from Day 1 and maintaining it through Day 20. In a parallel condition, Aβ 1–42 peptide was introduced concurrently with the differentiation medium beginning on Day 1; throughout the manuscript, this will be referred to as the Aβ condition. Cells from both conditions were collected on Days 0, 7, 13, and 20 (Fig. 1A). After quality filtering, the scRNA-seq datasets comprised a total of 42,575 cells collected across four time points: 5,013 Aβ condition and 4,751 control cells at day 7; 4,991 Aβ and 4,026 control cells at day 13; 13,450 Aβ and 4,729 control cells at day 20; and 5,615 cells under basal conditions (Supplementary Figure 1A–D). Similarly, the scATAC-seq datasets included 30,192 high-quality cells: 5,850 Aβ and 5,575 control cells at day 7; 6,457 Aβ and 6,343 control cells at day 13; 2,297 Aβ and 1,230 control cells at day 20; and 2,440 cells from basal samples.

Identification of major annotated cell groups. To support navigation and interpretation of the dataset, we generated broad cluster annotations based on established marker gene sets associated with neuronal and astrocytic programs. Marker expression patterns for SOX-family transcription factors[13] and additional lineage-associated genes were computed for the 16 clusters identified during global clustering (Fig. 1B; Supplementary Table 1). The distributions of these marker scores are shown in Fig. 1C–D and are included as annotation fields in the deposited dataset.

Cell-cycle phase classification was performed using standard marker-based scoring (Supplementary Figures 1E, F). Three clusters (clusters 9, 10, and 11) showed patterns consistent with proliferative cell-cycle states and corresponded to samples collected at Day 0 prior to the onset of differentiation. Because these clusters represent pre-differentiation populations, they were excluded from analyses focused on the differentiation time course.

These steps provided two broad annotation groups—neuronal-program–associated and astrocytic-program–associated clusters[14]—used throughout the dataset's metadata to facilitate downstream analysis by users (Fig. 1B).

NPCs culture, neuronal differentiation and Aβ peptide 1–42 administration. NPCs derived from XCL1 DCXpGFP (ACS5005™, American Type Culture Collection, ATCC, Manassas, VA, USA) were cultured following manufacturer recommendations. Briefly, 0.30×10^6 NPCs were seeded onto a CellMatrix Basement Membrane Gel (ATCC® ACS3035™) coated 12-well plate and incubated in NPC expansion medium: complete growth medium including DMEM/F-12 (Gibco, Fisher Scientific, Waltham, MA, USA), supplemented with the Growth Kit for Neural Progenitor Cell Expansion (ATCC® ACS3003) and then maintained in a humidified incubator (5% CO2, 37 °C).

Neuronal differentiation experiments were carried out for 7, 13 and 20 days by plating NPCs at a seeding density of 80,000 viable cells/cm2 in 6-well coated culture plates. First, NPCs were incubated in expansion medium (Day 0). From Day 1 (post-seeding), half of the medium was replaced with differentiation medium every 2-3 days throughout the culture period. Complete Differentiation Medium consisted of serum-free neuronal basal BrainPhys™ Neuronal Medium, formulated to improve the electrophysiological and synaptic properties of the neurons[15], NeuroCult™ SM1 Neuronal Supplement (1:50), N2 Supplement-A (1:100), Recombinant Human Brain-Derived Neurotrophic Factor (BDNF, 20 ng/mL), Recombinant Human Glial-Derived Neurotrophic Factor (GDNF, 20 ng/mL), Dibutyryl-cAMP (1 mM) and ascorbic acid (200 nM) (STEMCELL Technologies, Vancouver, BC, Canada). Half-fresh medium containing Amyloid β Protein Fragment 1–42 (50 nM; Sigma-Aldrich, St. Louis, MO, USA) or DMSO (Sigma-Aldrich) as vehicle was added once a week[16].

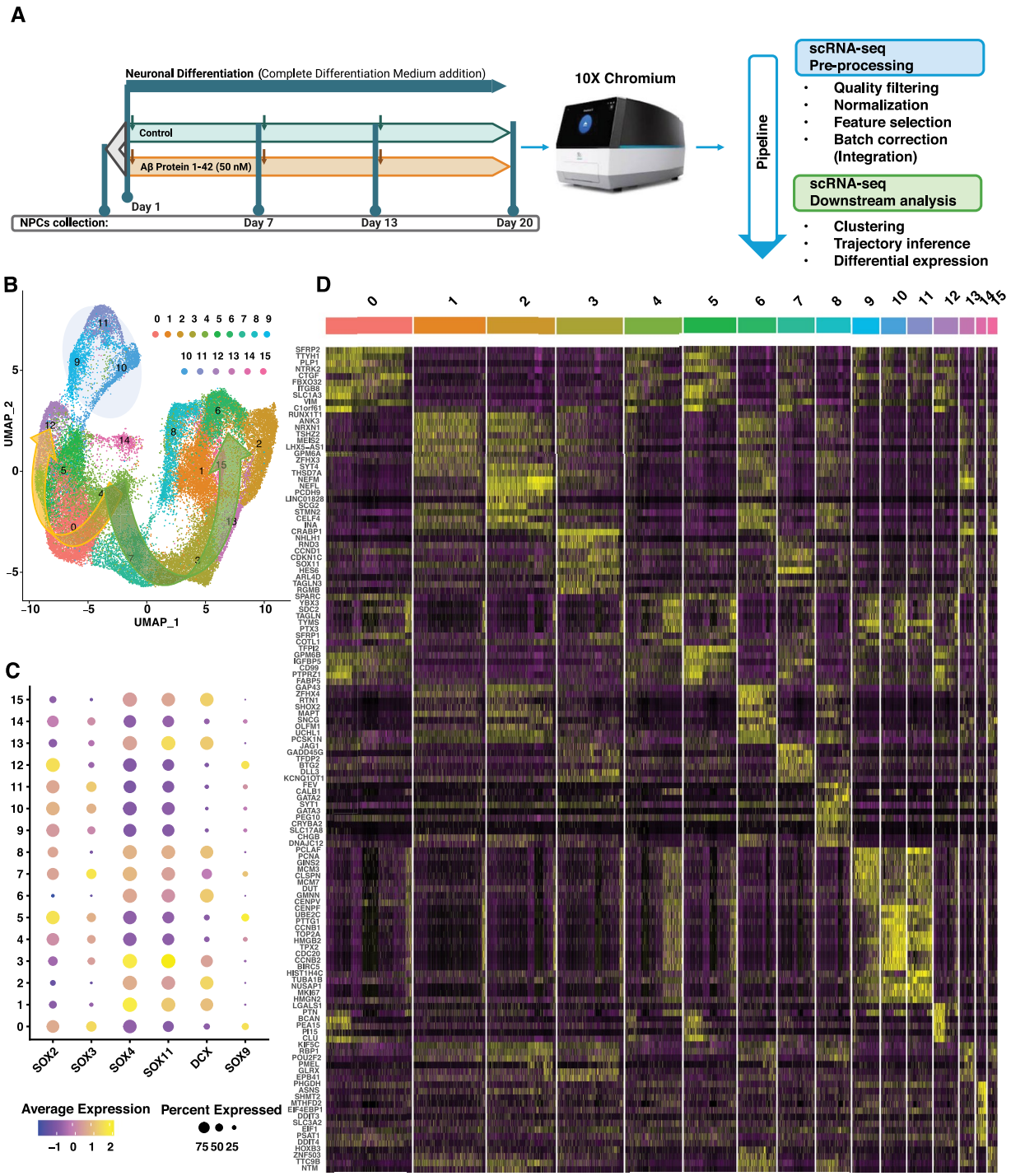

Fig. 1 Experimental design and overall finding: (**A**) Schematic representation of the experimental design for sample generation. The left panel outlines the biological workflow from sample collection to data acquisition. The right panel provides an overview of the bioinformatics pipeline, detailing the computational steps followed for data processing, quality control, and downstream analyses. (**B**) UMAP visualization of the non-filtered cells (based on quality-control analysis), with colors indicating the identified cell clusters (green indicating neurons and orange indicating astrocytes). (**C**) Dot plot showing the expression of selected astrocytogenesis- and neurogenesis-related biomarkers across clusters. Both expression level and the percentage of cells expressing each gene are indicated. (**D**) Heatmap displaying the top marker genes identified per cell cluster. Similar to panel C, yellow denotes high expression and purple denotes low expression.

NPCs were harvested on Day 0 and 7, 13, and 20 days of differentiation for both conditions by detaching them with Accutase (Innovative Cell Technologies, San Diego, CA), then washed with Dulbecco's phosphate-buffered

saline (DPBS, Sigma-Aldrich) and centrifuged at 13,000 rpm. NPCs pellet was resuspended in 100 µL of fresh and filtered buffer (0.05% BSA in PBS).

Cell sorting. To enrich the NPCs viability before the single-cell experiment, 1 µL of TO-PRO-3 (Invitrogen) was added to each sample to determine dead cells. Viable cells gated on the FSC/SCC after analysis with Summit V6.3.1.16945 software were sorted on the basis of the negative expression of TOPRO-3 using a MoFlo Astrios EQs Sorter (Beckman Coulter, Brea, CA, USA). Standard, strict FSC width versus area criteria were used to discriminate doublets and gate only singleton cells. Between 20.000 and 70.000 viable cells were collected in 0.5% BSA in PBS and immediately processed for scRNA-seq.

Single-cell RNA-sequencing (scRNA-seq). The single-cell transcriptome of NPCs was examined using NEXTGEM Single Cell 3′ Reagent Kits v3.1 from 10X Genomics, according to the manufacturer's instructions[16]. Between 6.000 and 20.000 cells were loaded at a concentration of 500–1.000 cells/µL on a Chromium Controller instrument (10X Genomics) to generate single-cell Gel Beads-in-emulsion (GEMs). In this step, each cell was encapsulated with primers containing a fixed Illumina (San Diego, CA, USA) Read 1 sequence, a cell-identifying 16 bp 10X barcode, a 10 bp Unique Molecular Identifier (UMI), and a poly-dT sequence. Upon cell lysis, reverse transcription yielded full-length, barcoded cDNA. This cDNA was then released from the GEMs, PCR-amplified, and purified with magnetic beads (SPRIselect, Beckman Coulter). Enzymatic Fragmentation and Size Selection was used to optimize cDNA size prior to library construction. Fragmented cDNA was then end-repaired, A-tailed, and ligated to Illumina adaptors. A final PCR-amplification with barcoded primers enabled sample indexing. Library quality control and quantification were performed using the Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and the 4200 TapeStation System (Agilent Technologies, Inc., Santa Clara, CA, USA), respectively. Sequencing was performed in a NextSeq. 500 (Illumina) (Read1: 28 cycles; Read55 cycles; i7 index: 8 cycles) at an average depth of 20,000 reads/cell.

scRNA-seq analysis: quality control, preprocessing, filtering, integration, and global clustering. Eleven scRNA-seq samples were generated, including one at Day 0 and paired samples for Days 7, 13, and 20 for both control and Aβ conditions (Supplementary Figure 1B; Fig. 1A). Raw sequencing data were processed using the 10x Genomics Cell Ranger 3.0.1 pipeline for de-multiplexing, barcode processing, and single-cell gene counting, with alignment performed against the GRCh38 Ensembl reference genome using STAR[17]. Quality control included removal of cells contributing to a low-quality bimodal distribution observed in one Day 20 control sample, exclusion of cells with fewer than 3,250 counts or 1,500 detected features, quantile-based filtering to remove cells below the 10th or above the 90th percentile for counts and features within each sample, retention of genes detected in at least three cells, inclusion of cells with at least 200 detected genes, and exclusion of cells with mitochondrial proportions $\geq$ 10% (Supplementary Figures 1C,D). Cell-cycle scoring was applied to identify clusters enriched for S and G2/M-phase signatures (clusters 9, 10, and 11; Supplementary Figure 1E–G), corresponding to Day 0 samples, and these were excluded from analyses of the differentiation time course. After filtering, samples were normalized using the SCTransform (SCT) workflow. Doublets were identified with DoubletFinder[18] using an estimated rate of 7.5% and removed prior to reprocessing by SCT (Supplementary Figure 1H). Dataset integration was performed using the Seurat v4 anchor-based integration strategy[18], in which weight matrices were constructed, Gaussian kernels applied, and transformation matrices iteratively computed for each dataset pair based on guide tree ordering. Integration was carried out using SCT-normalized data and 30 components. Global clustering was performed using the Louvain algorithm[19] (resolution 0.6) implemented in Seurat.

scRNA-seq analysis: clustering resolution within neuronal and glial differentiation trajectories. To organize the scRNA-seq dataset at multiple levels of granularity, we evaluated relationships among the clusters identified in the global analysis (Fig. 1B) using two computational approaches: (i) Spearman-based correlation of the top 2,000 marker genes per cluster (Supplementary Figure 1H) and (ii) pseudotime ordering using Monocle3 with default parameters (Supplementary Figure 1B). These approaches were used to assess similarity structure across clusters and to determine whether groups of clusters could be treated jointly in subsequent analyses.

Based on these similarity assessments, two levels of resolution were generated. The high-resolution level corresponds to the individual clusters shown in Fig. 1B. In addition, a lower-resolution grouping was created in which clusters 4, 0, 7, and 3 were assigned to groups A, B, C, and D, respectively, and clusters 1, 2, 6, 8, 13, and 15 were aggregated into group E (Fig. 2A). These group labels are provided as metadata to facilitate downstream organization of the dataset.

scRNA-seq analysis: differential expression. Differential expression analyses were performed using the limma framework[20–22]. Comparisons were carried out between adjacent stage groupings defined in Fig. 2A, using contrasts of the form C(i + 1) – C(i), where *i* refers to stages A–E in the control condition. To compute condition-dependent contrasts, we applied the formula (T(i + 1) – T(i)) – (C(i + 1) – C(i)), where *T* denotes samples from the Aβ 1–42 condition and *C* denotes control samples. Sample collection time was included as a covariate in all models. The resulting statistics and contrast outputs are provided in the deposited dataset[23].

Gene set Enrichment Analysis. ClusterProfiler[24] was employed to conduct both Gene Set Enrichment Analysis (GSEA) and Overrepresentation Analysis (ORA). For GSEA, data were ranked based on the fold change (FC). DOSE package has been used for disease over-representation analysis[25].

Proportion analysis of neurogenesis between control and beta conditions. Cells assigned to the five broad stage groups (A–E) were used for a comparative proportion analysis. Metadata fields, including sample

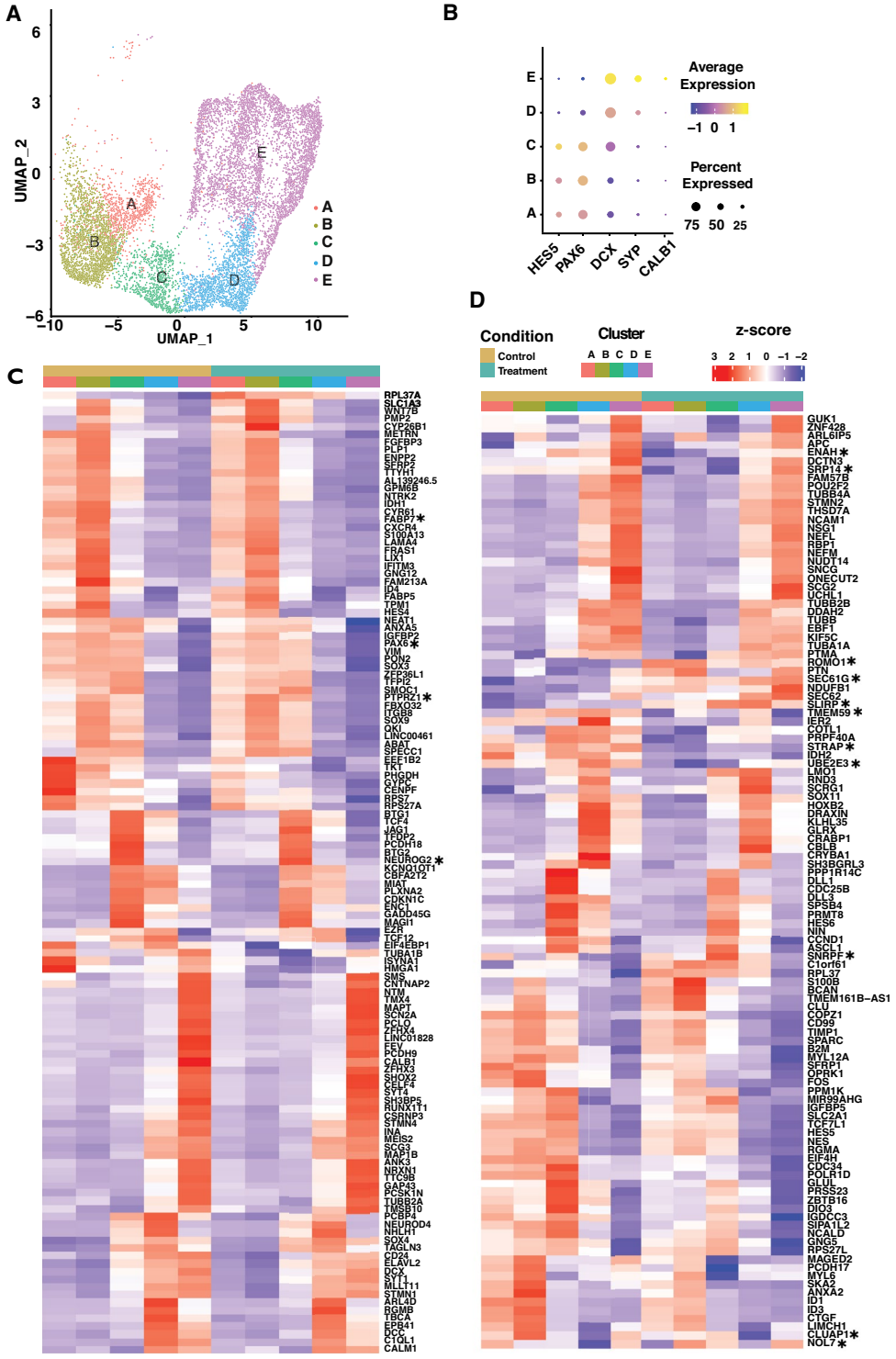

Fig. 2 Neurogenesis differentiation trajectory: (**A**) UMAP projection of neurogenesis-related cells, with ordered letters corresponding to sequential steps identified along the trajectory. (**B**) Dot plot presents known biomarkers following neurogenesis maturation based on bibliography. Both expression level and the percentage of cells expressing each gene are indicated. (**C**) Heatmap of significantly differentially expressed during neurogenesis, similarly in control and Aβ condition (Set 1). Colors represent expression levels. (**D**) Heatmap of significantly differentially expressed genes across each pair of steps in the differentiation trajectory, highlighting genes with distinct regulation under Aβ versus control conditions. Asterisks indicate genes uniquely differentially expressed in Aβ but not in control (Set 3). The remaining genes are differentially expressed in both Aβ and control, but with significant differences in magnitude (Set 2).

identity, condition, and cluster annotations, were extracted to define groups for comparison. To standardize representation across samples, the minimum number of cells available across all replicates (N = 3,054) was used as a target for random subsampling. Proportions of cells in each stage were then calculated within each sample.

To evaluate differences in cell-state proportions between conditions, chi-square tests were applied to contingency tables generated from subsampled counts. Ordinal logistic regression was additionally used to model transitions across the ordered stages (A–E). Permutation-based proportion testing was performed with the scProportionTest package, generating fold-change estimates and false-discovery–rate–adjusted p-values for each comparison. All proportion values, test statistics, and outputs are provided in the accompanying supplementary files (Supplementary Figure 2A–C).

Human brain data analysis. We downloaded the count matrix associated with the bulk RNA-seq data set from GEO: GSE173955[26]. In total it consists of 8 samples of Alzheimer's disease (AD) and 10 samples of controls. Dataset includes individuals of both sexes with a similar representation across conditions. All the samples correspond to adults between 72 and 100 years old (except for one patient aged 55).

The count matrix was normalized using TMM[27], and limma + voom was used to perform the differential analysis[28] between healthy and AD individuals. Age and sex were considered as covariates in the model. Genes with an adjusted p-value less than 0.05 and logFC (logarithmic fold change) greater than 1 or smaller than −1 were considered as significantly differential genes.

Single-cell ATAC-sequencing (scATAC-seq). scATAC-seq experiments were performed with the Chromium Next GEM Single Cell ATAC kit v1.1 from 10X Genomics, following the manufacturer's instructions. Nuclei isolation was done for samples with 100.000 cells or less, according to the 10X Genomics demonstrated protocol for low cell input[29]. Briefly, cells were collected in ice-cold lysis buffer (PBS + 0.04% BSA), counted and centrifuged to concentrate them in 50 µL. Isolated nuclei were resuspended in chilled diluted nuclei buffer and used immediately to generate a bulk transposition reaction following the manufacturer's instructions and conditions. In this step, adapter sequences are added to the ends of the DNA fragments.

After transposition, GEMs were generated in the Chromium Controller (10x Genomics). Oligonucleotides containing an Illumina P5 sequence, a 16nt 10X barcode and a read1 sequence are released and mixed with DNA fragments and master mix. After incubation, GEMs are broken and pooled fractions are recovered. P7, a sample index and read2 sequence are added during library construction via PCR. Library quality control and quantification were performed with the Qubit 3.0 Fluorometer (Life Technologies) and Agilent's 4200 TapeStation System. Sequencing was performed using a NextSeq. 500 (Illumina) (Rd1 50 cycles, i7 Index 8 cycles, i5 index 16 cycles and Rd2 50 cycles) to an average depth of 50.000 reads.

scATAC-seq analysis: preprocessing. Seven samples were produced using the mentioned scATAC-seq technology protocol. The design was similar and paired (at sample level) with the single-cell transcriptomic profiling.

For the processing of scATAC-seq samples, including read de-multiplexing, duplicate reads removing, background cell removal and tabular Peak/cell matrix generation we utilized the 10x Genomics Cell Ranger ATAC pipeline[30]. Reads were mapped to the Ensemble human reference genome GRCh38 peaks.

Initially, consensus peaks were identified across all samples and used as the reference peak list. Subsequently, the count matrix (i.e., the number of reads per peak per cell) was quantified using this set of consensus peaks.

To assess the presence of batch effects in the resulting scATAC-seq datasets, we followed the protocol outlined in the Signac package[31]. Initially, we implemented several filters to eliminate low-quality cells. The first filter excluded cells with fewer than 500 counts per cell. Subsequently, cells were removed if they had a number of fragments in the peak region less than 3,000 or greater than 100,000. Similarly, cells with a percentage of reads in peaks (FRIP) less than 40% were eliminated. Cells with a ratio of fragments falling in blacklisted regions equal to or greater than 0.025 were also discarded. Finally, cells with a ratio of mono-nucleosome cut fragments to nucleosome-free fragments (nucleosome signal) equal to or greater than 4, as well as cells with TSS enrichment less than or equal to 2, were eliminated.

After quality filtering, the data of each sample were normalized and embedded spaces were generated using the protocol proposed by Signac (Latent semantic indexing (LSI)). For this calculation, the 20th quantile was used as the most variable peak. Signac's 'ElbowPlot' was used to determine the number of dimensions to use and 'DepthCor' was employed to identify dimensions highly correlated with read depth. The resulting components excluding those correlated with read depth were used for downstream analyses.

To integrate the information from the different samples, the protocol for unpaired cells proposed in Seurat4 was used (Supplementary Figure 3A). To calculate the anchors between the cells of the different samples, 20.000 peaks randomly selected from the set of peaks overlapping in all samples were used. Following Signac guidelines, we computed gene activities across integrated samples and transferred metadata for cell groups.

scATAC-seq analysis: preprocessing and preparing Regulon Analysis. To generate inputs for enhancer–gene regulatory network workflows, additional preprocessing steps were performed following the SCENIC + framework[32], using the pycisTopic package[33]. Cells were filtered using a FRIP threshold of 50% and a TSS enrichment threshold of 5. Cells with a log-transformed count of unique fragments per barcode below 3.8 were removed. Potential doublets were identified and excluded with Scrublet using default parameters[34].

Dimensionality reduction was recomputed using latent Dirichlet allocation (LDA) implemented in pycisTopic. Multiple topic numbers were evaluated, and the final selection was based on the highest consensus score derived from the four metrics provided within pycisTopic.

For comparative preprocessing across experimental conditions, cells were subset into control and Aβ-treated groups, and previously transferred scRNA-seq cluster annotations were retained for organizational purposes.

Candidate regulatory regions were identified following SCENIC + procedures. Region–topic probability matrices were binarized using Otsu's method; drop-out imputation was performed by multiplying cell–topic and topic–region distributions; matrices were normalized using a scale factor of 10^4; and highly variable regions were selected. Differentially accessible regions (DARs) were computed using the Wilcoxon rank-sum test, using the same stage group order defined during scRNA-seq processing and preserved through Signac batch-correction workflows.

Regulon Analysis: eGRN. To generate inputs for enhancer–gene regulatory network (eGRN) workflows, scRNA-seq and scATAC-seq data were combined into a pseudo-multiome dataset. Cells were randomly selected from each dataset within each cell annotation category (using the differentiation clusters defined in the scRNA-seq analysis). For each annotation, average scRNA-seq counts and average scATAC-seq accessibility values were computed and consolidated into meta-cells using a K = 5 parameter. These aggregated matrices served as the input for downstream SCENIC + processing.

Motif enrichment and region-to-gene linkage analyses were carried out using the SCENIC + framework, based on the binarized topic matrices and differentially accessible regions produced by pycisTopic. Motif scoring was performed using cisTarget, which ranks genomic regions from the SCENIC + hg38 motif database (generated using Cluster-Buster). Input regions were intersected with database regions using a 40% minimum overlap criterion. For each motif, recovery curves were computed and areas under the curve (AUCs) were normalized across motifs to obtain normalized enrichment scores (NES). Motifs with NES >3 were retained.

Differential Enrichment of Motifs (DEM) was then applied using Wilcoxon rank-sum tests comparing foreground and background region sets. Adjusted p-values <0.05 (Bonferroni-corrected) and log2FC >0.5 were used as filtering thresholds. In these analyses, differentially accessible regions served as the foreground, and 500 regions from alternative DAR sets were used as background, with matched promoter proportions.

For each motif, motif-associated regions were assembled into cistromes. Cistromes annotated to the same transcription factor were merged to obtain the transcription factor–region sets. Region–gene distances and corresponding importance scores were computed following the gradient boosting machine (GBM)–based procedures specified in the SCENIC + workflow.

Data Records

All raw and processed data generated in this study are publicly available in the NCBI Gene Expression Omnibus (GEO) under accession number GSE307094 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE307094). The dataset is organized into two main components corresponding to the scRNA-seq and scATAC-seq experiments.

For the scRNA-seq dataset, we provide raw expression count matrices in Matrix Market (MTX) format, accompanied by gene annotation and cell barcode files in CSV format. Processed data, including normalized expression matrices and cell-level clustering assignments, are provided in matching formats to facilitate their direct use in downstream analyses without requiring additional preprocessing.

The scATAC-seq dataset is structured analogously. Raw peak-by-cell count matrices are provided in MTX format, together with corresponding feature and barcode annotation files. Normalized accessibility matrices and clustering information derived from the integrated analysis are also included. In line with the scope of this Data Descriptor, the scATAC-seq data are intended primarily to complement the transcriptomic dataset as a regulatory-prioritization resource, rather than to support peak-level differential accessibility analyses.

For technical validation, we additionally reference a human hippocampal bulk RNA-seq dataset obtained from GEO accession **GSE173955**, comprising eight samples from individuals with Alzheimer's disease and ten samples from age-matched controls. This dataset was used exclusively for validation analyses and is fully accessible through GEO.

Technical Validation

Validation of the neuronal lineage trajectory at single-cell level. To evaluate the internal consistency of the single-cell RNA-seq dataset, we examined the organization of control-condition cells across clusters identified during preprocessing (clusters 0, 1, 2, 3, 4, 6, 7, 8, 13 and 15; Fig. 1B). Correlation analysis of cluster-specific gene expression profiles, together with pseudotime ordering (Supplementary Figures 1G, 5A), supported the use of two resolution levels for downstream analyses. At the fine-resolution level, individual clusters were retained. At a broader resolution, clusters 1, 2, 6, 8, 13 and 15 were grouped together, yielding five aggregated states (A–E; Fig. 2A).

To characterize these aggregated states, expression distributions were visualized for reference marker genes commonly used to describe early neuronal differentiation (Fig. 2B; Supplementary Figure 6A). For each aggregated group, marker expression distributions were examined across replicates, providing users with reference plots for assessing annotation consistency.

An independent pseudotime analysis arranged cells in the same A–E order (Supplementary Figures 5A, 6B). To examine transcriptional changes associated with transitions between aggregated states, functional enrichment analyses were performed on genes varying between adjacent groups (Supplementary Figure 6C; Supplementary Table 2). These enrichments are provided as reference outputs accompanying the dataset.

Together, these analyses—cluster correlation summaries, marker distribution visualizations, pseudotime ordering, and transition-level enrichment outputs constitute a set of internal validation outputs for users working with the A–E grouping under control conditions.

To further assess dataset consistency across experimental conditions, we compared the distribution of cells across the A–E groups between control and Aβ 1–42 samples. Each replicate was subsampled to the minimum shared cell number (N = 3,054), and the proportions of cells in each group were estimated. Differences between

conditions were evaluated using chi-square tests, ordinal logistic regression, and a permutation-based approach (scProportionTest) (scProportionTest) (Supplementary Figure 2A–C). All proportion estimates and associated test statistics are provided in the Supplementary Information.

Assessment of condition-associated transcriptional differences. To evaluate whether the dataset supports detection of condition-associated transcriptional variation, we compared gene-expression profiles across the A–E differentiation stages between control and Aβ 1–42 samples[35]. Three categories of differentially expressed genes (DEGs) were defined: genes showing similar stage-dependent expression patterns in both conditions (Set1), genes showing stage-dependent patterns with statistically significant differences between conditions (Set2), and genes detected as differential only in the Aβ samples (Set3) (Fig. 2C,D; Supplementary Figure 5B,C; Supplementary Table 2). Set sizes were 124, 121, and 11 genes, respectively.

To assess structure within the identified DEGs, we performed functional enrichment analyses. For Set2 genes, Gene Set Enrichment Analysis and overrepresentation analysis identified enrichment for several curated functional categories across specific stage transitions (Supplementary Figure 7A–C). KEGG pathway analysis also identified additional enriched terms (Supplementary Figure 7B). These enrichment patterns indicated that the statistically defined gene sets corresponded to structured and internally consistent reference categories.

We examined transcription factors (TFs) represented among Set2 genes. TFs exhibiting stage-dependent variation in the control group generally showed similar patterns in the Aβ group, with a subset displaying statistically significant differences in expression levels between conditions (Supplementary Table 3; Fig. 2D). These TF-associated profiles further supported the internal consistency of the condition-based contrasts.

Together, these analyses show that the dataset enables identification of statistically defined gene-expression differences between experimental conditions, and that these differences form structured groups that can be evaluated using standard enrichment and annotation tools. Importantly, it mirrors the clinically relevant scenario where neurogenic niches remain structurally present but functionally impaired in AD patients.

Assessment of transcriptional variation in late-stage populations. To evaluate whether condition-associated transcriptional differences could also be detected in later-stage populations, we examined gene-expression patterns within subclusters corresponding to the maturing populations contained in cluster E (Fig. 3A–D). As in the A–E trajectory analysis, three categories of differentially expressed genes (DEGs) were identified when comparing control and Aβ 1–42 samples across transitions: genes showing comparable expression patterns across conditions (Set1m), genes showing statistically significant differences between conditions (Set2m), and genes detected as differential only in Aβ-treated samples (Set3m) (Supplementary Figure 8A,B; Supplementary Table 4). Set sizes were 101, 173, and 36 genes, respectively. These three groups provided a structure for summarizing condition-associated variation within the late-stage populations.

To assess whether the statistically defined Set2m groups corresponded to structured annotations, we conducted enrichment analyses using standard reference gene sets. Enrichment was observed for curated categories across specific transitions (Supplementary Figure 9A–C), and KEGG pathway analysis identified additional enriched terms (Supplementary Figure 9B). Transcription factors detected within Set2m were also examined: most transcription factors exhibited similar transition-associated patterns across the control and Aβ conditions, whereas a subset displayed statistically significant differences in expression levels between conditions (Supplementary Table 5). These results further supported internal consistency of the condition-based contrasts.

Together, these analyses indicate that the dataset supports detection of structured transcriptional variation within later-stage populations and that these differences can be summarized and validated using standard differential expression and enrichment workflows.

Assessment of chromatin accessibility consistency and regulatory annotations. To evaluate consistency of chromatin accessibility across experimental conditions, we compared scATAC-seq peak accessibility profiles between control and Aβ 1–42 samples along the A–E differentiation trajectory. Differential accessibility testing did not identify statistically significant peak-level differences between the two conditions, indicating comparable global accessibility profiles across experimental groups.

To further assess regulatory structure within the dataset, we conducted an integrated analysis of scATAC-seq–derived accessibility profiles with scRNA-seq expression data to compute gene regulatory network (GRN) components. Across differentiation stages, the integrated multi-modal analysis identified 24,579 candidate regulons, of which 3,840 met significance criteria. Regulons associated with transcription factors (TFs) that showed condition-dependent expression differences in the scRNA-seq dataset (Supplementary Tables 3, 5) were examined to compare regulatory annotations between conditions. Regulons linked to certain TFs were recovered primarily in control samples, whereas others were identified in both control and Aβ samples (Fig. 4A,B).

To assess structure within the regulon sets, we applied functional enrichment analysis. Regulon groupings showed enrichment for established annotation categories (Fig. 4C,D), providing an internal check that the integrated regulatory signatures mapped onto known functional groupings.

Together, these analyses confirm that the scATAC-seq data provides reproducible peak-level accessibility profiles across conditions and that integrated chromatin–transcriptome analyses yield regulon sets with structured and interpretable annotation patterns.

Consistent with the observations described above, we emphasize that differential peak-level accessibility between control and Aβ conditions was limited. Therefore, within this dataset the scATAC-seq modality should be interpreted primarily as a regulatory prioritization resource—providing TF activity inference, motif enrichment, and SCENIC + –derived regulon annotations—rather than as evidence of strong condition-specific chromatin-accessibility changes.

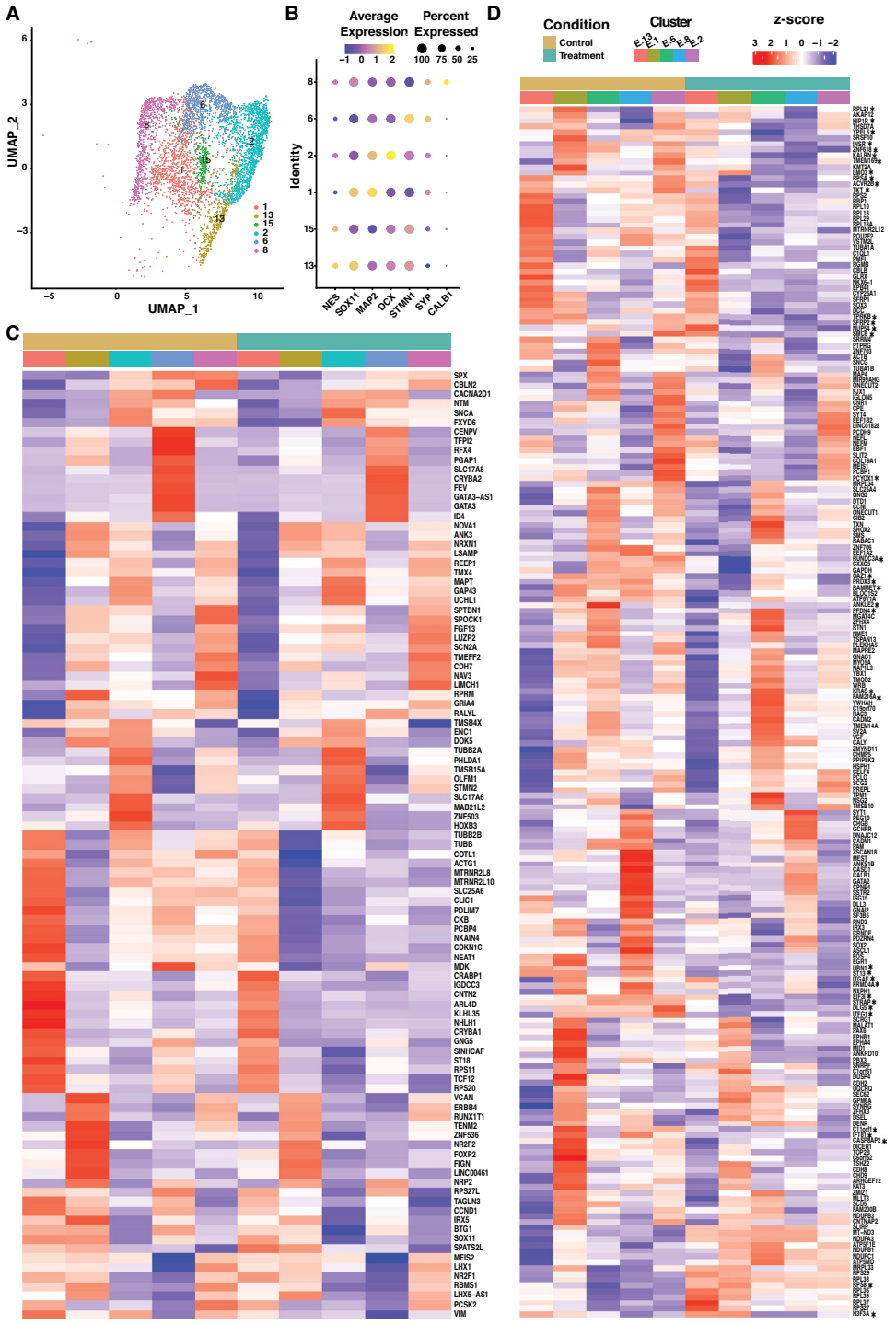

Fig. 3 Neurogenesis differentiation trajectory: mature cells. (**A**) UMAP projection of mature neurogenesis-related cells, with ordered numbers corresponding to sequential steps identified along the trajectory. Two differentiation trajectories were identified: one following the 13-1-6-8 and the second following 13-2. (**B**) Dot plot presents known biomarkers following neurogenesis advance maturation based on bibliography. Both expression level and the percentage of cells expressing each gene are indicated. Cluster 15 was not included in the plot as no differential gene expression was identified. (**C**) Heatmap of significantly regulated genes across each pair of steps in the advanced maturation differentiation trajectory, defined using only control cells. Colors represent expression levels. (**D**) Heatmap of significantly differentially expressed genes across each pair of steps in the advanced maturation differentiation trajectory, highlighting genes with distinct regulation under Aβ peptide 1–42-treated versus control conditions. Asterisks indicate genes uniquely differentially expressed in Aβ peptide 1–42-treated but not in control. The remaining genes are differentially expressed in both Aβ peptide 1–42-treated and control, but with significant differences in magnitude. Cluster 15 was not included in the plot as no differential gene expression was identified.

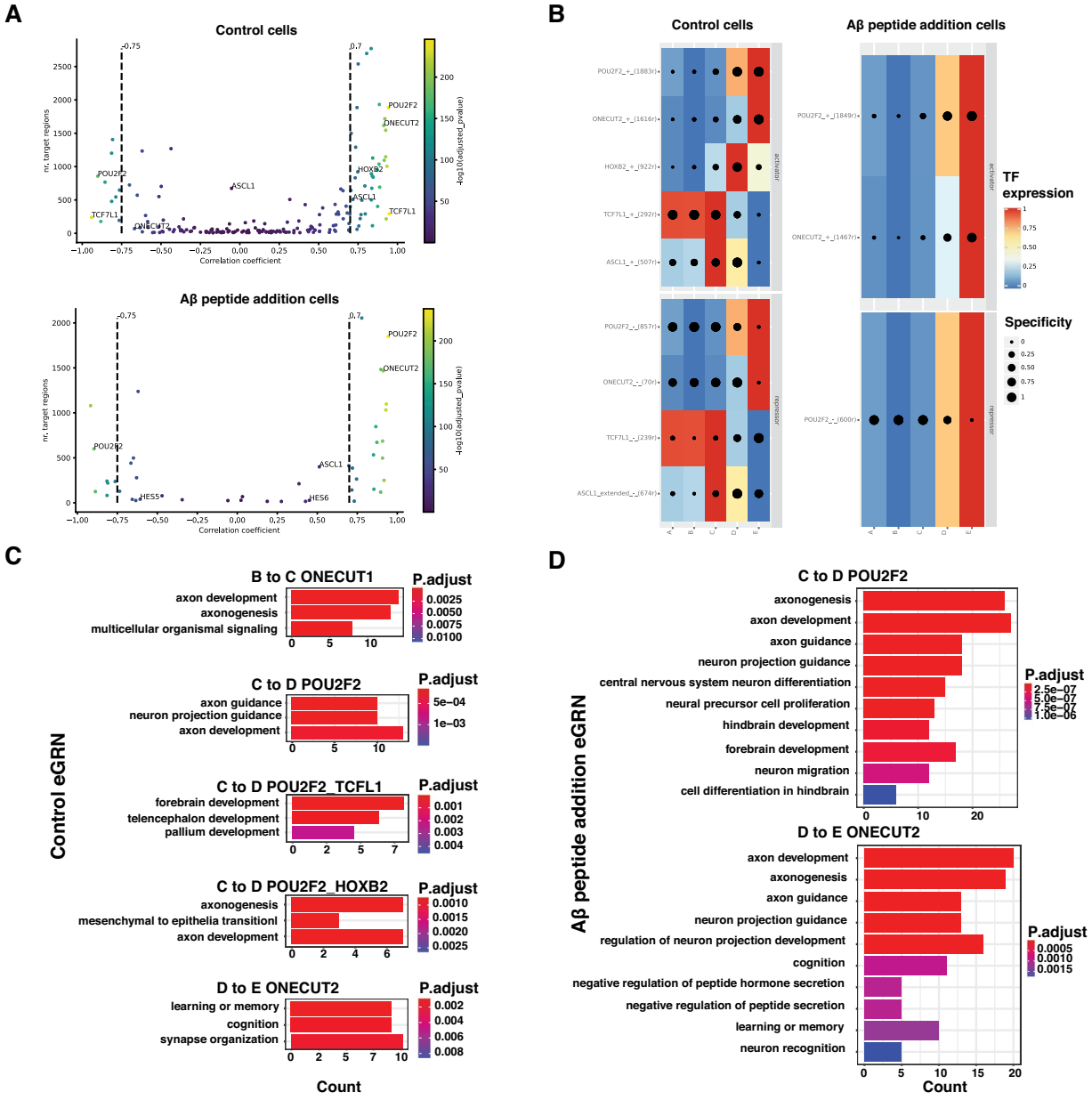

Fig. 4 Neurogenesis under chromatin accessibility robustness: (**A**) Scatterplots of significant regulons. Top panel: control cells; Bottom panel: Aβ peptide 1–42-treated cells. X-axis shows the correlation coefficient between transcription factor (TF) and target genes, Y-axis indicates the number of target regions. Dot color represents statistical significance. This panel illustrates the overall distribution of regulon activity and significance across conditions. (**B**) Heatmaps/dot plots of selected regulons associated with key TFs exhibiting stage-specific activity. Dot color reflects TF expression, and dot size represents regulon specificity. Regulons are ordered according to differentiation stages A to E to show dynamic changes in activity. (**C,D**) Functional enrichment of significant regulons across differentiation transitions. X-axis indicates the number of genes in each enriched pathway, and color represents pathway significance. Panel C shows results for control eGRNs, and Panel D for Aβ peptide 1–42-treated eGRNs. Only pathways with significant enrichment are displayed, illustrating the functional consequences of regulon activity throughout differentiation.

Cross-dataset comparison using external bulk RNA-seq data. To assess external consistency of the transcriptional signals derived from the single-cell dataset, we compared the Set2 gene groups identified from the control versus Aβ 1–42 contrasts with publicly available bulk hippocampal RNA-seq data from individuals with Alzheimer's disease and controls (GEO: GSE173955)[36]. A subset of genes identified as differential in the single-cell dataset also showed differential expression in the external dataset ($|\text{fold change}| > 1$, $p < 0.01$; Fig. 5A). This comparison served as an external check of gene-level variation identified in our analysis.

To further evaluate the relationship between stage-annotated gene sets and the external dataset, gene lists derived from each broad differentiation group (A–E) were tested for enrichment against differential expression

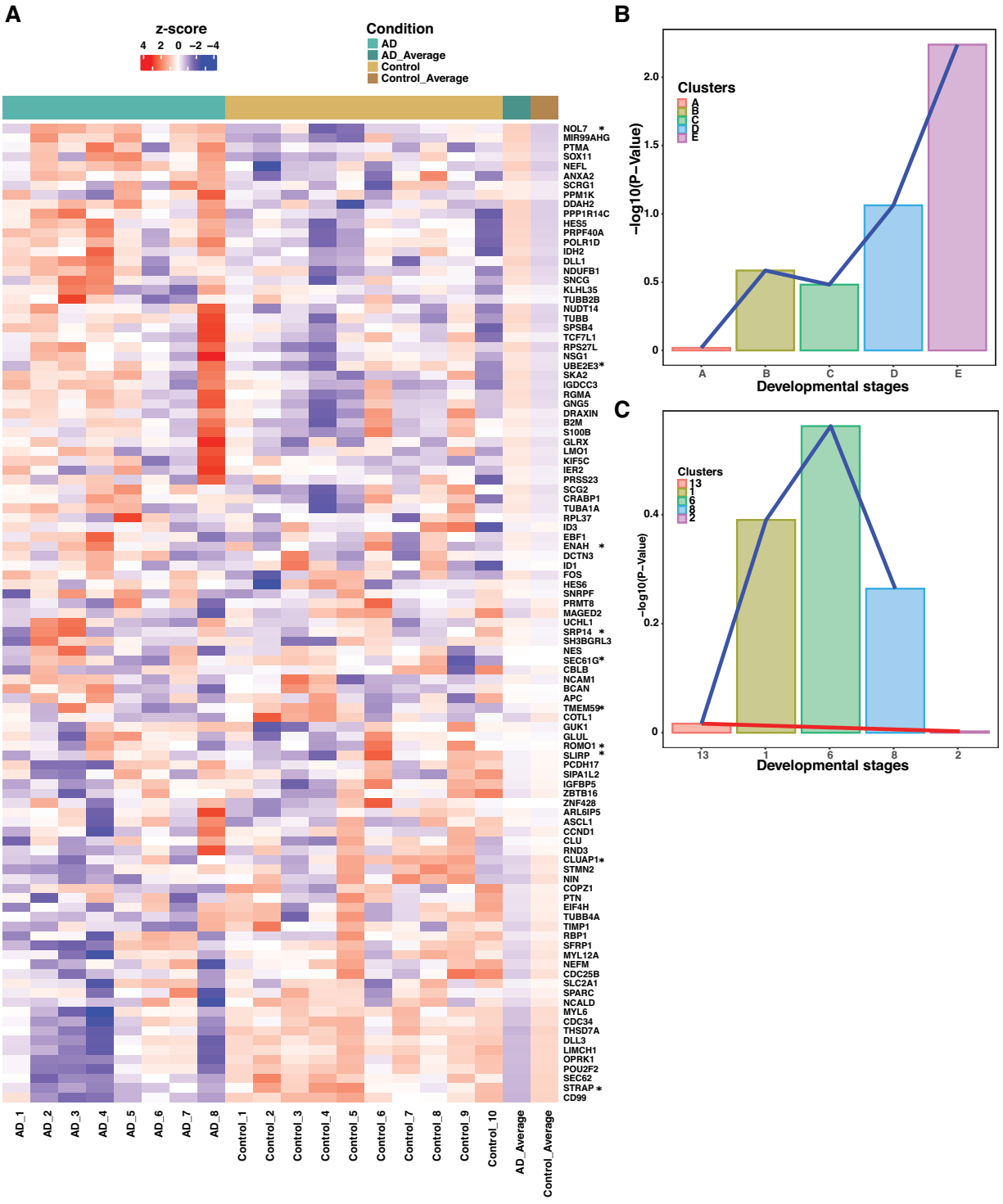

Fig. 5 Findings validation through human Alzheimer case study: (**A**) Heatmap of bulk gene expression. Genes were selected based on significant differential expression between control and Aβ peptide 1–42-treated single-cell conditions in broad and fine resolution analysis. Gene expression is derived from human brain dataset, GSE173955, including Healthy and Alzheimer samples. (**B**) Bar plot showing gene set enrichment p-values for healthy versus Alzheimer's disease patients in the GSE173955 dataset. Each gene set was derived from differentially expressed genes (DEGs) identified when comparing control and Aβ peptide 1–42-treated single-cell conditions resolution at each step of the broad neurogenesis trajectory. (**C**) Bar plot showing gene set enrichment p-values for healthy versus Alzheimer's disease patients in the GSE173955 dataset. Each gene set was derived from DEGs identified when comparing control and Aβ conditions across sub-steps of the mature neurogenesis trajectory (sub-clusters of cell cluster E). Within E, we identified two trajectories: 13-1-6-8 (blue line) and 13-2 (red line).

results from the bulk hippocampal dataset (Fig. 5B). Enrichment statistics for each group are provided in the accompanying materials. A similar analysis was performed for subclusters within the maturing neuronal populations (cluster E), with enrichment results shown in Fig. 5C.

These analyses provide an external reference-based assessment of the gene-expression variation identified in the dataset and demonstrate that specific gene groups obtained from the single-cell contrasts can be evaluated against independent transcriptomic resources.

Usage Notes

This dataset provides matched (multimodal) single-cell RNA-seq and ATAC-seq profiles generated from human neural progenitor cells (NPCs) differentiated *in vitro* and collected across four time points under two experimental conditions. The resource is intended to support users interested in exploring single-cell transcriptomic and chromatin-accessibility patterns across a controlled differentiation series, as well as those wishing to compare condition-specific variation using standard single-cell workflows.

Both raw and processed files are included in the GEO repository, together with metadata fields such as clustering annotations, cell-cycle scores, and trajectory groupings. These can be readily incorporated into popular single-cell analysis frameworks (e.g., Seurat, Scanpy, Signac). For users who wish to begin with processed objects, normalized matrices and integrated embeddings are provided to facilitate rapid data navigation.

To support interactive exploration of the scRNA-seq component, a Shiny application is available at: https://translationalbio.shinyapps.io/neurogenesis/. This browser allows users to visualize cluster structure, marker expression, and metadata categories without installing local software.

The accompanying GitHub repository contains all scripts used for preprocessing, integration, and annotation, enabling full reproducibility of the pipelines and providing a template for extending the analyses. Users may adapt these workflows to perform additional comparisons, integrate with other datasets, or benchmark computational tools.

The scATAC-seq data provide complementary chromatin-accessibility information that is intended primarily for regulatory prioritization and contextual interpretation. Because peak-level differential accessibility between conditions is limited, users should view the scATAC-seq modality as a resource for identifying candidate transcription factors, motifs, and regulatory programs rather than as a source of strong de novo accessibility changes. The provided outputs—including gene activity matrices, motif enrichment annotations, region-to-gene linkages, and SCENIC + –derived regulon files—allow users to explore TF activity, construct regulatory networks, and integrate chromatin information with transcriptomic patterns using tools such as Signac, ArchR, or SCENIC+.

As the dataset is derived from an established NPC differentiation system, it is suitable for applications such as comparative analyses of *in vitro* neuronal differentiation protocols, evaluation of computational trajectory or regulatory-network methods, and integration studies with external transcriptomic or epigenomic resources. Users should note that the system is not designed to model region-specific *in vivo* neurogenesis and should interpret comparisons accordingly[37].

Data availability

All raw and processed data files are publicly available in the Gene Expression Omnibus (GEO) repository (accession number: GSE307094). The dataset used for technical validation consists of human samples and can be downloaded from GEO: GSE173955. This includes 8 Alzheimer's disease samples and 10 control samples. All intermediate supplementary data are available in the Figshare repository: https://doi.org/10.6084/m9.figshare.30073657 and https://doi.org/10.6084/m9.figshare.30073594.

Code availability

All R and Python scripts are available in the GitHub repository (https://github.com/TranslationalBioinformaticsUnit/Human-neuronal-differentiation-under-AB-exposure-a-single-cell-transcriptomic-and-epigenomic-datase) to ensure reproducibility and facilitate further data analysis and extension.

Received: 25 September 2025; Accepted: 24 February 2026;
Published online: 10 March 2026

References

1. Hampel, H. *et al*. The Amyloid-β Pathway in Alzheimer's Disease. *Molecular Psychiatry* **26** 5481–5503, https://doi.org/10.1038/s41380-021-01249-0 (2021).
2. Amber, S. *et al*. Amyloid-beta Induced Neurotoxicity Impairs Cognition and Adult Hippocampal Neurogenesis in a Mouse Model for Alzheimer's Disease. *Curr. Alzheimer Res.* **17**, 1033–1042 (2021).
3. Culig, L., Chu, X. & Bohr, V. A. Neurogenesis in aging and age-related neurodegenerative diseases. *Ageing Research Reviews* **78**, 101636, https://doi.org/10.1016/j.arr.2022.101636 (2022).
4. Disouky, A. & Lazarov, O. Adult hippocampal neurogenesis in Alzheimer's disease. in *Progress in Molecular Biology and Translational Science* **177** 137–156 (Elsevier B.V., 2021).
5. Franjic, D. *et al*. Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* **110**, 452–469.e14 (2022).
6. Terreros-Roncal, J. *et al*. *Impact of Neurodegenerative Diseases on Human Adult Hippocampal Neurogenesis*. https://www.science.org.
7. Gonçalves, J. T., Schafer, S. T. & Gage, F. H. Adult Neurogenesis in the Hippocampus: From Stem Cells to Behavior. *Cell* **167**, 897–914, https://doi.org/10.1016/j.cell.2016.10.021 (2016).
8. Niklison-Chirou, M. V., Agostini, M., Amelio, I. & Melino, G. Regulation of adult neurogenesis in mammalian brain. *International Journal of Molecular Sciences* **21**, 1–21, https://doi.org/10.3390/ijms21144869 (2020).
9. Lam, M. *et al*. Single-cell study of neural stem cells derived from human iPSCs reveals distinct progenitor populations with neurogenic and gliogenic potential. *Genes to Cells* **24**, 836–847 (2019).

10. Artegiani, B. *et al*. A Single-Cell RNA Sequencing Study Reveals Cellular and Molecular Dynamics of the Hippocampal Neurogenic Niche. *Cell Rep.* **21**, 3271–3284 (2017).
11. Zhang, Q. *et al*. Single-cell analysis reveals dynamic changes of neural cells in developing human spinal cord. *EMBO Rep*. **22** (2021).
12. Blanco-Luquin, I. *et al*. NXN Gene Epigenetic Changes in an Adult Neurogenesis Model of Alzheimer's Disease. *Cells* **11** (2022).
13. Stevanovic, M. *et al*. SOX Transcription Factors as Important Regulators of Neuronal and Glial Differentiation During Nervous System Development and Adult Neurogenesis. *Frontiers in Molecular Neuroscience* **14**, 654031, https://doi.org/10.3389/fnmol.2021.654031 (2021).
14. Wang, Y. *et al*. SOX2 is essential for astrocyte maturation and its deletion leads to hyperactive behavior in mice. *Cell Rep*. **41** (2022).
15. Satir, T. M. *et al*. Accelerated neuronal and synaptic maturation by BrainPhys medium increases Aβ secretion and alters Aβ peptide ratios from iPSC-derived cortical neurons. *Sci. Rep*. **10** (2020).
16. Zheng, G. X. Y. *et al*. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun*. **8** (2017).
17. Dobin, A. *et al*. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
18. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
19. Tirosh, I. *et al*. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (1979)*. **352**, 189–196 (2016).
20. Nault, R. *et al*. Benchmarking of a Bayesian single cell RNAseq differential gene expression test for dose-response study designs. *Nucleic Acids Res.* **50**, E48 (2022).
21. Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255–261 (2018).
22. Ritchie, M. E. *et al*. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
23. Squair, J. W. *et al*. Confronting false discoveries in single-cell differential expression. *Nat. Commun*. **12** (2021).
24. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
25. Yu, G., Wang, L. G., Yan, G. R. & He, Q. Y. DOSE: An R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608–609 (2015).
26. Hokama, M. *et al*. Altered expression of diabetes-related genes in Alzheimer's disease brains: The Hisayama study. *Cerebral Cortex* **24**, 2476–2488 (2014).
27. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11** (2010).
28. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: Precision Weights Unlock Linear Model Analysis Tools for RNA-Seq Read Counts. *Genome Biology* **15**, http://genomebiology.com/2014/15/2/R29 (2014).
29. Ayhan, F., Douglas, C., Lega, B. C. & Konopka, G. Nuclei isolation from surgically resected human hippocampus. *STAR Protoc*. **2** (2021).
30. Satpathy, A. T. *et al*. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
31. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
32. Aibar, S. *et al*. SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
33. Bravo González-Blas, C. *et al*. cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16**, 397–400 (2019).
34. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281–291.e9 (2019).
35. Shang, Z. *et al*. Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation. *Gigascience* **7**, 1–19 (2018).
36. Pilot, F., Philippe, J. M., Lemmers, C., Chauvin, J. P. & Lecuit, T. Developmental control of nuclear morphogenesis and anchoring by charleston, identified in a functional genomic screen of Drosophila cellularisation. *Development* **133**, 711–723 (2006).
37. Blanco-Luquin, I., Martinez de Morentin, X., Mendioroz, M. & Gómez Cabrero, D. *GEO.* https://identifiers.org/geo/GSE307094 (2026).

Acknowledgements

We thank the members of the Neurogenesis and Neurodegeneration Laboratory for discussions and technical assistance throughout the project. We are also grateful to the Flow Cytometry Facility and the Sequencing Core Facility for their support with FACS sorting and single-cell library preparation. We acknowledge the administrative and institutional support provided by the NavarraBiomed research center. This work was supported by the Government of Navarra through two grants from the Department of Industry of the Government of Navarra (PI058 iBEAS-Plus and PI055 iBEAS-Plus). In addition, MM received a grant Programa de intensificación- (LCF/PR/PR15/51100006) founded by "la Caixa" Foundation and Fundación Caja-Navarra, and Contrato de intensificación from the Institute of Health Carlos III (INT19/00029). XMM, DGC, VL, LA, JT, and AM contributions were funded by King Abdullah University of Science and Technology (KAUST).

Author contributions

I.B.L., X.M.M., M.M., D.G.C. conceived the study, designed experiments, and interpreted the data and wrote the manuscript. I.B.L. and M.M. performed experiments. X.M.M. and D.G.C. performed data analysis. Mo.M., D.A., A.M., A.C.G., L.A., V.L., N.R., J.T., F.P. provided relevant intellectual input and edited the manuscript. M.M., D.G.C. obtained funding and supervised the whole project. All authors approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-026-06971-4.

Correspondence and requests for materials should be addressed to I.B.-L., X.M.-d.-M., M.M. or D.G.-C.