PNAS
www.pnas.org

Supplementary Information for
Altered cell and RNA isoform diversity in aging Down syndrome brains.

Carter R. Palmer[a,b,1], Christine S. Liu[a,b,1], William J. Romanow[a], Ming-Hsiang Lee[a], Jerold Chun[a,*]

Affiliations
[a]Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037, USA
[b]Biomedical Sciences Program, School of Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[1]C. Palmer and C. Liu contributed equally to this work.

Corresponding author, Jerold Chun
Email: jchun@sbpdiscovery.org

This PDF file includes:

Supplementary Methods
Supplementary text
Figures S1 to S9
Legends for Datasets S1 to S10
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S10

Supplementary Methods

Tissue sampling and preparation

Frozen tissue samples from Brodmann Area (BA) 8 or 9 of the prefrontal cortex were obtained from multiple sources and stored at -80ºC. Brain Bank Tissues sources included: The University of Maryland (N=16), Emory University (N=5), The London Neurodegenerative Diseases Brain Bank (N=3), The Newcastle Brain Tissue Resource (N=3), and the Southwest Dementia Brain Bank (N=2). Samples were sectioned in a cryostat set at -20ºC, and serial sections were taken: three 20μm sections for staining, followed by two 300μm sections for nuclei isolation, and one 20μm section for RNA integrity number (RIN) measurement.

RNA integrity measurement

RNA was isolated using an RNeasy isolation kit from Qiagen. RNA was diluted to a concentration of 2 ng/ul and evaluated on an Agilent 4200 TapeStation using an Agilent High Sensitivity RNA Screentape assay. Only samples that had a RIN ≥ 6 were used for sequencing analysis.

Thioflavin S staining

Tissue sections (20μm) were stained using thioflavin S to visualize amyloid plaques and tau tangles as hallmarks of Alzheimer's disease (AD) pathology. Briefly, tissue was placed on slides and fixed in 10% neutral buffered formalin (NBF) for 5 minutes. Slides were then washed with water and stained with 1% Thioflavin S before being washed twice in water and counterstained with DAPI (5μg/ml) for 3 minutes. Slides were washed

again, dehydrated in 70% EtOH for 4 minutes, 95% EtOH for 3 min and 100% EtOH for 1 min, before being submerged in Xylene for 3 min and coverslipped. Slides were imaged at 10X magnification using the BZ-X810 fluorescent microscope from Keyence.

Nissl staining

Tissue sections (20μm) were stained using Cresyl Violet to visualize the cortical layers of each section. Briefly, tissue was placed on slides and fixed in 10% NBF for 5 minutes. Samples were then incubated for 3 minutes at 70% then 95% EtOH, before placing them in 100% EtOH for 20 minutes. Slides were rehydrated with 95% EtOH, then 70% EtOH, for 3 minutes each, rinsed with water, stained for 5 minutes in 0.2% Cresyl Violet Acetate, rinsed again, and submerged in 70% and 95% EtOH for 1 minute each, followed by Xylene for 3 minutes. Slides were then coverslipped and imaged at 10X magnification using the BZ-X810 fluorescent microscope from Keyence.

Nuclei isolation and generation of amplified cDNA libraries

DS and control samples were randomized and processed in groups of four to negate potential batch process variation. Tissue sections (300μm) were removed from frozen storage and immediately submerged in 1 mL of nuclei isolation buffer (20mM Tris, 320mM Sucrose, 5mM CaCl2, 3mM MgAc2, 0.1mM EDTA, 0.1% Triton-X 100, 0.2% RNase Inhibitor)(*1, 2*). Extracted nuclei were washed twice in PBS + 0.25mM EGTA + 1% BSA + 0.2% RNase inhibitors (Takara Bio, Mountain View, CA). They were then suspended in PBSE + BSA + RNase inhibitors + 1.25ug/mL 4',6-diamidino-2-phenylindole (DAPI) (Sigma, St. Louis, MO). FANS was performed on a FACSAria

Fusion (BD Biosciences, Franklin Lakes, NJ) gating out debris from FSC and SSC plots and selecting DAPI+ singlets. Samples were kept on ice until sorting was complete and were immediately processed after sorting. Sorted nuclei were diluted to ~700-1,500 nuclei/mL, and a final concentration was determined using a fluorescent cell counter. The 10X Genomics Single Cell 3' v3 kit was then used to prepare samples targeting 10,000 single nuclei GEMs. The protocol was followed without deviation prior to fragmentation of the cDNA libraries.

Fragmentation of cDNA libraries and short-read sequencing on Illumina NovaSeq

Twenty percent of the pre-fragmented cDNA library was used for short-read sequencing. The standard protocol was followed including fragmentation, barcode adapter ligation, and Illumina adapter ligation. The libraries were sequenced by GENEWIZ using a Novaseq 6000 to an average depth of 333 million reads per sample and 26,682 mean reads per cell.

cDNA preparation and long-read sequencing

Fifty percent of the pre-fragmented cDNA library was used for long-read sequencing. If the cDNA input concentration was too low for Pacific Biosciences (PacBio) library preparation, the cDNA library was re-amplified (Supplemental Table 1) using the same reagents and concentrations as outlined in the 10X Genomics kit protocol. Subsequently, 100ng of cDNA was used in the PacBio procedure for "Preparing SMRTbell libraries using PacBio barcoded overhang adapters for multiplexing amplicons" (part no 101-791-700). The protocol was followed without variation to ligate barcoded SMRTbells to

sample specific cDNA libraries, bind polymerase, and load samples. Express Prep Kit 2.0, Binding Kit 2.1, and sequencing primer v4 were used. Each sample was sequenced in an individual SMRTcell containing 8 million available zero-mode waveguides (ZMWs). An average of 6.003 million polymerase reads were obtained per sample.

Selective amplification of cDNA libraries and subsequent long-read sequencing

Selective amplification of the genes *APP*, *SPP1*, and *BIN1* was pursued using custom designed primers (*SPP1* 5' UTR – CATCGTCGGGACCAGACTCGT, *APP* 5' UTR – TCAGTTTCCTCGGCAGCG, *BIN1* 5' UTR - AAGATCTCCCCGCGCGAGAGC) and the Read 1 primer from the10X Genomics preparation (Read 1 Primer – CTACACGACGCTCTTCCGATCT). The same cDNA libraries used for long-read analysis were linearly amplified with only the 5' UTR primer present.

PCR mix:

10µl Master Mix (Invitrogen Platinum Super Fi II Cat# 12368010)

0.4ul gene specific 5' UTR primer (10µM stock)

5µl cDNA library (starting concentration of 4ng/ul)

4.6µl Nuclease free water

And were amplified using the following protocol:

98°C – 30 sec

98°C – 5 sec

60°C – 10 sec

72ºC – 3.5 min

X20 cycles

Read1 Primer was added (0.4µl 10µM stock) and 20 (*SPP1*) or 25(*APP* and *BIN1*) cycles were run to amplify the genes

98ºC – 5 sec

60ºC – 10 sec

72ºC – 3.5 min

X20 or 25 cycles

72ºC – 5 min

4ºC Hold

Samples were cleaned with Pronex beads and were sequenced with PacBio Sequel II as outlined above.

Short-read snRNA-seq data processing and filtering

10X Genomics CellRanger software (v3.0.2) was used to demultiplex samples, align reads, quantify unique molecular identifiers (UMIs), and generate cell count matrices. Default parameters were used, with the exception of a pre-mRNA reference file (ENSEMBL GRCh38) to capture intronic reads originating from pre-mRNA species present in the nuclei.

Using Seurat (v3.0.3), sample matrices were filtered to remove nuclei with fewer than 300 genes expressed, greater than 1% of UMIs mapping to mitochondrial RNA, and

number of UMIs exceeding a cutoff set by calculating the interquartile range of UMIs detected and determining outliers. Matrices were normalized by the default global-scaling method in Seurat.

Clustering and UMAP visualization

Lake *et al.*'s (*2*) dataset was used as a reference with Seurat's TransferData function to label cell types in our samples. Seurat objects from the samples in the same disease/age group were merged (Seurat merge function). For comparisons between two groups, differential expression analysis and pseudotime analysis, merged samples within a group were integrated (Seurat IntegrateData function). For example, all DS-young samples were merged, all Ctrl-young samples were merged together, and the resulting merged samples, DS-young merged and Ctrl-young merged, were then integrated for further analysis. The integrated data was then scaled and UMAP embeddings were generated using the top 30 principal components. Additionally, cell types were subset from the larger datasets and independently clustered to identify any subclusters that were specific to age or disease state.

GAD67/NeuN staining

Tissue sections (20μm) from DS-young and age matched Ctrl-young samples were placed on slides and fixed in 10% neutral buffered formalin (NBF) vapor for 5 minutes. Slides were washed in wash buffer (1x PBS + 0.2% TritonX-100) three times. Tissue samples were then blocked in wash buffer + 2% bovine serum albumin (BSA) for 2 hours at room temperature. Slides were incubated in primary antibody (GAD67 (Novus NBP2-79803,

1:100), NeuN (Millipore MABN140, 1:1,000)) in wash buffer + 2% BSA overnight at 4ºC. Slides were again washed 3 times before incubation in secondary antibody (AF-488 Goat anti-mouse (ThermoFisher, A11029), AF-647 Goat anti-rabbit (Thermo Fisher, A-20991)) diluted in wash buffer + 2% BSA for 2h at room temperature. After 3 washes, slides were stained with DAPI and coverslipped. Slides were imaged at 40X magnification using the BZ-X810 fluorescent microscope from Keyence. After imaging, all images were matched in size, and for each section, a clear region spanning from white matter to the pial surface was identified and the rest of the image was removed. Images were equally magnified, and a matching grid was overlayed utilizing ImageJ. Cells were counted in three separate rows spanning from white matter to pial surface.

Multiple linear regression of inhibitory:excitatory ratios

For DS and control cohorts, data on sex, RIN, age, and DS vs control status were collected. Data was input into tables and Prism was utilized to calculate a multiple linear least squares regression for which the independent variable was ex:in ratio. Sex and DS status were assigned binary indicator variables. No weighting was utilized and no 2-way or 3-way interactions were accounted for.

Differential Gene Expression (DEG) analysis

Seurat was used to identify differentially expressed genes (DEGs) in DS compared to control samples by cell type and between age groups. Default parameters for FindMarkers were used to identify DEGs that were expressed in at least 10% of either of

the populations being compared, had at least a 0.25 log fold difference, and were significant based on a Wilcoxon Rank Sum test.

Gene enrichment analysis

Gene Ontology (GO) analysis was conducted using PANTHER (*3-5*). All DEGs for a specific comparison were loaded into the PANTHER interface and "biological process" was selected. The -log10 of the false discovery rate (FDR) was used to assess significance enrichment of DEGs.

Pseudotime analysis

Count matrices and UMAP projections of specific cell types from Seurat analysis were loaded into Monocle3 (v0.2.1). Cells were partitioned, and pseudotime trajectories were learned and plotted. Endpoints that clustered with the youngest samples' cells were chosen as the roots for each graph. Differential expression analysis was completed to determine which genes had expression that varied as a function of pseudotime. Astrocytes separated into two partitions that were analyzed individually.

Processing of long reads and isoform calling

Samples from both untargeted and targeted long-read datasets were demultiplexed and barcodes were removed using lima (v1.10.0). Following the recommendations in the cDNA_Cupcake repository (version updated 02/07/2020) for single-cell isoform analysis, CCS reads were generated using ccs (v4.2.0) with the following parameters: --minPasses 1 --min-rq 0.8 --minLength 50 --maxLength 21000. 10X Genomics R1 and TSO primer

sequences and reads with improper primer orientation were removed using lima with the parameter --isoseq. UMIs and cellular barcodes were identified for each read. Isoseq3 refine (v3.2.2) was used to remove poly-A tails and artificial concatemers before mapping to the human reference genome (GRCh38). cDNA_Cupcake's (v9.0.1) collapse_isoforms_by_sam.py was used to collapse redundant isoforms. SQANTI2 (v7.3.2) was used to filter out mono-exon isoforms and artifacts of intra-priming and annotate the identified isoforms. Scripts from cDNA_Cupcake were used to assign UMIs/barcodes and isoforms back to specific reads. Original scripts were written to match specific reads back to sample and cell type, summarize which samples each isoform was detected in, and visualize the resulting isoforms in UCSC Genome Browser.

Differential isoform expression and usage analysis

The protocol for using tappAS (v0.99.15) for "Data from Long-read Sequencing Technology" was followed. Briefly, IsoAnnotLite (v1.2) was used to generate a tappAS gff3 reference file from the SQANTI2 output; the gff3 reference file was loaded into tappAS along with the UMI counts matrix and experiment design file. Differential expression analysis (DEA) and differential isoform usage (DIU) were both run on transcripts using default parameters.

RNAscope for microglial gene markers

Sections of tissue (20μm) were cut and adhered to slides. Sections were fixed for 15 minutes in pre-chilled 10% Neutral Buffered Formalin and then washed in 1x PBS at room temperature for 2 minutes. Samples were then dried and processed using the

recommended kit protocol (2.5 HD Duplex Assay, Advanced Cell Diagnostics 322500). Probes applied were *C1QA* (485451-C2) and *CX3CR1* (411251). Slides were imaged at 40X magnification using the BZ-X810 fluorescent microscope from Keyence.

Western blot analysis

Sections (200 mm) of 5 control and 5 DS brains (including 1 with a RIN below the cutoff for sequencing analysis) were lysed in RIPA buffer containing 1x Roche cOmplete Protease Inhibitors, 1x Roche PhosSTOP phosphatase inhibitors and 1mM DTT. Extracts were incubated in Invitrogen LDS sample buffer containing DTT for 30 minutes at room temperature, separated on an Invitrogen Bolt 4-12% Bis-Tris protein gel and transferred to a PVDF membrane. The blot was probed with antibodies to CX3CR1 (Invitrogen #14-6093-81) and GAPDH (Invitrogen #AM4300) and visualized using a LI-COR Biosciences CLx Imager. Bands were quantitated using the LI-COR Image Studio Lite software.

Clustering long reads

A gene-cell matrix for one sample (DSY2) was generated by summing up the UMIs for all the isoforms (including mono-exon isoforms) of each gene for every identifiable cell barcode. Cellular barcodes with fewer than 50 UMIs were removed from the matrix. Seurat was used to normalize counts, cluster cells, and transfer cell type labels from the Lake *et al*. reference dataset. Cell type predictions were compared to the cell type assignments from short-read data for DSY2. Additionally, accuracy of predictions for cellular barcodes with 50+ UMIs and 100+ UMIs were compared.

Supplementary Text

Identification of cell types via long reads

The use of only long-read data for cell type identification was assessed. UMI counts for all isoforms of every gene were summed, and the gene count was generated for each unique cellular barcode in a sample. This resulted in a gene-cell matrix derived entirely from long reads that was analyzed by Seurat and utilized the same key genes as the short-read analysis. Using a cutoff of 50UMI/cell, we identified 2,709/3,779 (72%) of cell types accurately when using the cell-type assignments from short-read sequencing analysis as a reference. A 100UMI/cell cutoff, which was reached in 3,197 of the 7,791 cells analyzed (Fig. S9A), resulted in 2,258/2,612 (86.4%) of cells called accurately (Fig. S9B-C). This effort displays the potential to utilize only long-read sequencing to profile and cluster single-nucleus cDNA libraries.

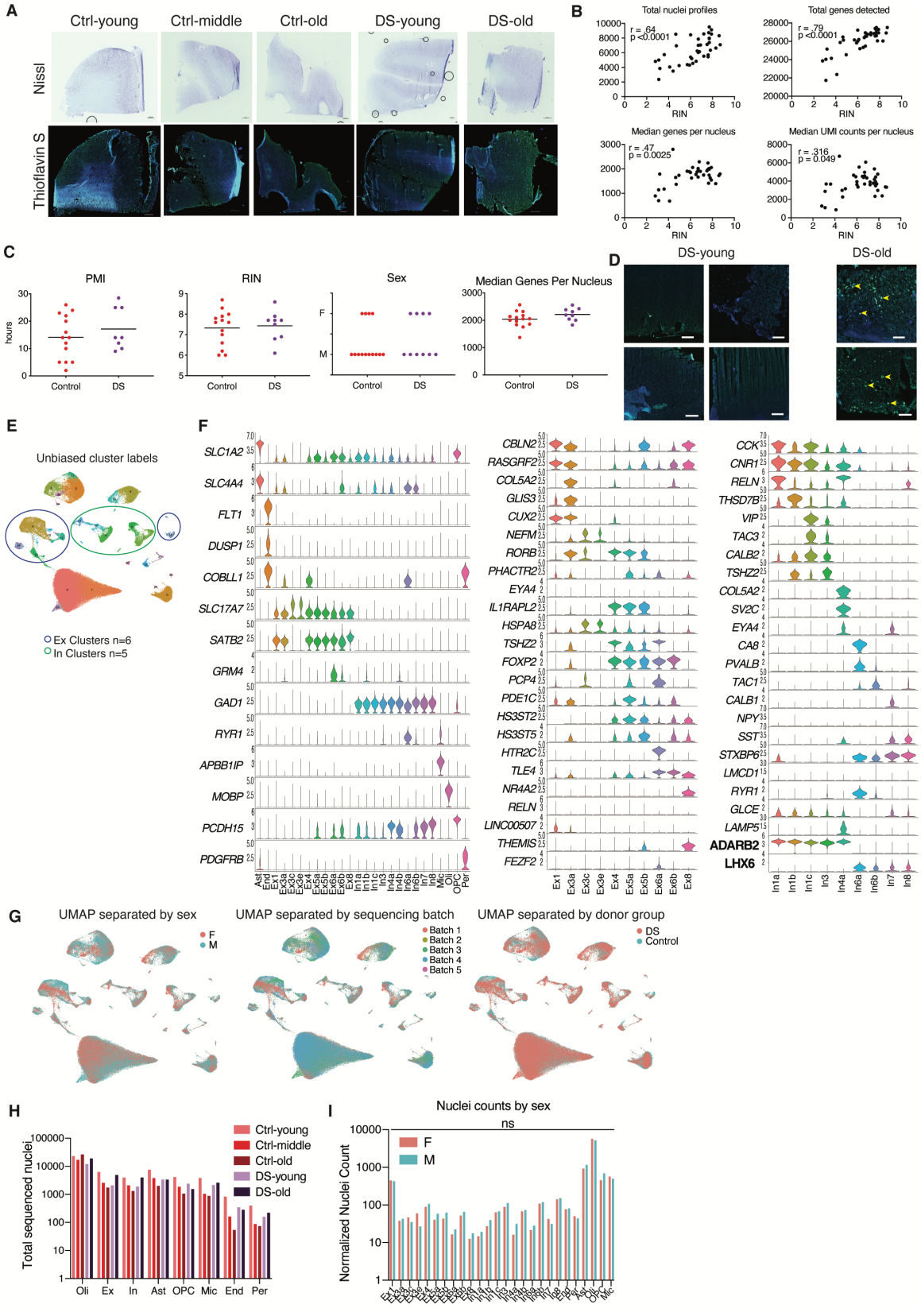

Fig. S1. Sampling and clustering details for sequenced tissues. (**A**) Representative

images from each group stained with Nissl (cell bodies) or Thioflavin S (plaques). Scale bars, 1 mm. (**B**) Effect of RIN on key snRNA-seq quality parameters with measured Pearson correlation coefficient (r) and statistical significance (p). (**C**) PMI, RIN, sex, and median genes per nucleus for DS vs control. (**D**) Detailed Thioflavin S staining on DS samples. Scale bars, 500 μm. (**E**) Unbiased clustering from Seurat and the number of excitatory or inhibitory clusters. (**F**) Violin plots displaying expression of key genes from Lake *et al.* (*2*) used to identify cell types and neuronal subtypes. (**G**) UMAPs of samples separated by sex, sequencing batch, and DS vs control. (**H**) Total number of cells clustered into each major cell type for each group. (**I**) Normalized cell counts for each cell type for male and female samples. No sex differences are significant as measured by t-test.

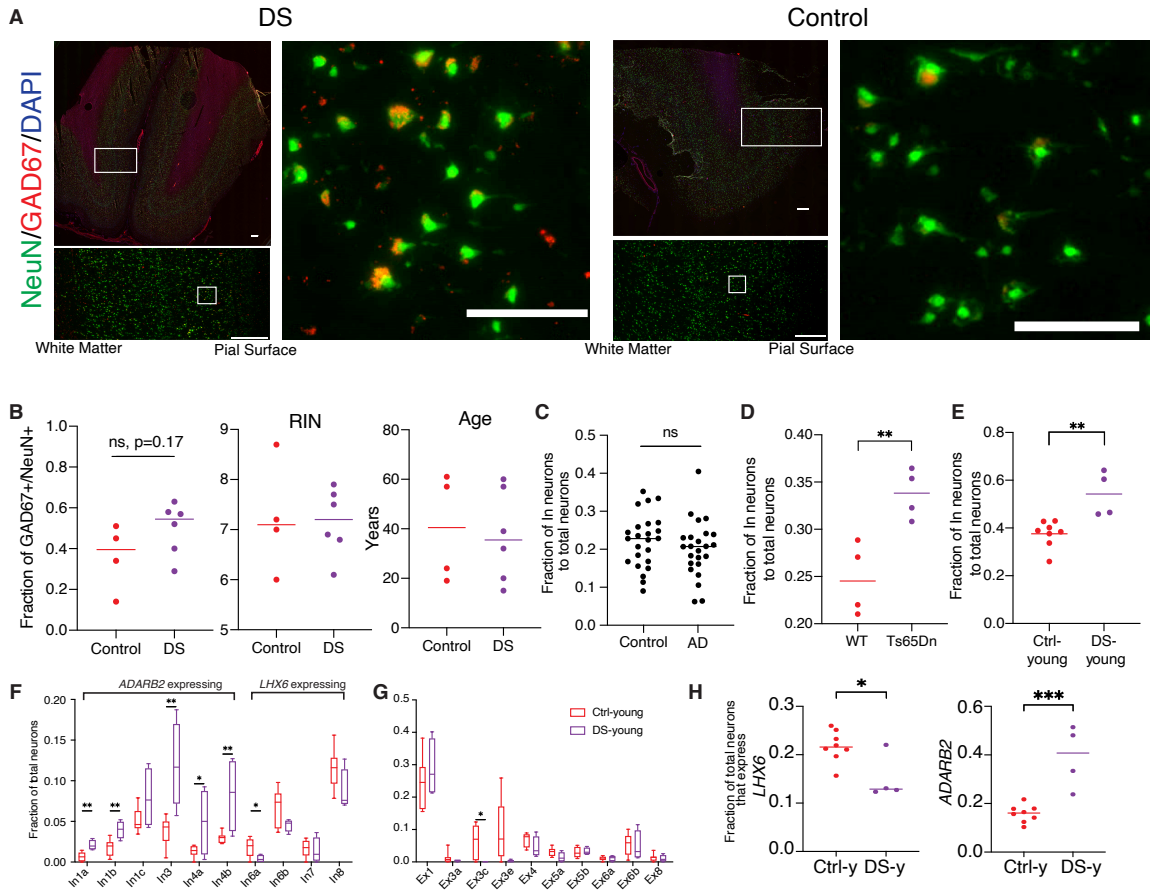

Fig. S2. Confirmation of altered neuronal fractions in DS. (**A**) Representative staining of DS and control brains. Scale bars – 100µm in highest magnification, 500µm in others. DAPI channel has been excluded from magnified images to aid in visualization. (**B**) Fraction of total NeuN+ cells that were also GAD67+, as well as RINs and ages of the samples analyzed by immunofluorescence, p=0.17 from unpaired t-test. (**C**) Fraction of total neurons that cluster as inhibitory (In) from previously published dataset in Alzheimer's disease (AD) (*6*) (not significant as determined by unpaired t-test $p = 0.3$). (**D**) Fraction of total neurons identified as inhibitory in WT and Ts65Dn mouse models of DS. Data from previously published scATAC-seq data (*7*). (**E**) Fraction of total neurons identified as inhibitory in control (Ctrl) and DS-young brains. (**F** and **G**) Fraction of inhibitory (F) and excitatory (Ex) (G) neuronal subtypes in Ctrl and DS-young brains. For

(E and F) boxes extend from the 25th to 75th percentiles and whiskers are the minimum to maximum values. (**H**) Fraction of inhibitory neurons that expressed *LHX6* or *ADARB2*. For (C-H) asterisks denote statistical significance in unpaired t-test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

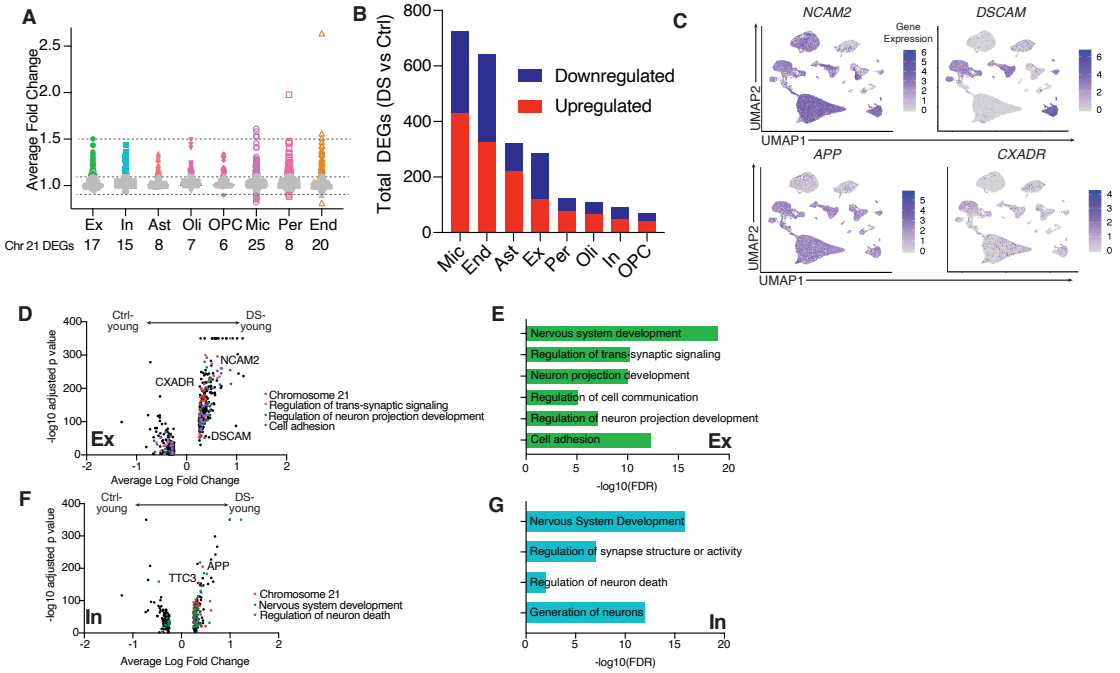

Fig. S3. Details of altered gene expression in DS-young samples. (A) Average fold change in expression for genes on human chromosome 21 (HSA21) in each detected cell type. Gray color signifies an absolute fold change less than 1.1. **(B)** Differentially expressed genes (DEGs) from the full transcriptome detected in each cell type categorized as up or downregulated in DS compared to controls. (**C**) Feature plots displaying expression of key HSA21 genes for complete DS and Ctrl clustering. **(D)** Volcano plots for total DEGs in excitatory neurons for DS-young vs Ctrl-young groups. (**E**) Key biological processes determined by Gene Ontology (GO) analysis for excitatory neuronal DEGs of DS-young vs Ctrl-young. **(F)** Volcano plots for total DEGs in inhibitory neurons for DS-young vs Ctrl-young groups. (**G**) Key biological processes determined by GO analysis for inhibitory neuronal DEGs of DS-young vs Ctrl-young.

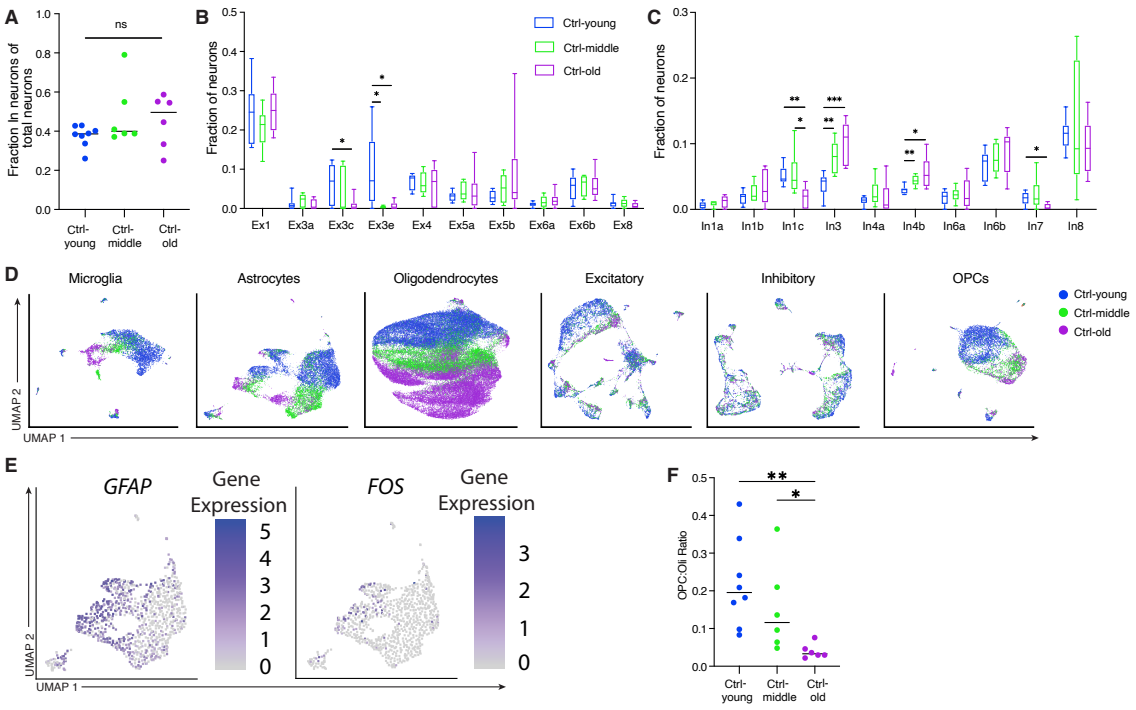

Fig. S4. Changes in neuronal subtypes and cell-type-specific clustering during aging. (**A**) Proportion of total neurons identified as inhibitory (In) across control (Ctrl) age groups. Non-significant averages as determined by ANOVA ($p = 0.326$). (**B** and **C**) Proportions of excitatory (Ex) (B) and inhibitory (C) neuronal subtypes in Ctrl brains. For (B and C) boxes extend from the 25th to 75th percentiles and whiskers extend from minimum to maximum values. Asterisks denote statistical significance in unpaired *t*-test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). (**D**) Unsupervised clustering of major cell types in Ctrl datasets. (**E**) *GFAP* and *FOS* expression in Ctrl astrocytes. (**F**) Ratio of OPCs to oligodendrocytes across Ctrl age groups.

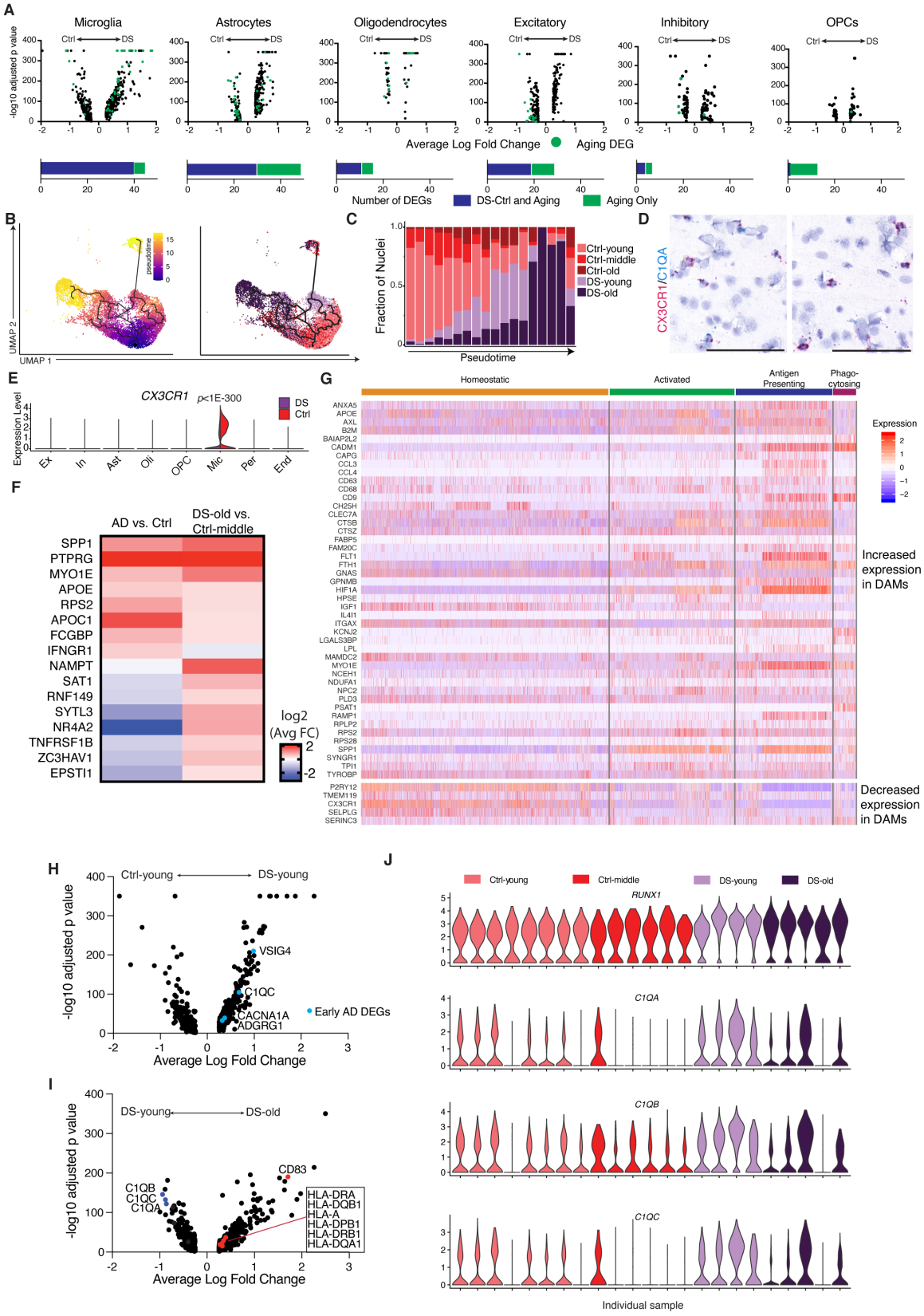

Fig. S5. Microglial transcriptomic alterations in DS. (**A**) Volcano plots for total

differentially expressed genes (DEGs) in DS vs control (Ctrl) for each cell type and bar graphs displaying quantity of DEGs for each cell type. (**B**) Unsupervised pseudotime trajectories with cells colored by pseudotime assignment (left) and cohort (right). (**C**) Fraction of nuclei that comprise discrete intervals of pseudotime. (**D**) RNAscope of tissue sections for *CX3CR1* (red) and *C1QA* (blue). Scale bar 100μm (**E**) Cell-type-specific expression of *CX3CR1* in DS vs Ctrl brains. (**F**) Heatmap of microglial DEGs from a previously published dataset in Alzheimer's disease (AD) (*6*) comparing expression changes in AD vs Ctrl to DS-old vs age-matched Ctrl-middle. (**G**) Heatmap of DAM (*8*) differentially expressed genes and expression levels in DS and control microglia profiled here. (**H**) Volcano plot of Ctrl-young vs DS-young microglia showing AD DEGs identified in (*6*) overlapping with DS-young dataset. (**I**) Volcano plot of DS-young vs DS-old microglia with highlighted genes of interest. (**J**) Sample-specific microglial expression of *RUNX1* and C1q components.

Fig. S6. Categorization of Novel Not in Catalog (NNC) variants using reads from the gene *PTGDS*. **(A)** Samples utilized for long-read sequencing (squares). **(B)** Filtering process for long reads and the read numbers remaining after each step. **(C)** Example of isoforms with a novel exon. **(D)** Examples of isoforms with an intra-exonic junctions (IEJs). **(E)** Examples of isoforms with intron retention. (C-E) Reads are derived from the untargeted dataset. The full gene annotation is aligned below.

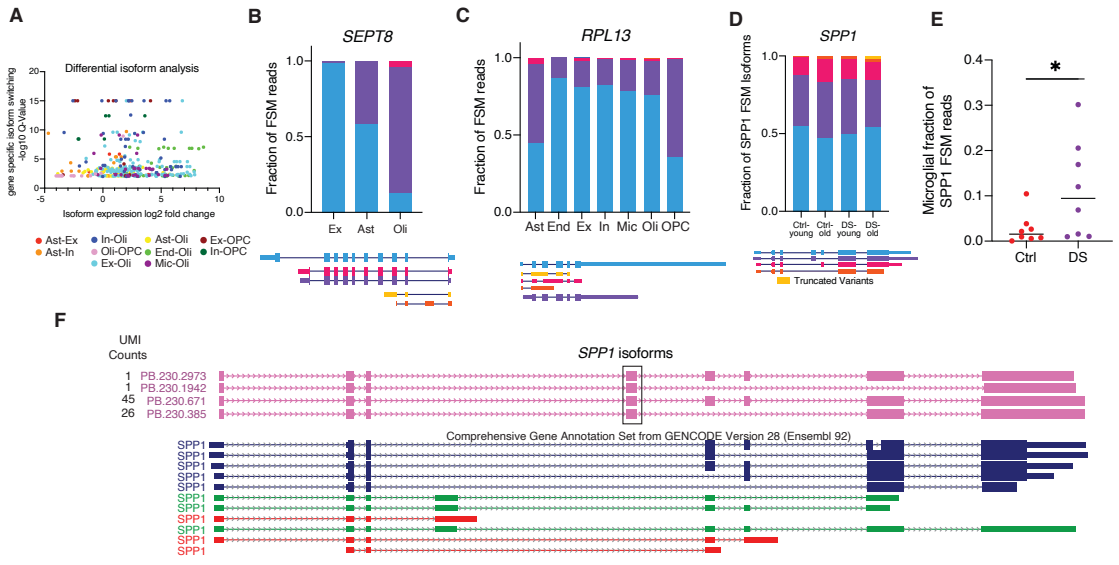

Fig. S7. Differentially utilized isoforms and variation in *SPP1*. (A) Plot of significantly changed isoforms between multiple cell types. Q-value represents the false discovery rate for which at least one isoform of a particular gene displays differential proportionality across compared groups, signifying changes in isoform usage at the gene level. Only isoforms with <0.01 Q-value and non-zero expression in both cell types are displayed. (**B-C)** Cell-type-specific isoform fractions for cell types with over 50 unique reads mapping to full splice match (FSM) isoforms for *SEPT8* (B) and *RPL13* (C). (**D)** Cohort-specific FSM isoform fractions for *SPP1*. (**E)** Fraction of *SPP1* FSM reads originating in microglia for each sample in DS and control (Ctrl) cohorts (unpaired t-test, $p = 0.04$). (**F)** Isoforms including a novel exon sequenced in *SPP1*, including total unique molecular identifier (UMI) counts for each novel isoform. Reads are from the *SPP1*-targeted dataset.

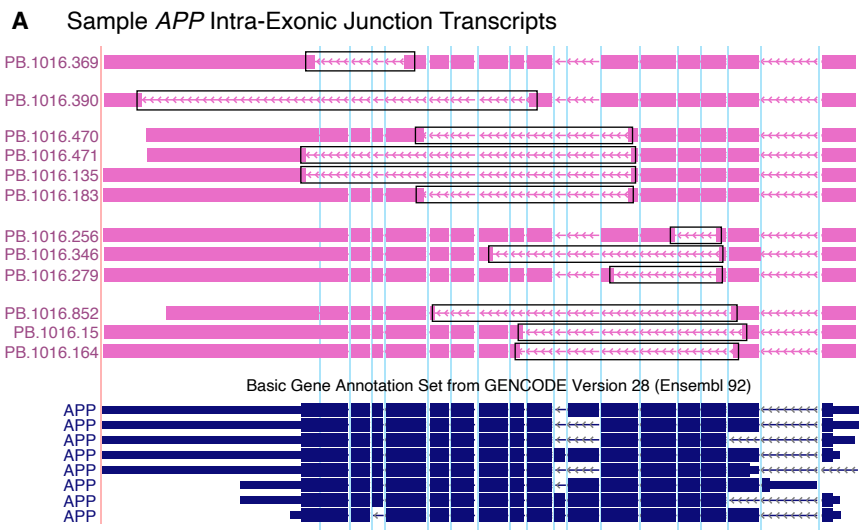

Fig. S8. *APP* isoforms with intra-exonic junctions (IEJs).

(A) *APP* isoforms with IEJs (boxes) identified using targeted long-read sequencing. Light blue vertical lines indicate intronic sequences.

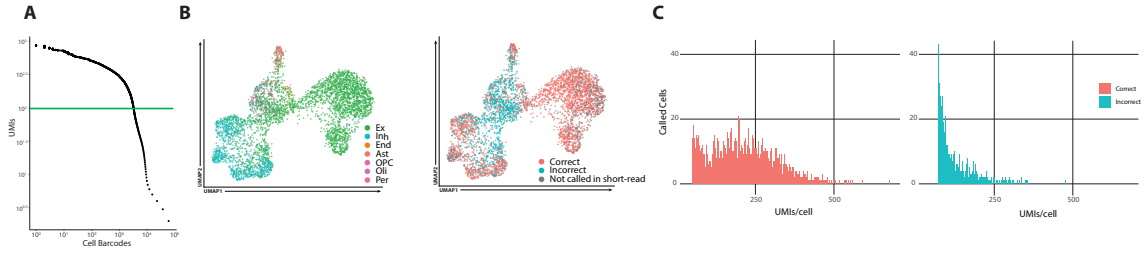

Fig. S9. Cell clustering based on long-read isoform sequencing

(A) Knee plot for cell inclusion based on UMI count from only long-reads for representative processed sample. Green line represents 100 UMI/cell threshold. **(B)** UMAP showing clustering of all cells (50 UMI/cell cutoff) from representative sample from long-read clustering, displayed by cell type and accuracy of cell type identification as compared to short-read sequencing cell type assignment from the same cDNA library. **(C)** Cell type identification accuracy as a function of UMIs/cell.

Dataset S1 (separate file) Sample information for DS and Control brains

Dataset S2 (separate file) Number of cells in each cluster identified per sample

Dataset S3 (separate file) Differentially expressed genes in DS vs Control samples. Each tab corresponds to a particular cell type

Dataset S4 (separate file) Differentially expressed genes in DS-young vs Control-young samples. Each tab corresponds to a particular cell type

Dataset S5 (separate file) Differentially expressed genes in DS-old vs Control-middle samples. Each tab corresponds to a particular cell type

Dataset S6 (separate file) Differentially expressed genes in Control-old vs Control-young samples. Each tab corresponds to a particular cell type

Dataset S7 (separate file) Differentially expressed genes in Control-old vs Control-middle samples. Each tab corresponds to a particular cell type

Dataset S8 (separate file) Differentially expressed genes in Control-middle vs Control-young samples. Each tab corresponds to a particular cell type

Dataset S9 (separate file) Differentially expressed genes over pseudotime

Dataset S10 (separate file) Genes with at least one transcript with an IEJ identified through long-read sequencing

References

1. M. H. Lee *et al.*, Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **563**, 639-645 (2018).
2. B. B. Lake *et al.*, Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* **36**, 70-80 (2018).
3. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330-d338 (2019).
4. M. Ashburner *et al.*, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
5. H. Mi, A. Muruganujan, D. Ebert, X. Huang, P. D. Thomas, PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* **47**, D419-d426 (2019).
6. H. Mathys *et al.*, Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).

7. R. Spektor, J. W. Yang, S. Lee, P. D. Soloway, Single cell ATAC-seq identifies broad changes in neuronal abundance and chromatin accessibility in Down Syndrome. *bioRxiv*, 561191 (2019).
8. H. Keren-Shaul *et al.*, A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290.e1217 (2017).