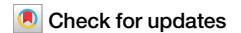


<https://doi.org/10.1038/s42003-026-09759-9>

RNA isoform diversity, splicing variants and switching in single cells of the Alzheimer's disease brain



Anis Shahnaee^{1,2}, Christine S. Liu^{1,2}, Tony Ngo¹, Carter R. Palmer¹, Derya Ziomek¹, Chris Park¹, Valerie P. Tan¹, Natalia L. Jimenez¹ & Jerold Chun¹ ✉

Alzheimer's disease (AD) is the most common cause of dementia, yet its molecular causes remain incompletely understood. RNA diversity in part arising from dysregulated splicing may contribute to AD pathogenesis, however the ability to interrogate the resulting full-length isoforms in single brain cells has been technologically limited. Here we report the use of PacBio Kinnex long-read sequencing combined with 10X Genomics single-cell preparations to identify both annotated and unannotated RNA isoforms. Eight AD and seven non-diseased post-mortem human brains yield ~70,000 single nuclei showing diverse, differentially expressed and switched transcripts in multiple genes. Cell-type-specific isoform expression and variants with intra-exonic junctions associated with reverse transcriptase-mediated somatic gene recombination are also detected. Novel isoforms are altered in AD, with examples of *CHI3L1* and *SEPTIN4*. Kinnex detection of RNA isoforms from single nuclei highlights vast isoform diversity amongst brain cell types, representing an under-explored element in AD and other brain disorders.

Alzheimer's disease (AD) is the most common form of dementia¹, currently affecting more than 30 million individuals worldwide². Brain pathological features such as amyloid-beta plaques, neurofibrillary tau tangles, synaptic dysfunction, and neurodegeneration characterize AD^{3–5}, however the fundamental etiology and underlying molecular mechanisms remain incompletely known. Transcriptomic studies are providing new insights into the normal and diseased brain^{6–8}, including through the use of single-nucleus RNA-seq^{9,10} that provides cell-type level resolution^{11,12}. A vast majority of single-cell studies have utilized short-read sequencing that is suboptimal for detecting full-length mRNA isoforms, while providing an important albeit incomplete view of the transcriptome^{13,14}.

Alternative splicing leads to diversification of the transcriptome and the proteome^{15,16}. In the nervous system, the number of splice variants arising from a single gene can be extreme, as illustrated by Down syndrome cell adhesion molecule (*Dscam1* in *Drosophila*) that can generate over 38,000 splice variants^{17,18}, while human genes can also show thousands of isoforms, as demonstrated by Neurexin 1 (*NRXN1*)^{19–21}. Dysregulated splicing has been implicated in neurological disorders, including AD, ALS, and schizophrenia^{22–24}. A further facet of RNA diversity first reported in the AD brain is reverse transcriptase-mediated somatic gene recombination (RT-SGR) affecting the amyloid precursor protein (*APP*) gene. Dysregulation of RT-SGR generates myriad diverse and novel

mRNAs, including those with pathogenic mutations, which can be reverse transcribed and somatically retroinserted into the genome as genomic cDNAs (gencDNAs)^{25–29}. Somatic gencDNAs are analogous to germline processed pseudogenes^{30–34} but can have both coding potential and enormous sequence diversity that is generated somatically within even a single individual.

Long-read sequencing enables full-length transcript identification, including the classification of novel isoforms, splicing events, and disease-related transcript switching patterns³⁵ as successfully applied to bulk tissue^{36–38}. However, bulk sequencing cannot resolve cell-type-specific isoform expression, which requires single-cell approaches. Early efforts to integrate long-read sequencing technologies with single-cell profiling^{29,39,40} were limited by insufficient read depth. PacBio's Kinnex⁴¹ overcomes this limitation through cDNA amplicon concatenation for increased sequencing depth, which was used here to investigate cell-type-resolved RNA isoform expression in single cells of the AD and non-diseased human brain.

Results

Methodology overview and sample characteristics

To identify isoforms and splicing alterations on a cell-type level, we performed single-nucleus short-read RNA-seq in combination with single-

¹Center for Neurologic Diseases, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA. ²These authors contributed equally: Anis Shahnaee, Christine S. Liu. ✉e-mail: jchun@sbpdiscovery.org

nucleus Kinnex long-read sequencing on prefrontal cerebral cortex tissue samples from seven non-diseased (ND) and eight Alzheimer's disease (AD) brains (Fig. 1a). We focused on the prefrontal cortex because this region is among the most vulnerable areas in Alzheimer's disease, showing pronounced synaptic loss and strong correlations with cognitive impairment⁴² as well as significant gray- and white-matter atrophy compared with healthy aging⁴³. Samples were matched for age, sex, and RNA integrity number (RIN) (Mann-Whitney U test, P value > 0.05) (Supplementary Fig. 1a, Supplementary Data 1). We confirmed that tissue sections contained all cortical layers using Nissl staining and that AD brains had evidence of amyloid pathology using thioflavin S staining (Supplementary Fig. 2a). Single-nucleus barcoded cDNA libraries were generated using the 10X Genomics Single Cell 3' v3.1 kit. After barcoding, the amplified cDNA was split into two aliquots for short-read or long-read sequencing. A portion of the amplified cDNA was enzymatically fragmented and used to generate the short-read Illumina libraries for cell-type identification. The other aliquot of the same barcoded cDNA pool was left unfragmented and used to prepare long-read libraries with the Kinnex method and sequenced on the PacBio Sequel II to obtain full-length cDNAs. Kinnex increases sequencing throughput by concatenating cDNA molecules to form a 16-plex array that is sequenced as a single molecule (Fig. 1a). After sequencing, the reads were then split into individual reads representing each barcoded cDNA.

Cell-type-specific gene expression in the Alzheimer's disease (AD) brain

Quality control filtering resulted in over 70,000 cells that were used for short-read sequencing analysis (Supplementary Data 1). Seurat (v4.3) was used to cluster cells and annotate cell types using marker genes¹⁰ (Supplementary Fig. 2b). All major cell types in the brain were identified, including excitatory neurons (Exc), inhibitory neurons (Inh), oligodendrocytes (Oli), astrocytes (Ast), microglia (Mic), oligodendrocyte precursor cells (OPC), pericytes (Per), and endothelial cells (End) (Fig. 1b). PCA and UMAP visualizations colored by sample-level variables—RIN, sex, and disease—indicated that these factors did not drive clustering (Supplementary Fig. 1b, c). We quantified the cell-type proportions in ND and AD brain samples. Traditional statistical analysis showed no significant differences in the proportions of the specific brain cell types between AD and ND samples using the conventional threshold (two-sided t -test, $p > 0.05$; Fig. 1c). We additionally analyzed the cell-type proportions using propeller⁴⁴ and detected no significant shifts in cellular populations between AD and ND. We assessed differential gene expression in AD brains relative to ND across the major cell types, defining differentially expressed genes (DEGs) as having $|\log_2$ fold change > 0.25 and the adjusted p -value < 0.05 using the Wilcoxon rank-sum test with Bonferroni correction (Supplementary Data 2). The number of DEGs varied in each cell type, with the highest numbers identified in astrocytes and oligodendrocytes, with 491 and 465 DEGs, respectively (Fig. 1d). Gene ontology (GO) analysis⁴⁵ revealed cell-type-specific alterations in numerous biological processes in AD brains, which, importantly, were consistent with previous snRNA-seq studies^{12,46} (Fig. 1e–h, Supplementary Fig. 1d,e). For example, gene set enrichment in excitatory neurons included terms associated with “synapse organization”¹² and “glutamate signaling” (Fig. 1e), whereas microglia showed enriched terms “neuroinflammatory response”⁴⁷ and “response to type II interferon”, and astrocytes showed “gliosis”, all consistent with inflammatory processes in the AD brains (Fig. 1f,h). Oligodendrocytes showed changes to “amyloid-beta metabolic processes” (GO:0050435) and “amyloid precursor protein metabolic processes” (GO:0042982) (Fig. 1g), including alterations in *PSEN1*, providing evidence that oligodendrocytes contribute to amyloid-beta production^{48,49}. Terms related to “gliosis”, “forebrain development”, and “organophosphate catabolism” were also enriched in numerous cell types, providing evidence of widespread gene dysregulation. Importantly, these GO terms were consistent with those identified in previously published datasets, providing validation for the selected samples and sequencing efforts and establishing the samples as valid for parallel long-read isoform profiling.

Isoform characteristics

A major limitation of long-read sequencing is the lower depth per sequencing run, including a higher cost per read in comparison to short-read sequencing. Kinnex improves sequencing throughput by $\sim 16\times$ ⁴¹. We therefore used Kinnex and processed each sample using a complete PacBio Sequel II SMRTcell that yielded an average of 2.1 million HiFi reads (each comprising concatenated cDNAs), which were then segmented into an average of 32.5 million reads per sample (on average, 15.62 molecules per concatenated array). Reads were next processed using isoseq, a modified version of SQANTI3³⁹, and isoSeQL⁵⁰ to quantify and annotate isoforms using the GENCODE reference annotation (v44). SQANTI3 isoform categories were assigned to each of the isoforms to distinguish novel isoforms (novel in catalog, NIC; novel not in catalog, NNC) from known isoforms (full-splice match, FSM; and incomplete splice match, ISM which can be produced by incomplete reverse transcription, degradation or use of new transcription start or termination sites (TSS/TTS)) (Fig. 2a). There were 858,171 total isoforms detected (Supplementary Data 3), which were filtered to generate an operational set of “confident” isoforms detected in four or more samples with normalized expression of counts per million (CPM) ≥ 1 (Fig. 2b, c). This confidence cutoff resulted in 53,536 confident isoforms ($\sim 6\%$ of all detected isoforms, supported by $\sim 75\%$ of all reads; Supplementary Data 4). These cutoffs were chosen primarily to filter out isoforms supported by single reads in single samples and to link isoform expression changes to disease by requiring that isoforms detected in only one sample group be expressed in approximately half the samples of that group (4/7 ND samples, 4/8 AD samples). The median length of confident transcripts was 671 nucleotides long (reflecting 10X Genomics library construction limitations), and isoforms ranged from 1–31 exons (Fig. 2d, e). Novel confident isoforms were also identified (3995 isoforms ($\sim 7\%$ of total isoforms) comprising 2133 NIC isoforms and 1862 NNC isoforms (Fig. 2f)). Annotated FSM (20,386) and ISM isoforms (28,906) were a majority of the confident isoforms, although ISM isoforms were not the focus of downstream analyses. Multiple isoforms were identified for expressed genes; notably, *DST* (dystonin gene) had the most with 66 (Fig. 2g).

Inclusion of isoforms that didn't meet the confidence cutoffs would have resulted in massive increases in diversity with an additional 83,927 NIC and 235,157 NNC isoforms supported by an average of 0.977 and 0.611 reads per sample respectively (Fig. 2h–j; Supplementary Data 3). Low-abundance isoforms may arise biologically from cell-type-specific alternative splicing in cell populations or through mechanisms like reverse transcriptase-mediated somatic gene recombination (RT-SGR) producing somatic genomic mosaicism⁵¹. Technical factors like template switching may also contribute to rare isoforms. *APP* and *TARDBP* isoforms with intra-exonic junctions (IEJs) that can arise through reverse RT-SGR events^{25,26,28,29,52} (Fig. 3a,b) were also detected in our dataset, though they were not included in the confident set, reflecting their low prevalence and expression levels. Isoforms with IEJs, some of which were in the confident set, were identified for many other genes (13,618 gene/putative gene sequences; Supplementary Data 3) consistent with prior reports^{26,28,29,52}.

Known and novel isoforms identified in causal and risk-factor AD genes

Multiple genes have been implicated causally or as risk factors in the etiology of AD. Single-nucleotide mutations in *PSEN1*, *PSEN2*, and *APP* cause familial forms of AD (*APP* triplication in Down syndrome also causes AD)^{53,54}, with somatic mutations also occurring in sporadic AD for *APP*²⁵. Genome-wide association studies (GWAS) have identified multiple risk-factor germline gene mutations in sporadic AD. To profile how isoform expression changes could contribute to AD pathology, we interrogated multiple AD genes predicted by GWAS to contribute to AD (Supplementary Data 5). For example, several GWAS have linked mutations in *BIN1* to sporadic AD^{55,56}, and we detected five *BIN1* FSM isoforms and one NNC isoform containing a novel exon and 5' truncation; the novel exon was validated by RT-PCR and detected in all samples (Fig. 4a,b, Supplementary Fig. 3). *BIN1* isoform expression was notably variable across samples

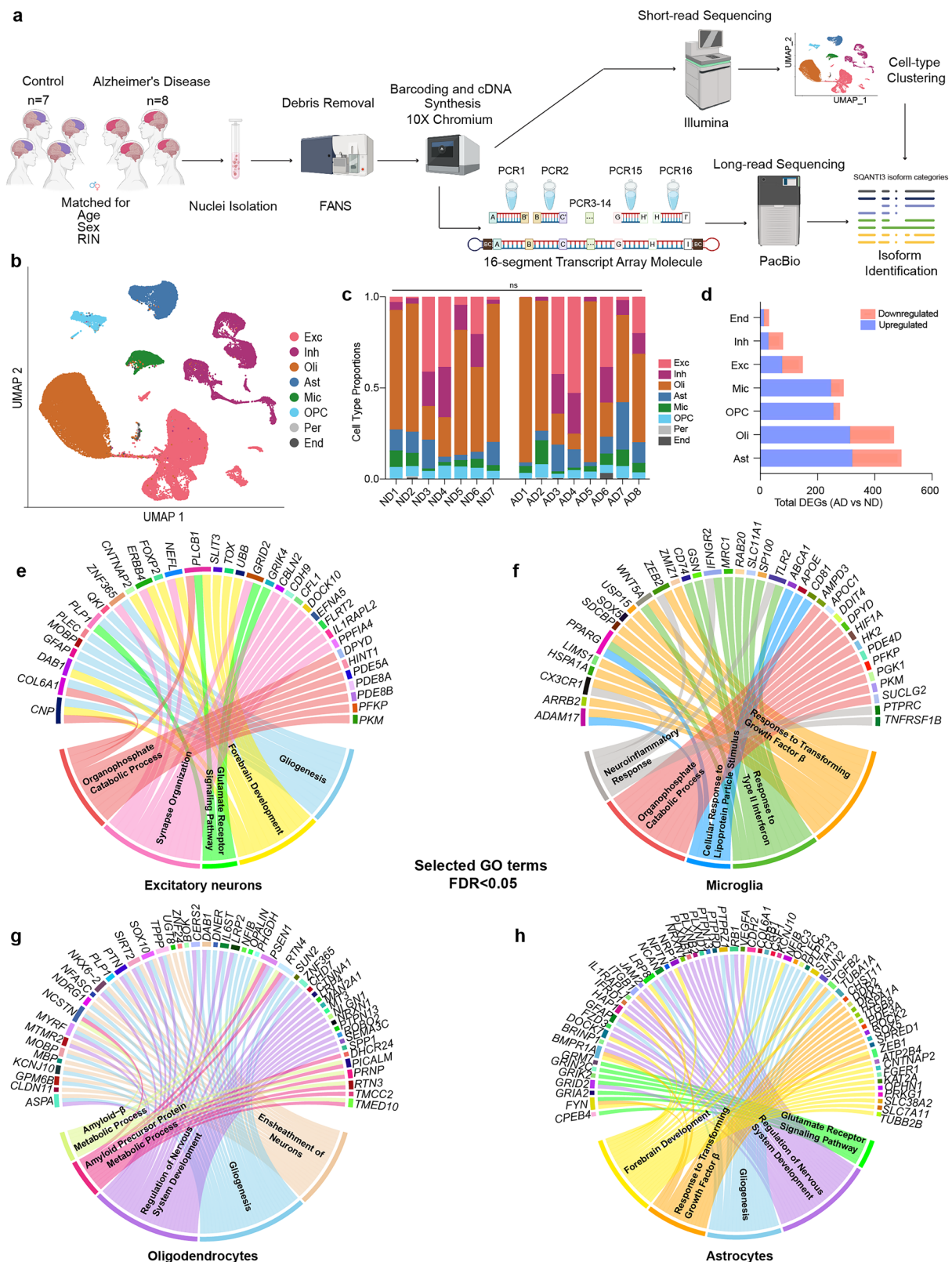


Fig. 1 | Single-nucleus short-read RNA-sequencing and Kinnex long-read sequencing workflow and cell type-specific analyses. **a** Schematic representation of experimental workflow for single-nucleus short-read RNA-sequencing and Kinnex long-read RNA-sequencing. Figure 1a. was partially created with BioRender. <https://BioRender.com/c44n540>. **b** UMAP plot colored by cell type assignments. **c** Bar plot showing the proportions of each cell type in AD and ND samples (ns: not significant; Two-sided t-test). **d** Differentially expressed genes (DEGs) identified across different

cell types in AD and ND samples (absolute log₂ fold change > 0.25 and adjusted *p*-value < 0.05, Wilcoxon Rank Sum test with Bonferroni correction). **e–h** Circos plots of five selected Gene Ontology (GO) terms for excitatory neurons, microglia, oligodendrocytes, and astrocytes with associated genes (false discovery rate (FDR) of <0.05 with the Benjamini-Hochberg (BH) test). Conserved GO terms are shown in the same color.

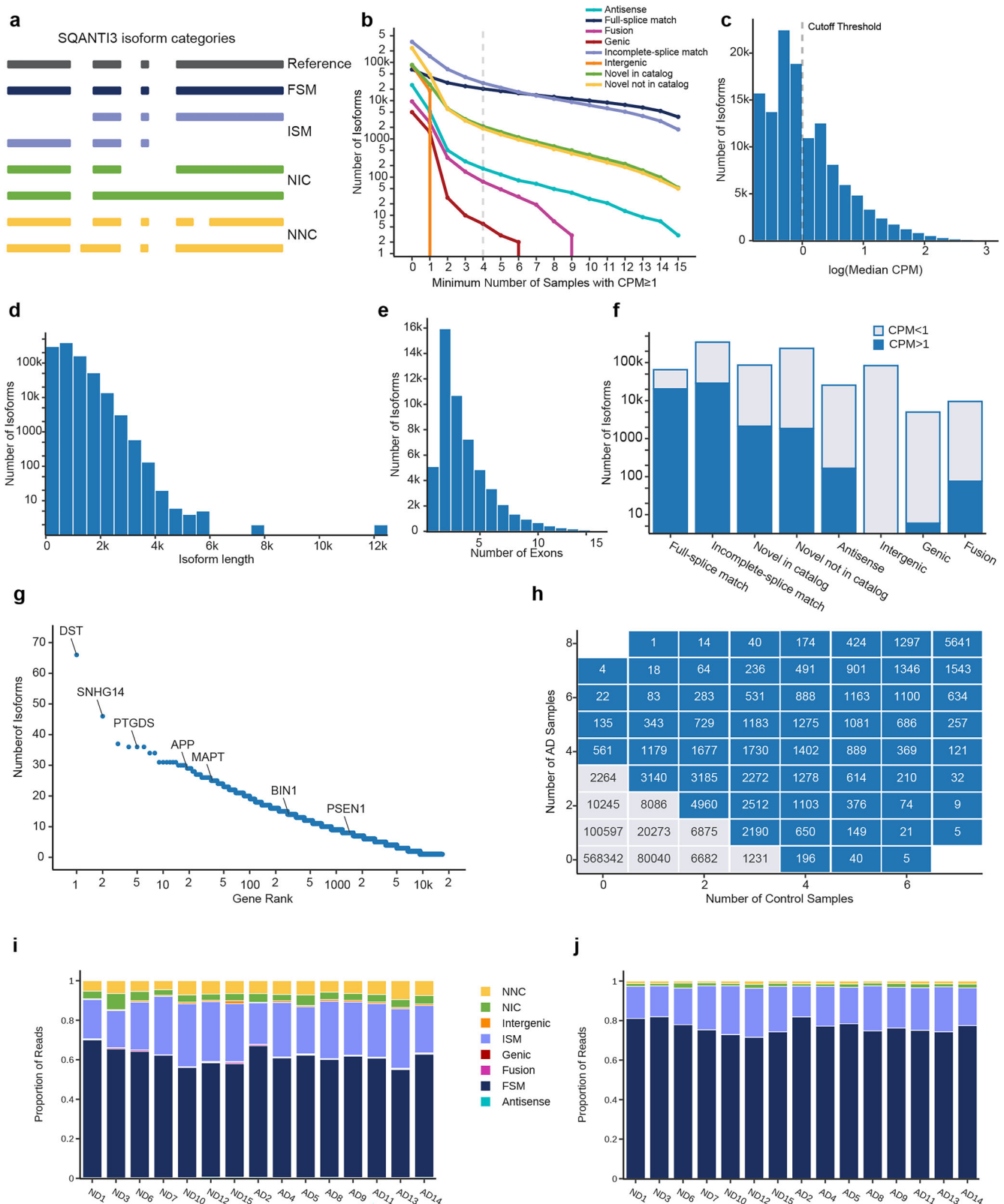


Fig. 2 | Overview of Kinnex long-read sequencing quality control and isoform filtering criteria. a Schematic representation of SQANTI3 categories used for isoform classification. **b** Scatter plot showing the number of isoforms with CPM ≥ 1 in different numbers of samples. **c** Distribution of isoforms by highest CPM cutoff that would meet the four-sample cutoff. **d** Distribution of isoform lengths in the confident Kinnex dataset. **e** Bar plot showing the distribution of the number of exons in each confident isoform. **f** Bar plot showing the number of isoforms in each SQANTI3 category that met or did not meet the confidence cut-off. **g** The number of confident

isoforms identified per gene. **h** The number of isoforms per sample group that had CPM ≥ 1 . Zero samples means that isoforms were detected with CPM < 1. Gray boxes indicate isoforms that did not meet the confidence cut-off. **i** Bar plot showing the proportion of reads assigned to each SQANTI3 category in AD and ND samples without applying the confidence cut-off. **j** Bar plot showing the proportion of reads assigned to each SQANTI3 category in AD and ND samples after the confidence cut-off. These isoforms were used for further analyses.

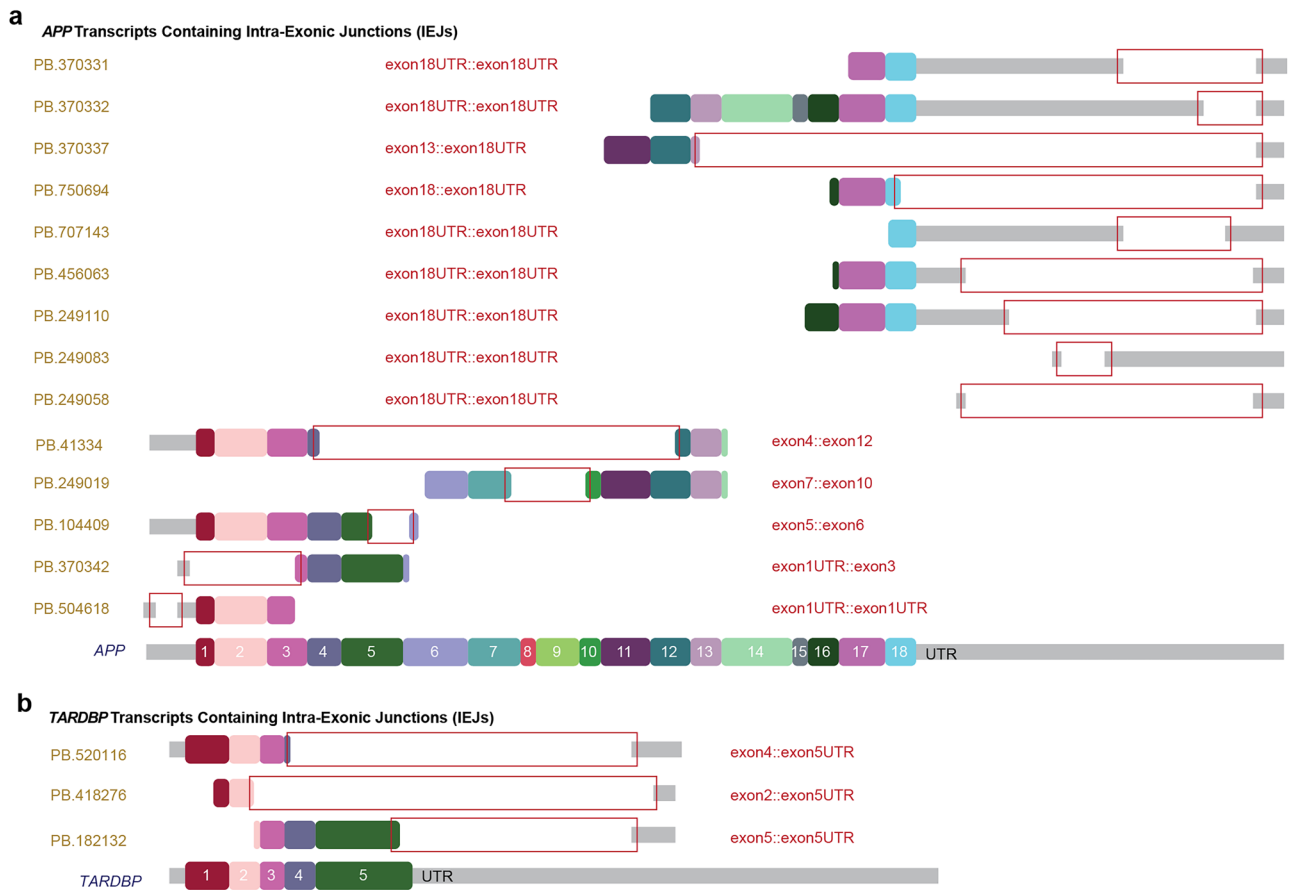


Fig. 3 | Isoforms with intra-exonic junctions. a Novel *APP* transcripts containing intra-exonic junctions (IEJs, boxes). **b** Novel *TARDBP* transcripts containing intra-exonic junctions (IEJs, boxes).

(Fig. 4c) contrasting with results from short-read sequencing that did not report significant variability between AD and ND (Fig. 4d).

In addition to familial and GWAS AD genes, candidate gene approaches identified the well-known AD risk-factor gene *APOE* that has distinct risk levels tied to polymorphic alleles⁵⁷. We detected two *APOE* isoforms: an FSM and a RT-PCR-validated NNC that was missing multiple exons and was expressed at significantly lower levels (Fig. 4e-g, Supplementary Fig. 4). *APOE* gene expression from short-read sequencing showed a distinct cell-type-specific trend across AD and ND brains, wherein AD brains had increased expression in microglia, consistent with neuroinflammation and aging⁵⁸, but decreased expression in astrocytes (Fig. 4h). The different *APOE* gene expression patterns suggested that *APOE* and its isoforms may have distinct roles depending on cell type and disease state.

Previous long-read RNA-seq publications profiling the human brain in AD and healthy controls enabled a comparison to our data^{36,37,59}. A caveat is that our study differed from the prior studies in multiple ways: sequencing technology, library preparation method, exact neuroanatomical location, number of samples, and median sequencing depth (i.e., Heberle et al: 35 million Q10 reads per sample; Leung et al: 322,000 HiFi reads per sample; PacBio: 3.9 million HiFi reads) along with known inter-individual biological variability. Published datasets were bioinformatically processed similarly to our confident set utilizing isoSeQL (Methods) and results compared to one another (Fig. 4i-k). A majority of FSM isoforms from the confident set were detected in at least one other dataset (16,846/20,386; 82.6%), with 4268 isoforms detected in all three (Fig. 4i); 1272 NIC (59.6%) and 847 NNC (45.5%) isoforms were also detected in at least one other dataset (Fig. 4j,k). The overlap confirmed that despite the technical and sample differences, both known and novel isoforms were consistently observed across platforms and samples. In addition, many isoforms unique to each dataset were

observed as well, with 3,540 FSM, 861 NIC, and 1015 NNC isoforms only identified in our confident isoform dataset (Fig. 4i-k).

Differential transcript expression and switching in AD

To examine transcriptomic changes in AD, we assessed differential isoform expression and disease-related switching (via pseudobulk data analyses). Isoform switching analysis identifies the proportion of transcripts for a gene that changes with conditions, since isoform proportions can vary even if overall gene expression remains constant. Reads were aggregated by sample across all cell types, and count matrices were generated using isoSeQL for input into DESeq2 to identify differentially expressed transcripts (DETs) and DEXseq to identify switched isoforms. We identified multiple DETs in AD (Supplementary Data 6), including both known and novel isoforms (Fig. 5a), and four isoforms that were significantly switched for *GTPBP6*, *LUC7L2*, *STX7*, and *SEPTIN4* (Fig. 5b-e; Supplementary Data 7). Of the four isoforms that were switched, three were also down-regulated DETs (*GTPBP6* ENST00000711232.1, *LUC7L2* ENST00000482860.1, and *SEPTIN4* ENST00000317256.10). Interestingly, the switched isoforms' genes were not among the DEGs detected by short-read sequencing (Fig. 5f-i).

For upregulated isoforms in AD, the most significant was an FSM of *TNFRSF25*, ENST00000453260.6 (Fig. 5a). Activation of *TNFRSF25*, also known as death receptor 3, leads to inflammation and apoptosis⁶⁰. The upregulated isoform is truncated on the 5' end relative to the canonical transcript, which utilizes all exons of the gene, and contains a retained intron between the sixth and seventh exons. While the final three exons are identical to the canonical transcript, the transcript is predicted to be non-coding but could still have a regulatory function. This prediction is based on GENCODE annotation indicating the lack of a complete coding sequence, including the absence of the canonical start codon and premature termination codons introduced by the retained intron. Another top upregulated

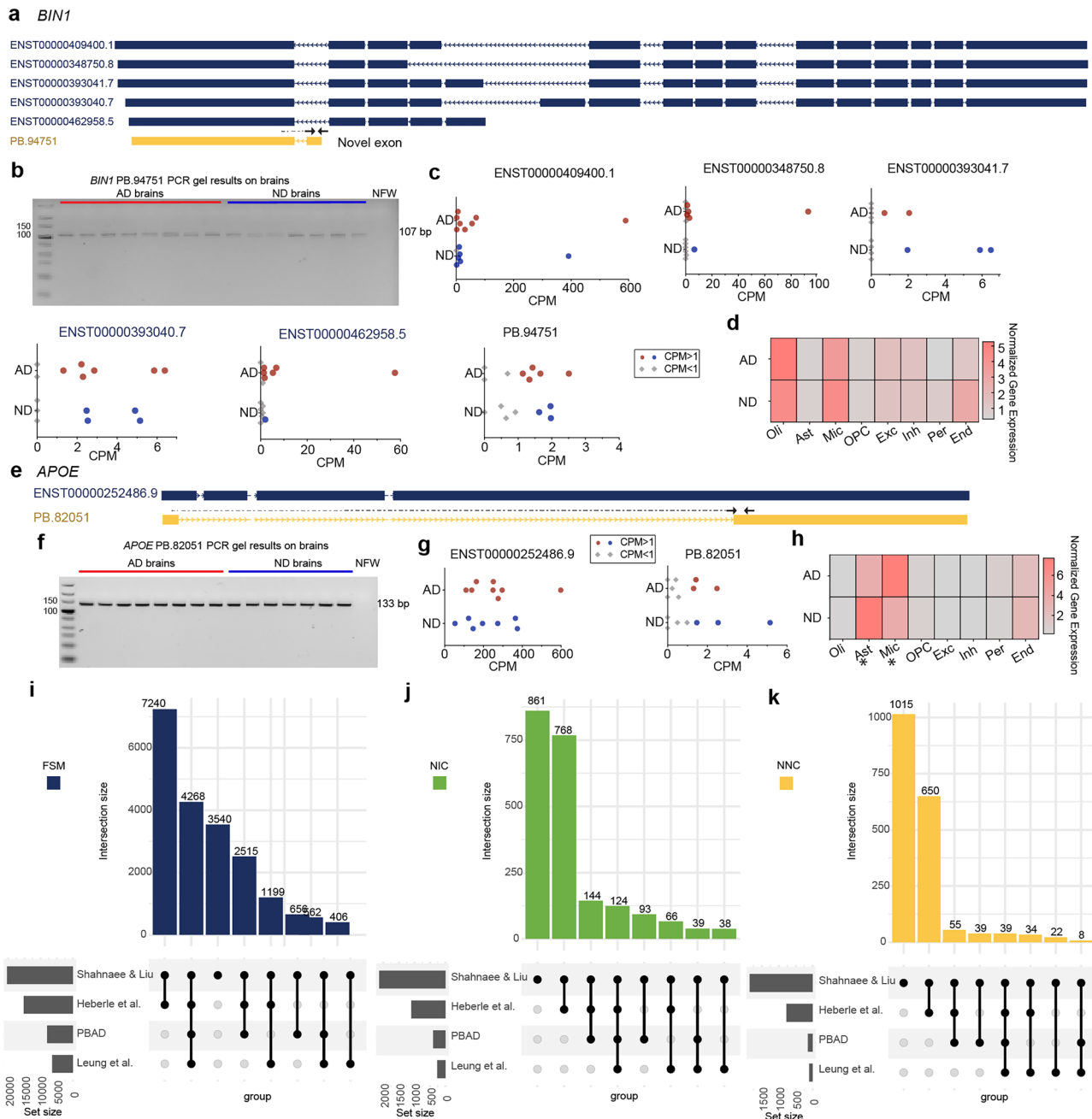


Fig. 4 | Discovery of novel transcripts of AD-related genes via Kinnex sequencing with cross-dataset comparisons. **a** Confident FSM *BIN1* transcripts (navy blue) and novel transcript PB.94751 (yellow). Arrows indicate the locations of the forward and reverse primers used for amplification. Dashed arrow indicates a primer that spans the exon-exon junction. **b** RT-PCR gel validation of the novel isoform, PB.94751. The band at 107 bp confirms that PB.94751 was detected in all 16 samples. Nuclease-free water (NFW) serves as the negative control. **c** *BIN1* isoform expression in individual samples from Kinnex dataset. AD samples with CPM ≥ 1 are represented as red circles; ND samples with CPM ≥ 1 are represented as blue circles; and AD and ND samples with CPM < 1 are represented as gray diamonds. **d** Heatmap showing normalized gene expression of *BIN1* gene across different cell types from short-read sequencing data. Expression was not significantly changed between AD and ND in any cell type (Wilcoxon rank-sum test with Bonferroni correction). **e** Confident FSM

APOE transcript (navy blue) and novel transcript PB.82051 (yellow). Arrows indicate the locations of the forward and reverse primers used for amplification. Dashed arrow indicates a primer that spans the exon-exon junction. **f** RT-PCR gel validation of the novel isoform, showing a band at 133 bp confirming the presence of the isoform in all 16 samples. Nuclease-free water (NFW) serves as the negative control. **g** *APOE* isoform expression in individual samples from Kinnex dataset. AD samples with CPM ≥ 1 are represented as red circles; ND samples with CPM ≥ 1 are represented as blue circles; and AD and ND samples with CPM < 1 are represented as gray diamonds. **h** Normalized gene expression of *APOE* gene across different cell types from short-read sequencing data. Asterisks (*) indicate DEGs (Wilcoxon rank-sum test with Bonferroni correction). UpSet plots showing FSM (**i**), NIC (**j**), and NNC (**k**) isoforms from our confident set detected in other datasets.

DET (PB.178972) was from the gene *CHI3L1* (Fig. 5a). *CHI3L1* encodes for chitinase-3-like protein 1 (also known as YKL-40) and is a known biomarker for inflammation^{61,62}. At the transcript level, the NIC isoform (PB.178972), which has two retained introns, was significantly upregulated in AD brains (Fig. 5a), while the FSM and NNC transcripts were not

differentially expressed. Consistent with our DET analysis, the FSM transcript was not detected as being expressed at different levels in AD, while the NIC isoform was significantly upregulated in AD brains (Fig. 5a)

For downregulated isoforms, multiple DETs were identified in genes that have been previously linked to brain function. The FSM isoform of

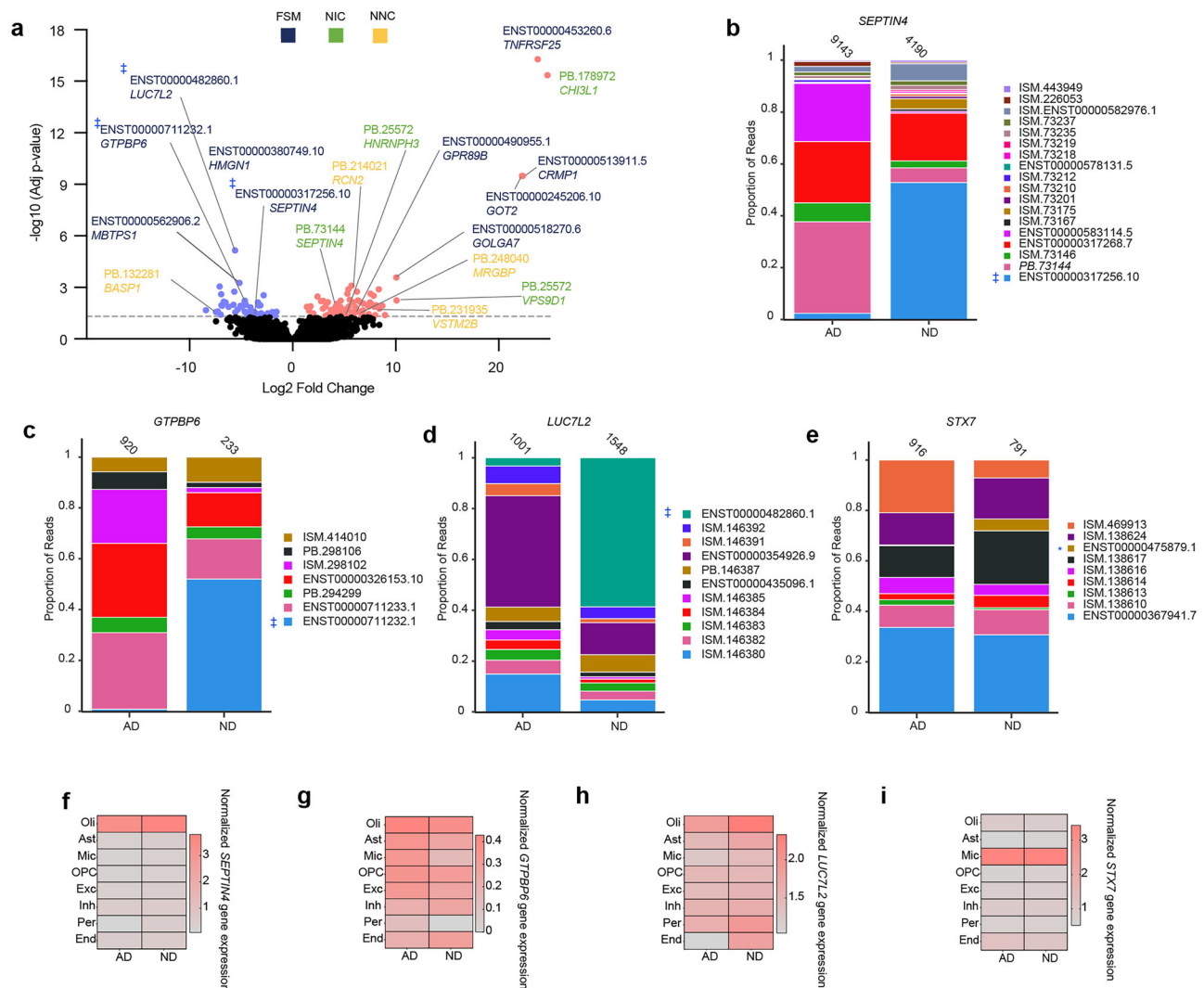


Fig. 5 | Disease-associated novel isoforms expression and transcript switching in AD. **a** Volcano plot of differentially expressed transcripts (DETs). Red dots represent up-regulated DETs in AD (L2FC > 1 and adjusted p -value < 0.05); blue dots represent down-regulated DETs in AD (L2FC < -1 and adjusted p -value < 0.05); and black dots indicate non-significant transcripts (adjusted p -value > 0.05). Transcripts marked with (*) sign represent the same isoform that is both differentially expressed and switched. Isoform names are colored by SQANTI3 category: FSM (black), NIC (green), NNC (yellow). **b** Proportion of the reads for each transcript of *SEPTIN4* gene in AD and ND samples. ENST00000317256.10 (light blue) is differentially expressed and significantly switched between ND and AD brains. Numbers above each bar indicate total read count. **c** Proportion of reads for each

transcript of *GTPBP6* gene in AD and ND samples. ENST00000711232.1 transcript (light blue) is differentially expressed and significantly switched between ND and AD brains. Numbers above each bar indicate total read count. **d** Proportion of the reads for each transcript of *LUC7L2* gene in AD and ND samples. ENST00000482860.1 (teal) differentially expressed and significantly switched between ND and AD brains. Numbers above each bar indicate total read count. **e** Proportion of the reads for each transcript of *STX7* gene in AD and ND samples. ENST00000475879.1 (tan) is switched between ND and AD brains. Numbers above each bar indicate total read count. Normalized gene expression across cell types from short-read sequencing data for *SEPTIN4* (**f**), *GTPBP6* (**g**), *LUC7L2* (**h**), and *STX7* (**i**).

HMGNI, ENST00000380749.10, was significantly downregulated in AD (Fig. 5a). *HMGNI* is a chromatin-binding protein involved in gene expression regulation and neural development, and altered *HMGNI* levels and function have been shown to contribute to cognitive deficits in Down syndrome^{63,64}. A novel isoform of *BASPI* was also found to be down-regulated in AD. *BASPI* is a signaling protein involved in neurite outgrowth and synaptic plasticity⁶⁵ that has been explored as a potential CSF biomarker for Alzheimer's disease⁶⁶. The *BASPI* NNC isoform (PB.132281) contained a novel exon on the 5' end. Its predicted protein sequence matches previously annotated *BASPI* isoforms, ENST00000322611.4 and ENST00000616743.1, suggesting that the novel exon serves as an alternate 5'UTR that may affect its stability and translation efficiency.

SEPTIN4 had multiple DETs in AD: a down-regulated and significantly switched FSM isoform (ENST00000317256.10) and an up-regulated NIC isoform (PB.73144) with part of exon 2 missing (Fig. 6a),

both of which were validated by RT-PCR (Fig. 6b,c, Supplementary Fig. 5-6). *SEPTIN4* encodes a GTP-binding cytoskeletal protein involved in neuronal structure and synaptic function⁶⁷. It has also been found in pathological protein aggregates of both Parkinson's disease and AD^{68,69}. Another gene example expressed by reactive astrocytes⁷⁰ was *CHI3L1* that encodes YKL-40 and is found at higher levels in the cerebrospinal fluid, positively correlating with disease severity and cognitive decline in AD⁶¹. We identified three *CHI3L1* transcripts (Fig. 6d) with varying expression across AD and ND brains (Fig. 6e). Based on gene-level analysis from short-read sequencing, *CHI3L1* was significantly upregulated in AD astrocytes and had limited expression in other cell types (Fig. 6f). Both FSM and NIC transcripts were validated by RT-PCR (Fig. 6g, Supplementary Fig. 7-8, Supplementary Data 9). Consistent with our DET analysis, the FSM transcript was not detected as being expressed at different levels in AD compared to ND, while the NIC isoform was significantly upregulated in AD brains (Fig. 6h).

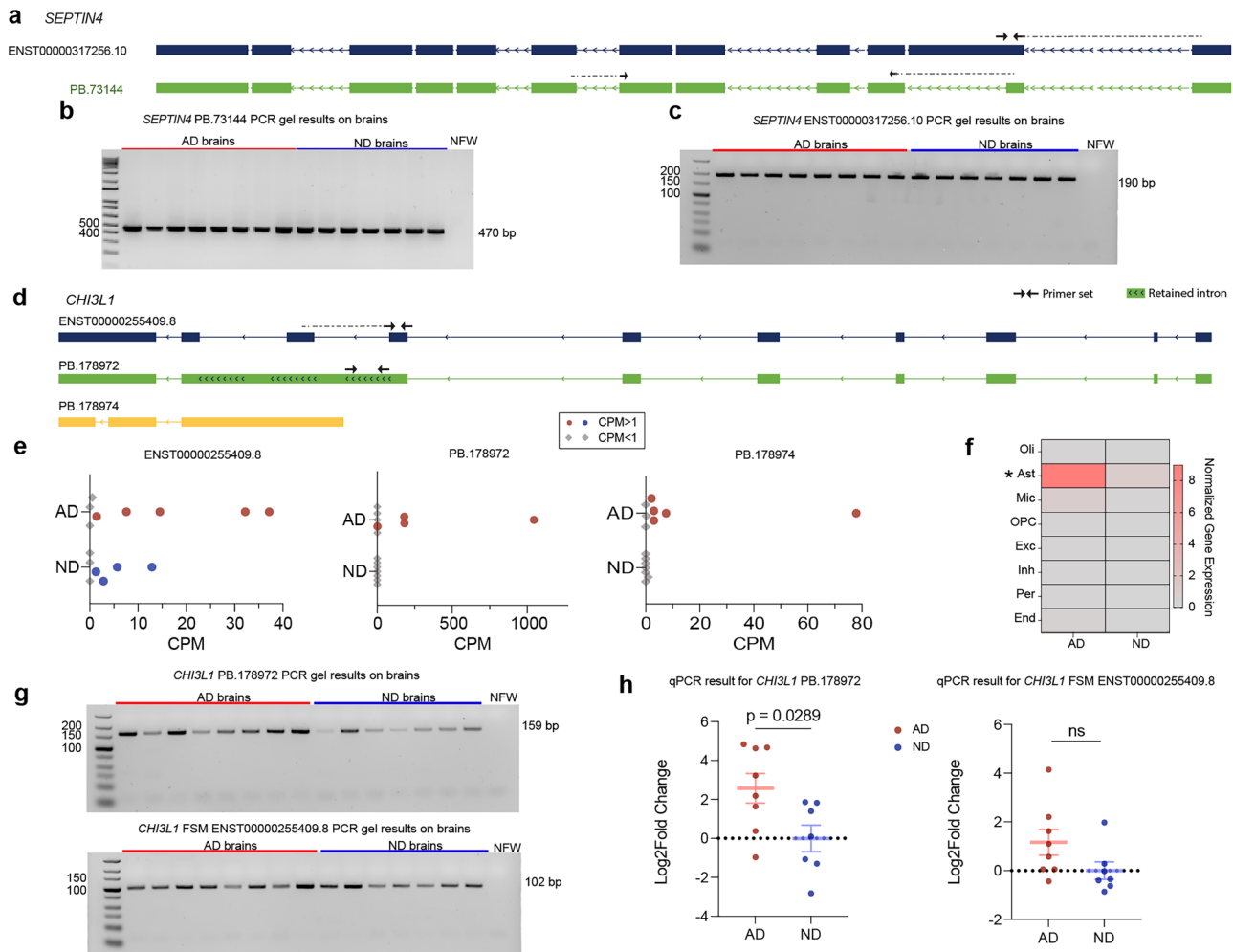


Fig. 6 | DET isoforms expression. **a** Confident FSM isoform (navy blue) and NIC isoform (green) of *SEPTIN4*. Arrows indicate forward and reverse primers. Dashed arrows indicate primers spanning an exon-exon junction. Gel electrophoresis validation of RT-PCR for *SEPTIN4* PB.73144 NIC isoform (**b**) and *SEPTIN4* ENST00000317256.10 isoform (**c**). Nuclease-free water (NFW) was used as a negative control. **d** Confident FSM isoform (navy blue), NIC isoform (green), and NNC isoform (yellow) of *CHI3L1*. Arrows indicate forward and reverse primers. Dashed arrows indicate primers spanning an exon-exon junction. **e** *CHI3L1* isoform expression in individual samples from Kinex dataset. AD samples with CPM ≥ 1 are represented as red circles; ND samples with CPM ≥ 1 are represented as blue

circles; and AD and ND samples with CPM < 1 are represented as gray diamonds. **f** Heatmap showing normalized gene expression of *CHI3L1* gene across different cell types from short-read sequencing data. Asterisk (*) indicates DEGs (Wilcoxon rank-sum test with Bonferroni correction). **g** Gel electrophoresis validation of RT-PCR for *CHI3L1* PB.178972 NIC isoform and *CHI3L1* ENST00000255409.8 isoform. Nuclease-free water (NFW) was used as a negative control. **h** RT-qPCR validation of the FSM and NIC transcript. NIC was significantly upregulated in AD brains (p -value: 0.0289; Mann-Whitney U test, (two-tailed)), whereas FSM isoform showed no difference (not significant: ns = 0.0721). Error bars indicate SEM.

Upregulation of *CHI3L1* observed at the gene level is likely driven by the increased expression of isoform PB.178972, given that the other isoforms were minimally expressed in comparison and did not change in expression in AD.

Distinct transcript expression patterns across cell types

Cell type assignments for isoforms were made by using short-read analysis by employing libraries sharing the same cellular barcodes (Supplementary Data 8). An isoform was considered expressed in a given cell type if at least 5% of all cells in which that isoform was detected belonged to that cell type. In all cell types, FSM isoforms comprised a higher proportion of reads, while NIC and NNC had the lowest read proportions (excluding fusion, antisense, and genic sequences) (Fig. 7a-f). In both AD and ND, most isoforms were present in neuronal and non-neuronal cell types (Fig. 7g). Also, while most isoforms were expressed in more than one cell type, each of the major cell types possessed a unique set of isoforms, suggesting specialized functions for transcript isoforms within that cell type population (Fig. 7h-l).

A focused examination of six AD-related genes—*APOE*, *APP*, *BIN1*, *CLU*, *MAPT* and *PSEN1*—and *SMARCA5*, which has been linked to neurological disorders, revealed different expression patterns across cell types and AD (Fig. 7m). For example, a novel *APOE* isoform (PB.82051; Fig. 4e) was not detected in oligodendrocytes of ND brains. In contrast, the *CLU* PB.149064 isoform (Fig. 7m, Fig. 8a-c) was not detected in AD oligodendrocytes. Two different *APP* isoforms (ENST00000348990.9/APP-695, ENST00000357903.7/APP-751, Fig. 8d) with varying expression across AD and ND brains but not among DEGs (Fig. 8e,f), were detected in inhibitory neurons of AD brains but were absent from ND samples. Moreover, the APP-695 isoform was predominantly expressed in AD excitatory neurons, whereas the APP-751 isoform was predominantly in ND excitatory neurons (Fig. 7m). In contrast, an inverse pattern was observed in oligodendrocytes (Fig. 7m). Two novel isoforms of *MAPT* (PB.4075 and PB.4078, Fig. 8g-i) were both detected in excitatory neurons whereas only one, PB.4078, was found in inhibitory neurons. Furthermore, other genes beyond the AD set had expression patterns that varied by cell type. Genes with DETs identified

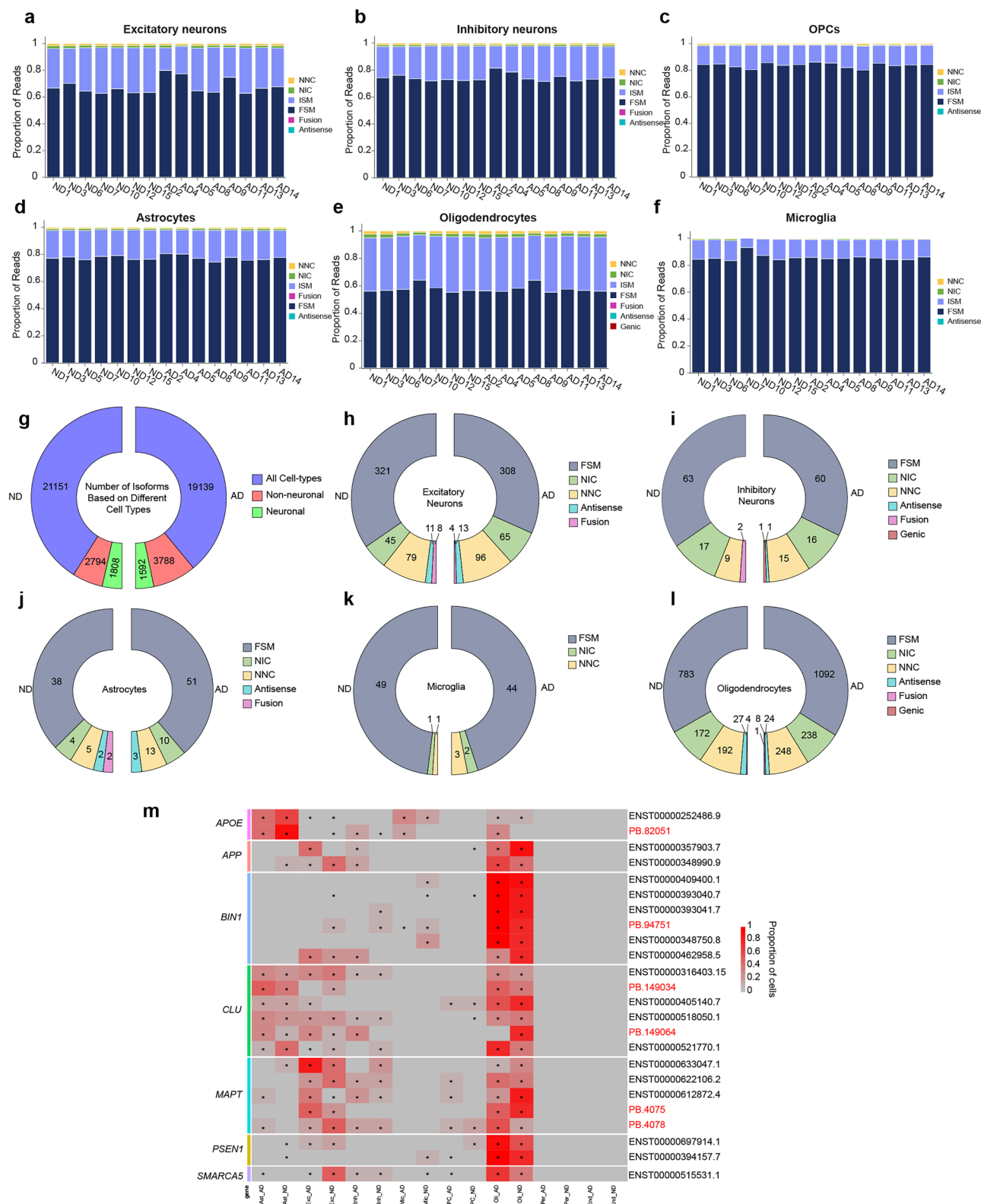


Fig. 7 | Isoform expression pattern across cell types in AD and ND brains. a–f Proportions of reads for different isoform categories (FSM, ISM, NIC, NNC, genic, antisense, and fusion) across different cell types in AD and ND. **g** Number of isoforms detected in each cell type, separated by neuronal (excitatory and inhibitory neurons) and non-neuronal (astrocytes, microglia, oligodendrocytes, and OPC) and

shared across the neuronal and non-neuronal. **h–l** Number of unique isoforms expressed in only one cell type, suggesting transcript specificity across cell types. **m** Heatmap showing cell-type proportion in which each isoform was detected from neurological genes across neurons and glia.

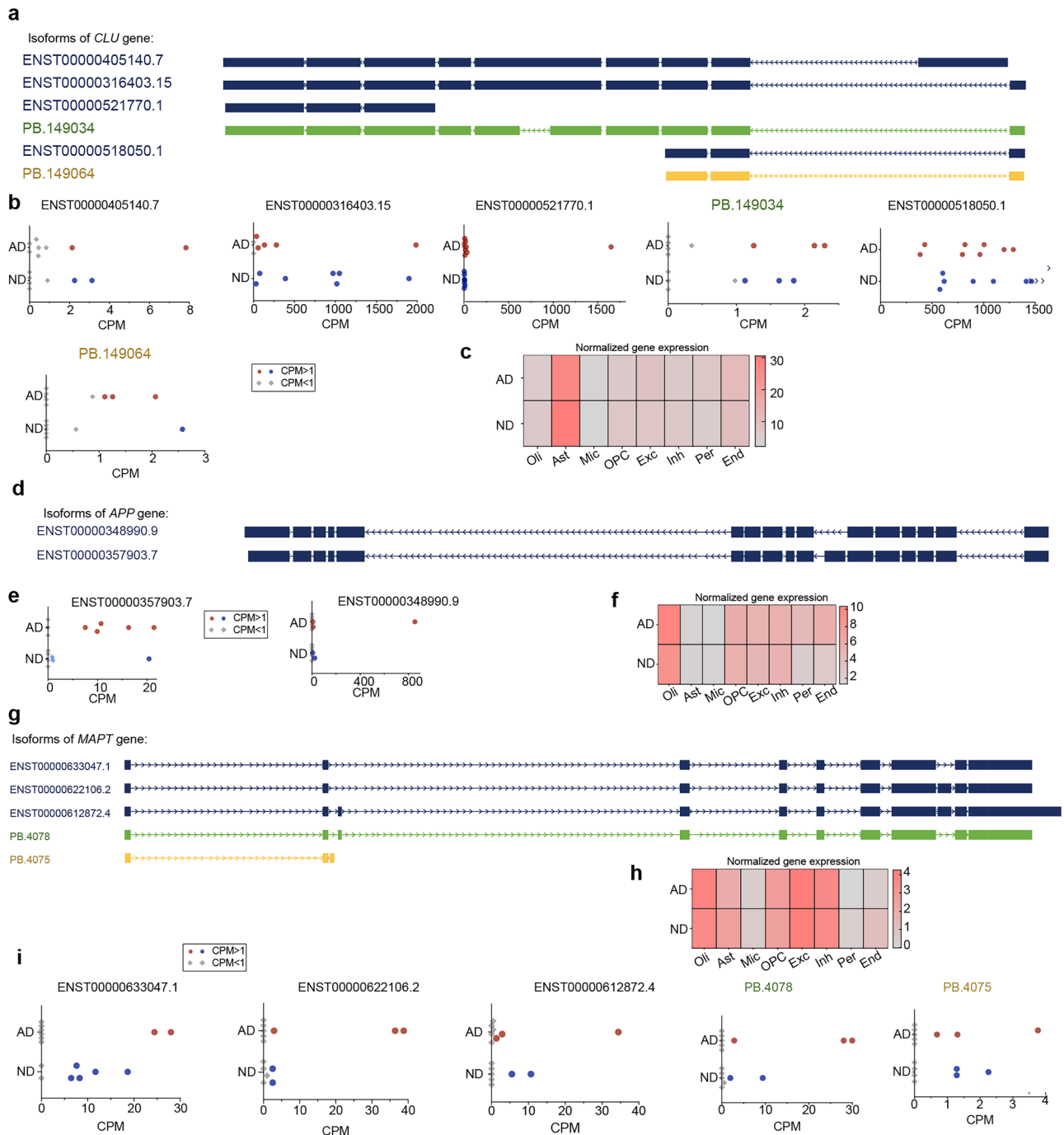


Fig. 8 | Transcripts of AD-related genes. a Confident FSM *CLU* transcripts (navy blue), PB.149034 (green) NIC isoform, and PB.149064 (yellow) NNC isoform.

b *CLU* isoform expression in individual samples from Kinnex dataset. AD samples with CPM ≥ 1 are represented as red circles; ND samples with CPM ≥ 1 are represented as blue circles; and AD and ND samples with CPM < 1 are represented as gray diamonds. **c** Heatmap showing normalized gene expression of *CLU* gene across different cell types from short-read sequencing data. Expression was not significantly changed between AD and ND in any cell type (Wilcoxon rank-sum test with Bonferroni correction). **d** Confident FSM *APP* transcripts (navy blue). **e** *APP* isoform expression in individual samples from Kinnex dataset. AD samples with CPM ≥ 1 are represented as red circles; ND samples with CPM ≥ 1 are represented as blue circles; and AD and ND samples with CPM < 1 are represented as gray diamonds. **f** Heatmap showing normalized gene expression of

APP gene across different cell types from short-read sequencing data. Expression was not significantly changed between AD and ND in any cell type (Wilcoxon rank-sum test with Bonferroni correction). **g** Confident FSM *MAPT* transcripts (navy blue), PB.4087 NIC isoform (green), and PB.4075 NNC isoform (yellow).

h Heatmap showing normalized gene expression of *MAPT* gene across different cell types from short-read sequencing data. Expression was not significantly changed between AD and ND in any cell type (Wilcoxon rank-sum test with Bonferroni correction). **i** *MAPT* isoform expression in individual samples from Kinnex dataset. AD samples with CPM ≥ 1 are represented as red circles; ND samples with CPM ≥ 1 are represented as blue circles; and AD and ND samples with CPM < 1 are represented as gray diamonds.

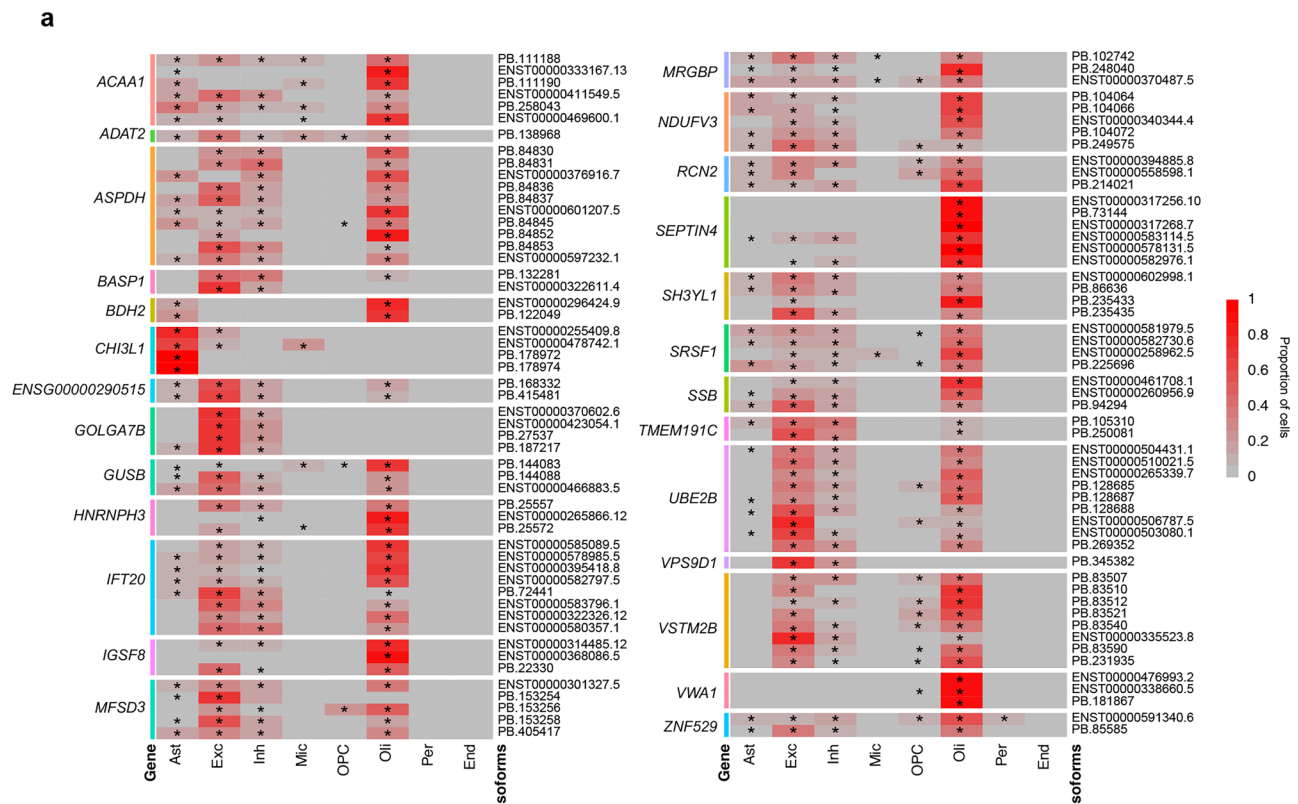


Fig. 9 | DET Isoform expression pattern across cell-types in AD and ND brains. a Heatmap showing proportion of cells in which each isoform was detected from DET genes across neurons and glia.

from the pseudobulk analysis displayed differing patterns with some isoforms detected in single cell types while others showed widespread expression (Fig. 9a). For example, *CHI3L1* PB.178972 and PB.178974 (Fig. 6d) were detected only in astrocytes (Fig. 9a), and *SEPTIN4* PB.73144 only in oligodendrocytes (Fig. 9a). Overall, these isoform data complement and augment prior short-read transcriptomic studies and demonstrate isoform diversity varying amongst cell types and AD state.

Discussion

RNA sequence diversity generated by mechanisms that include splicing has been implicated in AD; however, its range of isoforms and relationships to specific cell types and disease remain unclear and/or unassessed by prior short-read single-cell sequencing approaches. Long-read sequencing studies have implicated RNA diversity in single brain cells, albeit hampered by limited sequencing depth²⁹. PacBio Kinnex addresses this limitation by increasing long-read sequencing depth by ~16X through template concatenation. Its application here to single-nucleus RNA-sequencing from human brain revealed a vast level of diverse RNA isoforms, ranging from known splice variants albeit not previously linked to specific cell types, to unknown transcripts. Differential transcript expression and switching analysis of pseudobulk isoforms revealed isoform changes that would be overlooked by only short-read methods. Limitations, not unique to this study, included sample size for disease vs. non-disease comparisons, 10X library preparations that limit cDNA length, along with overall cost. Additionally, this study focused on the pre-frontal cortex which may not be representative of other brain regions affected by AD, but is more readily obtained compared to samples from the entorhinal cortex or hippocampus. Obtaining long, intact RNA is challenging to achieve from frozen and FFPE postmortem tissues. Ongoing optimization in extraction and library preparation chemistry will improve our ability to use FFPE samples alongside frozen tissues⁷¹. Other considerations of this study that warrant further detail are discussed below.

The use of nuclei rather than whole cells is standard for single-cell techniques to study human brain tissue^{9,10}, an approach that has been shown to be representative of gene expression in the cytosol^{72,73}. However, as with bulk long-read sequencing of total RNA, concerns arise regarding the inclusion of premature mRNA isoforms. Other groups have tried to limit the capture of premature mRNA through the use of exome enrichments with probes targeting exonic sequences⁷⁴. To limit the inclusion of incompletely spliced transcripts in our analysis, we applied confidence filtering, assuming that partially spliced or fully unspliced transcripts would be detected only transiently and not consistently across multiple samples. Importantly, transcripts with retained introns are included in GENCODE/Ensembl annotations, and intron retention is a known regulated form of alternative splicing rather than solely a marker of premature mRNA.

Filtering transcriptomic data is essential for distinguishing biologically meaningful isoforms from technical artifacts³⁶. We chose a confidence cutoff to generate a set of high-quality isoforms that we could compare between our AD and ND groups in addition to other studies, assuming that confident transcripts would be identified more consistently across multiple samples than artifacts of library preparation and sequencing. Somatic mosaicism, however, may introduce isoform diversity consistent with the brain's vast genomic, inter-individual heterogeneity that is excluded using such filtering methods^{31,75,76}. RT-SGR of genes within single neurons produces additional copies of a gene that, if transcribed, may only be expressed in a few cells or even a single cell within an individual brain^{25,77}. As an example, several isoforms of *SMARCA5* did not meet confidence cutoffs but may be indicative of their potential to undergo RT-SGR (Supplementary Data 3). Future analyses will be needed to identify and interpret the many isoforms removed in the filtering process.

Disease-related transcript differences were observed for many genes, with specific isoforms being expressed in defined cell types. Two notable examples that showed AD-dependent and significant changes in specific RNA isoforms were *SEPTIN4* and *CHI3L1*, both of which highlight the

importance of examining isoform-level expression. *SEPTIN4* was not a DEG (differentially expressed gene) in any cell type, likely explained by opposing changes in two isoforms—one upregulated, the other downregulated—that masked detection of differential expression at the gene level, underscoring a need for isoform resolution. *CHI3L1* was differentially expressed at the gene level in astrocytes, however our analysis revealed that most of its expression could be attributed to a novel isoform (PB.178972) rather than the known transcript. This novel transcript itself was significantly upregulated in AD, with no change detected in the expression of the known isoform, as validated with RT-qPCR. Given these isoforms' disease-associated expression patterns, they may represent novel mechanisms in AD and/or useful biomarker candidates, but require further study.

This study highlights the advantages of long-read RNA sequencing for resolving transcript-level changes that are often obscured in gene-level analyses of single brain cells. Different isoforms of the same gene can have distinct, and sometimes opposing, expression patterns, underscoring a need to capture isoforms to better understand the molecular complexities of the brain and diseases like AD. Future studies integrating long-read sequencing with functional assays will be essential to further characterize isoforms' relevance to disease mechanisms and their potential uses as biomarkers or targets for therapeutic intervention.

METHODS

Tissue selection and preparation

Fifteen prefrontal cortices were selected from various brain bank sources (eight female and seven male donors. The sources are: Emory University ($n = 9$), UK Imperial ($n = 3$), Southwest Neurodegenerative Diseases Brain Bank ($n = 2$), Pittsburgh Neurodegenerative Diseases Brain Bank ($n = 1$), and Sepulveda Neurodegenerative Diseases Brain Bank ($n = 1$). Samples were transferred from -80°C freezers to a cryostat set at -18°C . The samples were sectioned serially to obtain 200–300 μm sections for nuclei isolation and 20 μm sections for RNA integrity number (RIN) value analysis, as well as Nissl, Thioflavin S staining, RT-PCR, and RT-qPCR. Samples were matched for age, sex and RIN (Mann-Whitney U test). All ethical regulations relevant to human brain sample use were followed.

Nissl staining

Brain sections (20 μm) were mounted on glass slides [Fisherbrand™ Tissue Path Superfrost™ Plus Gold Slides, 22-035813] and fixed in 10% neutral buffered formalin (NBF) [Epredia, 22-050-104] for 5 min at room temperature. Sections were incubated in 70% and 95% ethanol for 3 min each, then 100% ethanol [Decon™ Labs, 04-355-222] for 20 min, followed by rehydration steps in 95% and 70% ethanol, and rinsed in deionized (DI) water. Sections were incubated in 0.2% Cresyl violet solution for 5 min. Excess stain was removed with DI water ($3 \times 30\text{ s}$), 70% ethanol and 95% ethanol for 1 min each, and xylene [Fisher Chemical, X5SK-4] for 3 min. Coverslips were used with Cytoseal 60 mounting medium [Epredia, 23-244257] to seal the slides.

Thioflavin S staining

Brain sections (20 μm) were mounted on glass slides and fixed in 10% NBF for 5 minutes at room temperature. Sections were washed with DI water twice and then incubated in 1% Thioflavin S solution for 8 min. Sections were washed again with DI water and counterstained with 5 $\mu\text{g}/\text{mL}$ DAPI [Sigma-Aldrich, D9542-5MG] for 3 min at room temperature, followed by a DI water rinse. Sections were dehydrated in 70% ethanol for 4 min, 95% ethanol for 3 minutes, 100% ethanol for 1 min, and then incubated in xylene for 3 min. Coverslips were used with Cytoseal 60 mounting medium to seal the slides.

RIN measurement

RNA was isolated from 20 μm sectioned brain samples using the RNeasy Mini Kit [Qiagen, 74106]. Samples were then diluted to 3 $\text{ng}/\mu\text{L}$, and their RIN values were measured on an Agilent 4200 TapeStation using Agilent High Sensitivity RNA ScreenTape at the Sanford Burnham Prebys

Genomics Core. Samples with $\text{RIN} \geq 6$ were included for further analysis (Supplementary Data 1).

Nuclei isolation

Human postmortem brain tissues were randomly sorted into four groups for processing to negate potential batch effects. To each sample, 1 mL of chilled nuclei buffer (250 mM sucrose [Fisher Chemical, S3-212], 1 mM DTT [Invitrogen, D1532], 10 mM Tris-HCl pH 7.5 [Invitrogen, 15567], 25 mM KCl [Sigma-Aldrich, P9333], 5 mM $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ [Sigma-Aldrich, M9272], 0.1% Triton X-100 [Sigma-Aldrich, T8787], 1U/ μL Protector RNase Inhibitor [Millipore Sigma, 3335402001], and cOmplete EDTA-free protease inhibitors [Roche, 11836170001]) were added and incubated for 5 min. Samples were gently homogenized using a tissue grinder [Wheaton, 358005] with 20 strokes. A 50 μm cell filter [Sysmex, 04-0042-2317] was used to filter the homogenate and remove debris. The samples were then centrifuged at 820 RCF for 5 min at 4°C . The supernatant was aspirated, and the pellet was resuspended in 500 μL of cold sorting buffer (PBS with 10% BSA [Gemini Bioproducts, 700-107P], 5 $\mu\text{g}/\text{mL}$ 7-AAD [Sigma, A9400], cOmplete EDTA-free protease inhibitors, and Protector RNase Inhibitor) and mixed gently five times. FACS was performed using a FACSaria Fusion [BD Biosciences] by gating out debris and selecting 7-AAD+ singlets to isolate an average of 200,000 nuclei per sample. Following sorting, nuclei were resuspended in PBS with 10% BSA and 5 U/ μL Protector RNase Inhibitor, then diluted to a final concentration of 200–1200 nuclei/mL.

10x Genomics 3' cDNA library generation

To generate 3' cDNA libraries, single-nuclei suspensions were processed using the Chromium Next GEM Single Cell 3' Reagent Kit v3.1 (Dual Index), according to the manufacturer's protocol (10x Genomics). From each sample, 7000–8000 nuclei were encapsulated into Gel Beads-in-Emulsion (GEMs) using the Chromium Controller, and the RT reaction took place. Barcoded full-length cDNA was purified using Dynabeads™ MyOne™ SILANE. The purified cDNA was then amplified by PCR. The pre-fragmented cDNA library was split between Illumina NovaSeq 6000 short-read (25%) and PacBio Sequel II long-read sequencing (75%).

3' cDNA library fragmentation for Illumina short-read sequencing

Amplified cDNA (25%) was fragmented using the fragmentation enzyme and the fragmentation buffer provided in the Chromium Library Construction Kit. End-repair and A-tailing were performed after fragmentation. Sequencing adapters and Illumina indices were ligated to the DNA fragments. The final libraries were purified, and the quality and size distribution were assessed using an Agilent Fragment Analyzer. The generated cDNA libraries were sent to Azenta/Genewiz for sequencing on the NovaSeq 6000 for an average of 62,612 reads per cell.

Kinnex library preparation and sequencing

Pre-fragmented cDNA libraries (75%) were used for MAS-Seq/Kinnex sequencing using “MAS-Seq for 10x Single Cell 3' kit” (PacBio) according to the manufacturer's instructions. Before starting the MAS/Kinnex protocol, the cDNA libraries underwent a quality control step to evaluate size distribution and quantity using a High Sensitivity DNA kit by Agilent Femto Pulse (FP-1101-0275). A PCR step was then performed to add biotin tags to DNA fragments, enabling the removal of template switching oligos (TSOs) that were generated during cDNA synthesis in the 10x protocol. In the next step, cDNA libraries were amplified in 16 separate PCR reactions using various primers (A–P MAS primer mix). The PCR-amplified fragments were then concatenated into a longer concatemer called the MAS array. DNA damage repair, nuclease treatment, and a final clean-up with SMRTbell beads were performed at the end of the library preparation. All final SMRTbell libraries were quality controlled using the Agilent Femto Pulse instrument.

Each sample was loaded onto its own 8 M SMRTCell and sequenced on the PacBio Sequel II. On average 2,077,210 HiFi reads were obtained per sample.

Short-read snRNA-seq analysis

Processing/QC. Using 10x Genomics Cell Ranger (v7.1.0), samples were demultiplexed, and reads were aligned to the reference genome to quantify unique molecular identifiers (UMIs) and create a cell-count matrix. Because sequencing libraries were prepared from isolated nuclei and not whole cells, a pre-mRNA reference file (ENSEMBL GRCh38) was used with the default parameters to capture intronic reads originating from nuclear pre-mRNAs. Seurat (v4.3) was used to process these cell-count matrices. Nuclei were removed if they expressed fewer than 300 genes, contained greater than 1% of reads mapping to mitochondrial genes, or exceeded an outlier cutoff of UMIs from interquartile range calculations (to remove potential doublets). Each sample was normalized using Seurat's SCTransform() function with vst.flavor = "v2" to prepare for integration.

Cell type identification and differentially expressed gene analysis.

SCTransform()-normalized datasets were integrated into a single Seurat object following the recommended workflow using PrepSCTIntegration(), FindIntegrationAnchors(), and IntegrateData(). To identify cell types, labels were transferred from a reference dataset generated from human cortical samples¹⁰ using TransferData() (dims=1:30). UMAP embeddings were generated (1:50), and clusters were generated using FindNeighbors() (dims=1:50) and FindClusters() (res=0.3). Differentially expressed genes between cell types in Alzheimer's disease and in non-diseased sample groups were identified using FindMarkers() with default parameters. Genes were labeled as DEGs if $|\log_2\text{fold}| > 0.25$ and the adjusted p -value < 0.05 using the Wilcoxon Rank Sum test with Bonferroni correction.

Gene ontology (GO) analysis

GO analysis was performed using WebGestalt. In the basic parameters, overrepresentation analysis as the method of interest, *Homo sapiens* as the organism of interest, and GO with the subcategory 'Biological Process (no redundant)' was selected, and the analysis type was set to gene/protein. Cell-type-specific DEGs comparing diseased to non-diseased samples were uploaded. A false discovery rate (FDR) of < 0.05 using the Benjamini-Hochberg (BH) test was used to assess significant enrichment of DEGs.

Long-read sequencing analysis

Quality control and isoform annotation. Each sample was processed individually from subreads to annotated isoforms. Starting from subreads off of the PacBio Sequel II, ccs (v6.4.0) was used to generate high-quality, consensus (HiFi) reads consisting of concatenated cDNAs. HiFi reads were then split into S-reads (individual, barcoded cDNAs) using skera (v0.1.0). S-reads were processed using lima (v2.7.1) for proper orientation and removal of primers. This step was followed by Isoseq (v3.8.2) tag to remove the cellular barcode and unique molecular identifier (UMI) sequences and add them to the metadata for each read. Isoseq refine was used to remove polyA tails and concatemers and generated full-length, non-concatemeric (FLNC) reads, followed by isoseq correct to correct errors in cell barcode identification. FLNC reads were deduplicated using isoseq dedup, and consensus reads generated from this deduplication step were mapped to the human reference genome (GRCh38) with pbmm2 (v1.12). Mapped reads were collapsed into non-redundant isoforms using cDNA_Cupcake (v29.0) (collapse_isoforms_by_sam.py). Unique, non-redundant transcripts were then annotated using a modified version of SQANTI3³⁹ and the GENCODE v44 reference annotation. SQANTI3 filtering was used to remove potential artifacts. cDNA_Cupcake's scripts make_csv_for_dedup.py and collate_FLNC_gene_info.py were used to generate a table with reads and their associated barcodes, UMIs, gene, isoform category from SQANTI3, celltype, etc.

isoSeQL. Each sample was processed sequentially with isoSeQL (v1.0.1)⁵⁰ to create an isoSeQL database that unifies all sample, isoform,

and read count information, enabling comparisons and differential expression/usage analyses. Required inputs included the filtered classification file output from SQANTI3; the genePred file from SQANTI3; the read, barcode, and cell type table from cDNA_Cupcake; and two config files, including information about how the data were processed and sample metadata. Standard and custom queries were used to extract information from the isoSeQL database. isoSeQL is available at: <https://github.com/christine-liu/isoSeQL>

Pseudobulk DET/DTU analysis. Isoform x read count tables were generated through querying isoSeQL. DESeq2 (version 3.21) was used to identify differentially expressed transcripts (Wald test with Benjamini-Hochberg adjustment). Only isoforms deemed "confident" through the established QC metrics were included in the analysis. DEXseq (version 1.32) was used to identify switched isoforms (generalized linear modeling with Benjamini-Hochberg adjustment). Input tables were set up such that isoforms represented features (typically exons) and grouped by gene.

Comparison to publicly available datasets. Fastq files from Heberle et al.'s study were downloaded from Synapse with accession number: syn52047893⁷⁸. They were processed via pychopper to remove primer sequences and correct the orientation of the reads. Transcripts were then mapped to the hg38 reference genome using minimap2 (nanopore parameters) and collapsed into unique, non-redundant isoforms using isoseq collapse with parameters --min-aln-coverage 0.85 --min-aln-identity 0.8 to account for the lower sequencing accuracy. Subread files (bam) from Leung et al.'s study were downloaded from SRA (PRJNA664117)⁷⁹. These subreads were processed nearly identically to our dataset. CCS generated HiFi reads; lima oriented reads and removed polyA tails; reads were mapped to the hg38 reference genome with pbmm2; transcripts were collapsed using isoseq collapse; and isoforms were annotated and filtered using our modified version of SQANTI3. Raw and processed files were made available from the PacBio Alzheimer's Iso-Seq dataset (<https://www.pacb.com/general/data-release-alzheimer-brain-isoform-sequencing-iso-seq-dataset/>). We downloaded the subreads.bam and processed it identically to our dataset: generated HiFi reads using CCS, oriented read and removed polyA tails with lima, mapped reads to the reference genome with pbmm2, collapsed transcripts with isoseq collapse; and annotated and filtered isoforms using a modified version of SQANTI3. Isoforms from all three of these datasets were added to the same isoSeQL database with the rest of our samples. Isoforms from our confident set that were observed in additional samples (regardless of level or number of samples) from these other datasets were flagged.

Cell type assignment based on short-read analysis. Barcodes from the long-read data were matched to their corresponding cell type determined from short-read sequencing analysis. For each isoform, the % of cells of each cell type that the isoform was detected in was calculated. The isoform was determined to be expressed in a particular cell type if that cell type made up 5% or more of the total cells it was detected in.

cDNA synthesis for RT-PCR and RT-qPCR

RNA was extracted from brain tissues using an RNeasy Plus Kit [Qiagen, 74134] according to the manufacturer's instructions to remove genomic DNA. RNA concentrations were measured and diluted using a NanoDrop. cDNA was synthesized using SuperScript IV VILO (SSIV VILO) Master Mix with ezDNase [Invitrogen, 11766050], following the manufacturer's instructions. Samples underwent two types of DNase treatment to ensure complete removal of genomic DNA. Final cDNA libraries were diluted to 1 ng/μl and were stored at -80°C .

RT-PCR confirmation of novel isoforms

PCR was performed using Platinum SuperFi II Master Mix [Invitrogen, 2368010], 10 μM primer mix, and NFW. The PCR run included the

following cycle settings in the thermocycler: 98 °C for 30 s, then 35 cycles of: 98 °C for 10 s, 60 °C for 10 s, and 72 °C for 30 s, with a final extension at 72 °C for 5 min. PCR products and Ultra Low Range DNA ladder (10–300 bp) [Invitrogen, 10597012] or the 1 Kb plus DNA ladder (100–15000 bp) [Invitrogen, 10787018] were loaded onto a 4% or 1.5% agarose gel prepared in 1× TAE buffer and stained with SYBR Safe DNA Gel Stain [ApexBio, A874310]. Then the gel was placed in an electrophoresis chamber and run at 100 V for 30 min.

Primer sequences ordered from IDT:

APOE_NNC_FWD: AGCGGAGGTGAAGGAGCA
 APOE_NNC_REV: CTGCATGTCTTCCACCAGG
 BIN1_NNC_FWD: GTG AAG TTC TCG GGG AAG
 BIN1_NNC_REV: GGAGGTGTCTCAAGGATGAAG
 CH13L1_NIC_FWD: CAGATGGCAGGTCTTGCG
 CH13L1_NIC_REV: TCCCAACACCTGGATTTC
 CH13L1_FSM_FWD: CACAGCATAGTCAGTGTTGC
 CH13L1_FSM_REV: TTTCATGGAGCCTGGCGTG
 SEPTIN4_FSM_FWD: GTCTGAACTCCAGGTCATCA
 SEPTIN4_FSM_REV: GCTGCAGCCATGATCAAGC
 SEPTIN4_NIC_FWD: ATCCAATGGCCGAGCCCAT
 SEPTIN4_NIC_REV: GGAGAACTGAGCAAGATGAC

RT-qPCR confirmation of DETs

The RT-qPCR assays were performed using the CFX384 Touch Real-Time PCR Detection System [BioRad]. Reactions were performed in 10 µl triplicate solutions with 5 µl of SYBR™ Green Universal Master Mix [Applied Biosystems, 4309155], 1 µl cDNA, 0.4 µl specific forward and reverse primer mix and 3.6 µl nuclease-free water. RT-qPCR conditions were as follows: 50 °C for 2 min, and 95 °C for 2 min following by 40 cycles of: 95 °C for 15 s, 58 °C for 20 s, and 72 °C for 30 s, with a final extension at 72 °C for 5 min. To confirm the specificity of amplification for each isoform, a melt curve analysis was performed at the end of the run. *GAPDH* was used as a housekeeping gene for normalization of each transcript's expression levels. The relative expression and the log₂Fold Change value were calculated using the $2^{-\Delta\Delta Ct}$ equation and log₂(relative expression), respectively.

Statistics and reproducibility

Figure 1c. Proportions of each cell type in AD and ND samples: Cell-type proportions were compared between Alzheimer's disease (AD, $n = 8$) and non-diseased (ND, $n = 7$) brain samples. A two-tailed t-test as a standard frequentist comparison was used ($p > 0.05$ considered not statistically significant). To complement this analysis and account for compositional constraints, we also used propeller, a tool from the speckle package, which identified no significant changes in cell-type proportions ($p > 0.05$).

Figure 1d–h. DEGs were defined as having $|\log_2 \text{fold change}| > 0.25$ and the adjusted p -value < 0.05 using the Wilcoxon Rank Sum test with Bonferroni correction.

Figure 5a. DESeq2 analysis utilized the Wald test and Benjamini-Hochberg adjustment to determine statistically significant changes in transcript expression.

Figure 5 b–e. DEXseq analysis used linear modeling and Benjamini-Hochberg adjustment to identify isoform switching.

Figure 6h. RT-qPCR: Expression values were log₂-transformed prior to analysis. Differences between AD ($n = 8$) and ND ($n = 7$) brain samples were evaluated using a two-tailed Mann–Whitney U test. A p -value < 0.05 was considered statistically significant.

Figures Figures were created with the aid of BioRender, Adobe Illustrator, and Prism.

Reporting summary

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

Data availability

Fastq files from short-read sequencing and bam files from long-read sequencing have been submitted to the European Genome-Phenome Archive (EGA) under accession code EGAS50000001476. For privacy reasons, these data are access-controlled and can be obtained after signing a Data Usage Agreement. All uncropped and unedited gels are included in the Supplementary Information.

Received: 2 September 2025; Accepted: 16 February 2026;

Published online: 03 March 2026

References

- Zhang, J. et al. Recent advances in Alzheimer's disease: mechanisms, clinical trials and new drug development strategies. *Signal Transduct. Target Ther.* **9**, 211 (2024).
- Gustavsson, A. et al. Global estimates on the number of persons across the Alzheimer's disease continuum. *Alzheimers Dement.* **19**, 658–670 (2023).
- Griffiths, J. & Grant, S. G. N. Synapse pathology in Alzheimer's disease. *Semin. Cell Dev. Biol.* **139**, 13–23 (2023).
- Serrano-Pozo, A., Frosch, M. P., Masliah, E. & Hyman, B. T. Neuropathological alterations in Alzheimer's disease. *Cold Spring Harb. Perspect. Med.* **1**, a006189 (2011).
- Wilson, D. M. et al. Hallmarks of neurodegenerative diseases. *Cell* **186**, 693–714 (2023).
- Crist, A. M. et al. Transcriptomic analysis to identify genes associated with selective hippocampal vulnerability in Alzheimer's disease. *Nat. Commun.* **12**, 2311 (2021).
- Guennewig, B. et al. Defining early changes in Alzheimer's disease from RNA sequencing of brain regions differentially affected by pathology. *Sci. Rep.* **11**, 4865 (2021).
- Aguzzoli Heberle, B. et al. Systematic review and meta-analysis of bulk RNAseq studies in human Alzheimer's disease brain tissue. *Alzheimer's Dement.* **21**, e70025 (2025).
- Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
- Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
- Mathys, H. et al. Single-cell multiregion dissection of Alzheimer's disease. *Nature* **632**, 858–868 (2024).
- Mathys, H. et al. Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365–4385.e4327 (2023).
- Gazestani, V. et al. Early Alzheimer's disease pathology in human cortex involves transient cell states. *Cell* **186**, 4438–4453.e4423 (2023).
- Luquez, T. et al. Cell type-specific changes identified by single-cell transcriptomics in Alzheimer's disease. *Genome Med.* **14**, 136 (2022).
- Tao, Y., Zhang, Q., Wang, H., Yang, X. & Mu, H. Alternative splicing and related RNA binding proteins in human health and disease. *Signal Transduct. Target. Ther.* **9**, 26 (2024).
- Verta, J. P. & Jacobs, A. The role of alternative splicing in adaptation and evolution. *Trends Ecol. Evol.* **37**, 299–308 (2022).
- Wojtowicz, W. M., Flanagan, J. J., Millard, S. S., Zipursky, S. L. & Clemens, J. C. Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* **118**, 619–633 (2004).
- Graveley, B. R. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**, 65–73 (2005).
- Iijima, T. et al. SAM68 regulates neuronal activity-dependent alternative splicing of neuroligin-1. *Cell* **147**, 1601–1614 (2011).

20. Treutlein, B., Gokce, O., Quake, S. R. & Sudhof, T. C. Cartography of neuexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proc. Natl. Acad. Sci. USA* **111**, E1291–E1299 (2014).
21. Schreiner, D. et al. Targeted combinatorial alternative splicing generates brain region-specific repertoires of neuexins. *Neuron* **84**, 386–398 (2014).
22. Li, D., McIntosh, C. S., Mastaglia, F. L., Wilton, S. D. & Aung-Htut, M. T. Neurodegenerative diseases: a hotbed for splicing defects and the potential therapies. *Transl. Neurodegener.* **10**, 16 (2021).
23. Hervoso, J. L. et al. Splicing-specific transcriptome-wide association uncovers genetic mechanisms for schizophrenia. *Am. J. Hum. Genet.* **111**, 1573–1587 (2024).
24. Verdile, V. et al. Dysregulation of alternative splicing underlies synaptic defects in familial amyotrophic lateral sclerosis. *Prog. Neurobiol.* **231**, 102529 (2023).
25. Lee, M. H. et al. Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature* **563**, 639–645 (2018).
26. Lee, M. H. et al. Reply to: APP gene copy number changes reflect exogenous contamination. *Nature* **584**, E29–E33 (2020).
27. Nicodemus, J. et al. Sequence diversity and encoded enzymatic differences of monocistronic L1 ORF2 mRNA variants in the aged normal and Alzheimer's disease brain. *J. Neurosci.* (2025).
28. Mitsunaga, S. et al. Detection of APP gene recombinant in human blood plasma. *Sci. Rep.* **13**, 21703 (2023).
29. Palmer, C. R., Liu, C. S., Romanow, W. J., Lee, M. H. & Chun, J. Altered cell and RNA isoform diversity in aging Down syndrome brains. *Proc. Natl. Acad. Sci. USA* **118**, (2021).
30. Kaeser, G. & Chun, J. Brain cell somatic gene recombination and its phylogenetic foundations. *J. Biol. Chem.* **295**, 12786–12795 (2020).
31. Sun, M. et al. Systematic functional interrogation of human pseudogenes using CRISPRi. *Genome Biol.* **22**, 240 (2021).
32. Harrison, P. M., Zheng, D., Zhang, Z., Carriero, N. & Gerstein, M. Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res.* **33**, 2374–2383 (2005).
33. Zhang, Z., Harrison, P. M., Liu, Y. & Gerstein, M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res.* **13**, 2541–2558 (2003).
34. Vanin, E. F. Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* **19**, 253–272 (1985).
35. Monzó, C., Liu, T. & Conesa, A. Transcriptomics in the era of long-read sequencing. *Nat. Rev. Genet.* **26**, 681–701 (2025).
36. Aguzzoli Heberle, B. et al. Mapping medically relevant RNA isoform diversity in the aged human frontal cortex with deep long-read RNA-seq. *Nat. Biotechnol.* **43**, 635–646 (2025).
37. Leung, S. K. et al. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Rep.* **37**, 110022 (2021).
38. Ransom, L. S. et al. Human brain small extracellular vesicles contain selectively packaged, full-length mRNA. *Cell Rep.* **43**, 114061 (2024).
39. Liu, C. S. et al. RNA isoform diversity in human neurodegenerative diseases. *eNeuro* **11**, ENEURO.0296-24.2024 (2024).
40. Mikheenko, A., Prjibelski, A. D., Joglekar, A. & Tilgner, H. U. Sequencing of individual barcoded cDNAs using Pacific Biosciences and Oxford Nanopore Technologies reveals platform-specific error patterns. *Genome Res.* **32**, 726–737 (2022).
41. Al'Khafaji, A. M. et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nat. Biotechnol.* **42**, 582–586 (2024).
42. DeKosky, S. T. & Scheff, S. W. Synapse loss in frontal cortex biopsies in Alzheimer's disease: correlation with cognitive severity. *Ann. Neurol.* **27**, 457–464 (1990).
43. Salat, D. H., Kaye, J. A. & Janowsky, J. S. Prefrontal gray and white matter volumes in healthy aging and Alzheimer's disease. *Arch. Neurol.* **56**, 338–344 (1999).
44. Phipson, B. et al. propeller: testing for differences in cell type proportions in single cell data. *Bioinformatics* **38**, 4720–4726 (2022).
45. Elizarraras, J. M. et al. WebGestalt 2024: faster gene set analysis and new support for metabolomics and multi-omics. *Nucleic Acids Res.* **52**, W415–W421 (2024).
46. Lau, S. F., Cao, H., Fu, A. K. Y. & Ip, N. Y. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *Proc. Natl. Acad. Sci. USA* **117**, 25800–25809 (2020).
47. Shippy, D. C. & Ulland, T. K. Genome-wide identification of murine interferon genes in microglial-mediated neuroinflammation in Alzheimer's disease. *J. Neuroimmunol.* **375**, 578031 (2023).
48. Ishii, A. et al. Contribution of amyloid deposition from oligodendrocytes in a mouse model of Alzheimer's disease. *Mol. Neurodegener.* **19**, 83 (2024).
49. Sasmita, A. O. et al. Oligodendrocytes produce amyloid- β and contribute to plaque formation alongside neurons in Alzheimer's disease model mice. *Nat. Neurosci.* **27**, 1668–1674 (2024).
50. Liu, C. S. & Chun, J. isoSeQL: comparing long-read isoforms across multiple datasets. *Bioinformatics* **42**, btaf680 (2026).
51. Costantino, I., Nicodemus, J. & Chun, J. Genomic mosaicism formed by somatic variation in the aging and diseased brain. *Genes* **12**, 1071(2021).
52. Park, J. S. et al. Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat. Commun.* **10**, 3090 (2019).
53. Xiao, X. et al. APP, PSEN1, and PSEN2 variants in Alzheimer's disease: systematic re-evaluation according to ACMG guidelines. *Front. Aging Neurosci.* **13**, 695808 (2021).
54. Lanoiselée, H. M. et al. APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: a genetic screening study of familial and sporadic cases. *PLoS Med.* **14**, e1002270 (2017).
55. Franzmeier, N., Rubinski, A., Neitzel, J. & Ewers, M. The BIN1 rs744373 SNP is associated with increased tau-PET levels and impaired memory. *Nat. Commun.* **10**, 1766 (2019).
56. Chapuis, J. et al. Increased expression of BIN1 mediates Alzheimer's genetic risk by modulating tau pathology. *Mol. Psychiatry* **18**, 1225–1234 (2013).
57. Liu, C. C., Liu, C. C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat. Rev. Neurol.* **9**, 106–118 (2013).
58. Yen, J. J. & Yu, I. I. The role of ApoE-mediated microglial lipid metabolism in brain aging and disease. *Immunometabolism* **5**, e00018 (2023).
59. Biosciences, P. (2019).
60. Guan, P., Zhu, D. & Wang, P. Meloxicam inhibits apoptosis in neurons by deactivating tumor necrosis factor receptor superfamily member 25, leading to the decreased cleavage of DNA fragmentation factor subunit α in Alzheimer's disease. *Mol. Neurobiol.* **60**, 395–412 (2023).
61. Antonell, A. et al. Cerebrospinal fluid level of YKL-40 protein in preclinical and prodromal Alzheimer's disease. *J. Alzheimer's Dis.* **42**, 901–908 (2014).
62. Connolly, K. et al. Potential role of chitinase-3-like protein 1 (CHI3L1/YKL-40) in neurodegeneration and Alzheimer's disease. *Alzheimer's Dement* **19**, 9–24 (2023).
63. Abuhatzira, L., Shamir, A., Schones, D. E., Schäffer, A. A. & Bustin, M. The chromatin-binding protein HMGN1 regulates the expression of methyl CpG-binding protein 2 (MECP2) and affects the behavior of mice. *J. Biol. Chem.* **286**, 42051–42062 (2011).
64. Farley, S. J., Grishok, A. & Zeldich, E. Shaking up the silence: consequences of HMGN1 antagonizing PRC2 in the Down syndrome brain. *Epigenet. Chromatin* **15**, 39 (2022).

65. Manganas, L. N. et al. Author Correction: BASP1 labels neural stem cells in the neurogenic niches of mammalian brain. *Sci. Rep.* **11**, 20008 (2021).
66. Zhou, M. et al. Targeted mass spectrometry to quantify brain-derived cerebrospinal fluid biomarkers in Alzheimer's disease. *Clin. Proteom.* **17**, 19 (2020).
67. Ihara, M. et al. Sept4, a component of presynaptic scaffold and Lewy bodies, is required for the suppression of alpha-synuclein neurotoxicity. *Neuron* **53**, 519–533 (2007).
68. Kinoshita, A. et al. Identification of septins in neurofibrillary tangles in Alzheimer's disease. *Am. J. Pathol.* **153**, 1551–1560 (1998).
69. Shehadeh, L., Mitsi, G., Adi, N., Bishopric, N. & Papapetropoulos, S. Expression of Lewy body protein septin 4 in postmortem brain of Parkinson's disease and control subjects. *Mov. Disord.* **24**, 204–210 (2009).
70. Li, F., Liu, A., Zhao, M. & Luo, L. Astrocytic Chitinase-3-like protein 1 in neurological diseases: Potential roles and future perspectives. *J. Neurochem.* **165**, 772–790 (2023).
71. Wang, T. et al. snPATHO-seq, a versatile FFPE single-nucleus RNA sequencing method to unlock pathology archives. *Commun. Biol.* **7**, 1340 (2024).
72. Grindberg, R. V. et al. RNA-sequencing from single nuclei. *Proc. Natl. Acad. Sci. USA* **110**, 19802–19807 (2013).
73. Krishnaswami, S. R. et al. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* **11**, 499–524 (2016).
74. Hardwick, S. A. et al. Single-nuclei isoform RNA sequencing unlocks barcoded exon connectivity in frozen brain tissue. *Nat. Biotechnol.* **40**, 1082–1092 (2022).
75. Chun, J. Somatic genomic mosaicism in the brain: Identified mutations provide challenges and opportunities for the clinic. *Medicines* **3**, 648–650 (2022).
76. Rohrback, S., Siddoway, B., Liu, C. S. & Chun, J. Genomic mosaicism in the developing and adult brain. *Dev. Neurobiol.* **78**, 1026–1048 (2018).
77. Kaeser, G. E. & Chun, J. Mosaic somatic gene recombination as a potentially unifying hypothesis for Alzheimer's disease. *Front. Genet.* **11**, 390 (2020).
78. Aguzzoli Heberle, B. et al. Mapping medically relevant RNA isoform diversity in the aged human frontal cortex with deep long-read RNA-seq. *Synapse*. <https://www.synapse.org/#!Synapse:syn52047893> (2025).
79. Leung, S. K. et al. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *SRA*. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA664117> (2021).

Acknowledgements

We are grateful for support by the National Institute on Aging of the NIH under Awards R01 AG065541, R01 AG071465; Department of Defense W81XWH-21-1-0642; and by non-federal funds from The Bruce Ford & Anne Smith Bundy Foundation, and the Larry L. Hillblom Foundation (J.C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Brain specimens were obtained from Emory University, UK Imperial, Southwest Neurodegenerative Diseases Brain Bank, Pittsburgh Neurodegenerative Diseases Brain Bank, and Sepulveda Neurodegenerative Diseases Brain Bank. We sincerely and gratefully thank the donors and their families who shared precious brain materials. We thank

Danielle Jones for administrative help; Laura Wolszon, PhD for helping to obtain human specimens; Kang Liu, PhD at the Sanford Burnham Prebys Medical Discovery Institute Genomics Core for RIN analysis of brain samples; Gwendolyn Kaeser, Ph.D., for assistance with visualization; Dong-Jin E. Kang Sim for statistics advice; and Deron Herr, Ph.D. and Yasuyuki Kihara, Ph.D. for comments on the manuscript.

Author contributions

A.S. performed all experiments; A.S. & C.S.L. analyzed the data; C.S.L. performed bioinformatic analysis on all sequencing data; T.N., C.R.P., D.Z., C.P., V.T., and N.J. aided experiments; A.S., C.S.L. & J.C. wrote the manuscript; J.C. conceived the study and secured funding.

Competing interests

J.C. has an employment relationship with Neurocrine Biosciences, a company that may potentially benefit from the research results. J.C.'s relationship with Neurocrine Biosciences has been reviewed and approved by Sanford Burnham Prebys Medical Discovery Institute in accordance with its conflict of interest policies. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-026-09759-9>.

Correspondence and requests for materials should be addressed to Jerold Chun.

Peer review information *Communications Biology* thanks Ziyang Liu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Dr. Nilanjan Banerjee and Dr. Kaliya Georgieva.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2026