

Droplet-based transcriptome profiling of individual synapses

Received: 25 June 2021

Accepted: 6 December 2022

Published online: 16 January 2023



Muchun Niu^{1,2}, Wenjian Cao^{1,3,11}, Yongcheng Wang^{4,5,12}, Qiangyuan Zhu^{1,11}, Jiayi Luo^{1,6}, Baiping Wang^{1,7}, Hui Zheng^{1,7}, David A. Weitz^{5,8}✉ & Chenghang Zong^{1,9,10}✉

Synapses are crucial structures that mediate signal transmission between neurons in complex neural circuits and display considerable morphological and electrophysiological heterogeneity. So far we still lack a high-throughput method to profile the molecular heterogeneity among individual synapses. In the present study, we develop a droplet-based single-cell (sc) total-RNA-sequencing platform, called Multiple-Annealing-and-Tailing-based Quantitative scRNA-seq in Droplets, for transcriptome profiling of individual neurites, primarily composed of synaptosomes. In the synaptosome transcriptome, or ‘synaptome’, profiling of both mouse and human brain samples, we detect subclusters among synaptosomes that are associated with neuronal subtypes and characterize the landscape of transcript splicing that occurs within synapses. We extend synaptome profiling to synaptopathy in an Alzheimer’s disease (AD) mouse model and discover AD-associated synaptic gene expression changes that cannot be detected by single-nucleus transcriptome profiling. Overall, our results show that this platform provides a high-throughput, single-synaptosome transcriptome profiling tool that will facilitate future discoveries in neuroscience.

Synapses are crucial structures that mediate signal transmission between neurons in complex neural circuits. Advances in microscopy and electrophysiology techniques have unveiled the morphological and electrophysiological heterogeneity existing among individual synapses^{1–5}. To facilitate the characterization of synaptic heterogeneity and the construction of a synapse transcriptome atlas, a high-throughput, transcriptome-profiling method of individual synaptosomes is greatly desired. However, to achieve successful profiling of gene expression

in individual synaptosomes, new technical features of transcriptome profiling beyond the state-of-art scRNA-seq platforms are required. First, individual synaptosomes contain smaller quantities of RNA molecules than single cells or single nuclei. Therefore, a high-sensitivity scRNA-seq assay is desired. Second, after synaptosomes are prepared, the materials require immediate fixation to prevent notable leakage of RNA molecules in downstream steps. Hence, RNA-seq chemistry compatible with fixed samples is demanded. Third, to characterize locally

¹Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. ²Integrative Molecular and Biomedical Sciences Graduate Program, Baylor College of Medicine, Houston, TX, USA. ³Genetics and Genomics Graduate Program, Baylor College of Medicine, Houston, TX, USA.

⁴Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. ⁵Wyss Institute of Bioinspired Engineering, Harvard University, Cambridge, MA, USA. ⁶Cancer and Cell Biology Graduate Program, Baylor College of Medicine, Houston, TX, USA. ⁷Huffington Center on Aging, Baylor College of Medicine, Houston, TX, USA. ⁸Department of Physics and School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. ⁹Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, USA. ¹⁰McNair Medical Institute, Baylor College of Medicine, Houston, TX, USA. ¹¹Present address: Research Center for Analytical Instrumentation, Institute of Cyber-Systems and Control, State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou, China. ¹²Present address: Liangzhu Laboratory, Zhejiang University Medical Center, Hangzhou, China. ✉e-mail: weitz@seas.harvard.edu; czong@bcm.edu

spliced genes in the synapses, a total-RNA-based assay that permits simultaneous detection of both mature and nascent RNA is desired.

To meet these specific technical demands, we report, in the present study, the development of a droplet-based single-cell total-RNA-seq platform. We refer to this assay as Multiple-Annealing-and-Tailing-based Quantitative scRNA-seq in Droplets (MATQ-Drop). The development of MATQ-Drop is based on the previous chemistry of MATQ-seq⁶. MATQ-Drop works with fixed samples and its effective detection of nascent RNA makes it suitable for characterizing local splicing in synaptosomes. It is worth noting that, although the commercial 10x Genomics Chromium platform is broadly accessible^{7–10}, SMART-seq-based chemistry¹¹ on this platform is mainly designed for quantifying mature messenger RNA levels in fresh samples, hence making it unsuitable for transcriptome profiling of synaptosomes.

Using the MATQ-Drop platform, we performed the transcriptome profiling of single synaptosomes of mouse and human brain samples. For convenience, we refer to the transcriptome of synaptosomes as the synaptome. In the synaptome data, we were able to identify various types of neurites, including different subtypes of synaptosomes and neuron–glia junctions. Among different subtypes of synaptosomes, we observed presynaptic and postsynaptic clusters, as well as a special subcluster associated with the synapses in the process of assembly and maturation. Transcriptomic differences between different subclusters can be readily detected. With the effective detection of nascent RNAs, we also characterized the landscape of intron retention for various clusters of synapses. In addition to synaptome profiling, we applied MATQ-Drop to profile the transcriptome of single nuclei for the same brain samples. With both synaptome and the single-nucleus transcriptome, we were able to connect subclusters of synapses to different types of neurons. The differential gene expression and splicing between the synapses and neuronal nuclei were then analyzed. Furthermore, we profiled the synaptosomes isolated from an AD mouse model. We characterized the synaptopathy-associated transcriptome and discovered AD-associated gene expression changes that cannot be detected by single-nucleus transcriptome profiling.

With the effective detection of total RNA, we also successfully generated the cell atlas using only long noncoding RNA (lncRNA) species. This result suggests that MATQ-Drop allows the large-scale identification of cell-type-specific lncRNA species. Furthermore, based on the single-nucleus (sn) transcriptome of the mouse brain, we also conducted a benchmark comparison between MATQ-Drop and 10x Chromium. Our result shows that MATQ-Drop demonstrated up to 135% improvement in gene detection sensitivity across different cell types when compared with the 10x v3.1 platform. Overall, as a total-RNA-based, high-throughput transcriptome platform, MATQ-Drop will provide an alternative high-throughput, high-sensitivity, single-cell transcriptome platform to the 10x Chromium platform.

Results

The chemistry of MATQ-Drop

In the chemistry of MATQ-Drop (Fig. 1a), we first applied 3% paraformaldehyde (PFA) to fix the nuclei. After the crosslinking, we permeabilized the nucleus membrane and performed ten cycles of multiple annealing with MALBAC (multiple annealing and looping–based amplification cycles) primers^{6,12}, which allow efficient hybridization to the internal regions of the transcripts (Fig. 1a). As a result, besides the reverse transcriptions initiated from the poly(adenylated) (poly(A)) tails at the 3′-end of the transcripts, substantial portions of reverse transcriptions were also initiated at the internal regions of transcripts, which warrants efficient total RNA capture. After the reverse transcription step, we washed away the excessive MALBAC primers. We then performed *in situ* poly(A) tailing for the complementary DNA molecules, which we refer to as the dA-tailing step. Next, we washed the processed nuclei and used microfluidic platforms to encapsulate single nuclei together with the barcoded dT20 hydrogel beads in droplets for multiplexed

second-strand synthesis. The barcoded dT20 hydrogel beads were prepared following the procedure described in the inDrop platform¹⁰.

It is worth noting that, different from the ultraviolet light (UV)-triggered release of the barcoded oligos from the beads in the inDrop platform, in the present study we introduced enzymatic release chemistry (Supplementary Fig. 1a and Methods). In this chemistry, we introduced a deoxyuridine (dU) base in the sequence near the 5′-end of the barcoded oligos. In the droplet reaction buffer for the second-strand synthesis, we included the USER enzyme that can cut the oligos at the dU site. As a result, on droplet encapsulation, the dT20 oligos with cell barcodes were efficiently released from the beads. Next, we performed RNA digestion and heat decrosslinking to release cDNA from the nuclei. The barcoded dT20 primers then hybridized to the poly(A) tail of the cDNA molecules to initiate the second-strand synthesis. After the second-strand synthesis was completed, we broke the droplets and collected the aqueous phase, followed by the PCR reaction to amplify the library for next-generation sequencing (Supplementary Fig. 1b).

To validate the successful single-cell barcoding in MATQ-Drop, we performed the standard species-mixing experiment as a control. We mixed equal numbers of fresh human HEK293T and mouse NIH/3T3 cells and then lysed them into nuclei. With the fixed nuclei, we performed the MATQ-Drop assay as described above. In the present study, we used a small aliquot of droplets to generate the sequencing library for technical evaluation. As shown in Fig. 1b, we identified 162 unique high-quality cell barcodes. Based on the species specificity, we unambiguously assigned them to 81 human 293T nuclei, 76 mouse 3T3 nuclei and 5 collision events (Fig. 1c). For each assigned cell barcode, we observed high species specificity of unique molecular identifiers (UMIs), as shown in Fig. 1d (99.7% for 293T nuclei and 99.4% for 3T3 nuclei). In addition, the 162 cell barcodes covered 89% of all uniquely mapped reads (Supplementary Fig. 2a–c), confirming an extremely low cross-barcode contamination rate. For the single-cell total-RNA-seq data generated by MATQ-Drop, we did not observe notable UMI inflation (Supplementary Fig. 2d and Supplementary Text 1).

It is worth pointing out that the major technical advantage of MATQ-Drop, in comparison to matured mRNA-based platforms such as 10x Genomics Chromium, is that we can effectively detect nascent RNAs using the reads mapped to intronic regions (Fig. 1e). With regard to gene detection sensitivity, at the average sequencing depth of ~70,000 raw reads per single nucleus, we detected a median of 21,192 UMIs and 6,575 genes for single 293T nuclei, and 11,286 UMIs and 4,220 genes for single 3T3 nuclei (Fig. 1f,g). As shown in Fig. 1h,i, the gene detection of MATQ-Drop is significantly higher than the sensitivity of other snRNA-seq methods^{8,13}. To further extend the benchmark comparison between MATQ-Drop and 10x Genomics Chromium for cell atlas construction with tissue samples, we also performed an equal footing comparison using the mouse brain samples described below.

Detection of synaptosome subtypes by transcriptome profiling

So far, the major approach in transcriptome profiling of synapses has been based on bulk samples¹⁴. Noticeably, microdissected neurites were used to profile the transcriptome of synapses localized at specific regions of rat hippocampus samples¹⁵. In the present study, in contrast to the bulk-based approach, with the development of MATQ-Drop, we would like to profile the transcriptome of individual synaptosomes. We refer to the transcriptome of individual synaptosomes as a synaptome and we profiled the transcriptome of individual synaptosomes isolated from both mouse and human brain samples on MATQ-Drop platform.

We first profiled the synaptome of the mouse hippocampus. To isolate synaptosomes, we ground out the snap-frozen brain tissue using a Dounce homogenizer (Supplementary Fig. 3a). We then performed FACS to enrich Hoechst-negative subcellular structures

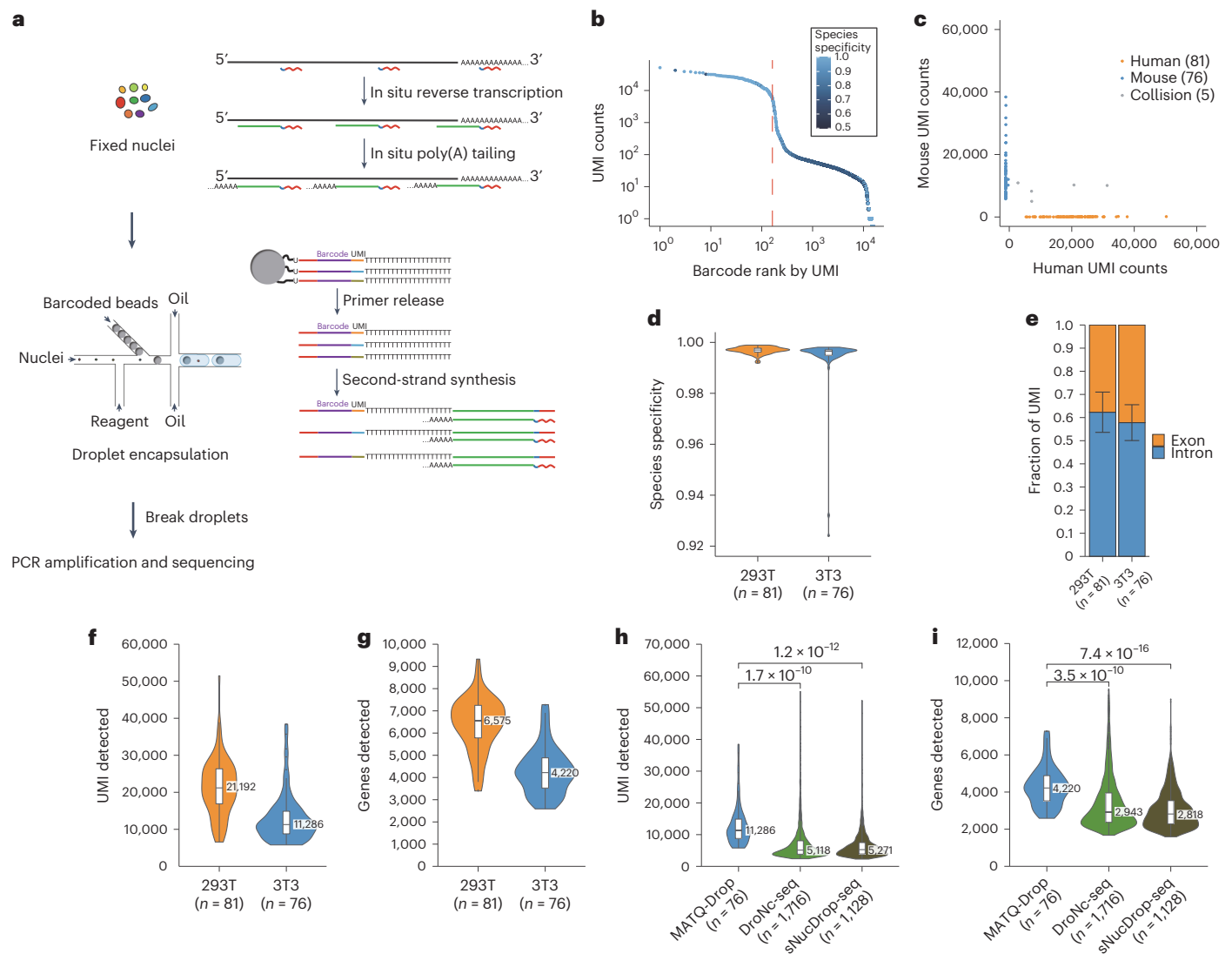


Fig. 1 | Overview of MATQ-Drop and the performance in species-mixing experiment. a, Reaction scheme of MATQ-Drop. In situ reverse transcription and poly(A) tailing are performed on the fixed nuclei, which are then encapsulated in droplets with barcoded hydrogel beads. Inside the droplet, barcoded dt20 primers are enzymatically released from the beads to capture the poly(A) tail of cDNA released from the nuclei. After the barcoded second-strand synthesis has been accomplished, the emulsion is broken and the product can be amplified and sequenced. **b**, Identification of the barcodes representing true nuclei in the species-mixing experiment. Barcodes are ordered from the largest to the smallest UMI counts. On the UMI counts versus barcode rank plot, the knee

point (162, red dashed line) indicates the threshold for true nuclei. **c**, Species annotation of the 162 nuclei identified. **d**, Species specificity of UMIs. **e**, Fractions of UMIs in exons and introns (mean $\pm$ s.d.). **f, g**, Detection sensitivity of MATQ-Drop in UMI counts (**f**) and gene counts (**g**). **h, i**, Comparison of detection sensitivity between MATQ-Drop and other major snRNA-seq methods^{8,13} for single NIH/3T3 nuclei. UMI detection (**h**) and gene detection (**i**) (*P* values calculated using two-sided Student's *t*-test). **d, f–i**, Boxplot shows the center line and median, box limits the upper and lower quartiles, whiskers the 1.5 $\times$ interquartile range (IQR) and points the outliers.

with sizes $<5 \mu\text{m}$ (Supplementary Fig. 3b,c). We want to point out that the main reason for conducting this rapid isolation of synaptosomes is to preserve RNA quantity and quality. In comparison to this sorting-based rapid isolation procedure, we also performed synaptosome profiling using synaptosomes isolated from the standard gradient centrifugation-based enrichment method. As a result, we observed a significant reduction in gene detection, leading to the poor resolution of synaptosome clustering as described below.

With the transcriptome data of individual synaptosomes, we first performed the unsupervised clustering analysis following the standard Seurat v.4 integration pipeline¹⁶. It is interesting that we identified 15 primary clusters, among which 12 clusters were synapse associated (Fig. 2a,b, Supplementary Figs. 4 and 5 and Supplementary Tables 1

and 2). Among the 12 synapse-associated clusters, we noticed that the Syn1 cluster exhibits a 3.5-fold increase of nascent RNA proportion compared with the rest of the synapses (average intronic fraction 29.9% versus 8.5%; Fig. 2c). In Syn2 and Syn4 clusters, we observed the upregulation of *Grin2b*, *Pclo* and *Bsn* (*Pclo* and *Bsn* are known presynaptic scaffold genes). In contrast, in the Syn3 cluster, we observed the upregulation of postsynaptic genes, including *Shank1* and *Shank3* (Fig. 2b and Supplementary Table 2).

Besides the overrepresentation of presynaptic features in Syn2 and Syn4 clusters and postsynaptic features in Syn3 cluster, we also observed additional synapse subclusters that are defined by specific markers: Syn5: *Zbtb20*; Syn6: *Chd9*; Syn7: *Purg*; Syn8: *Nopchap1*; Syn9: *Apc*; Syn10: *Hivep3*; Syn11: *Kmt2d*; and Syn12: *Ksr2* (Fig. 2a,b). Among

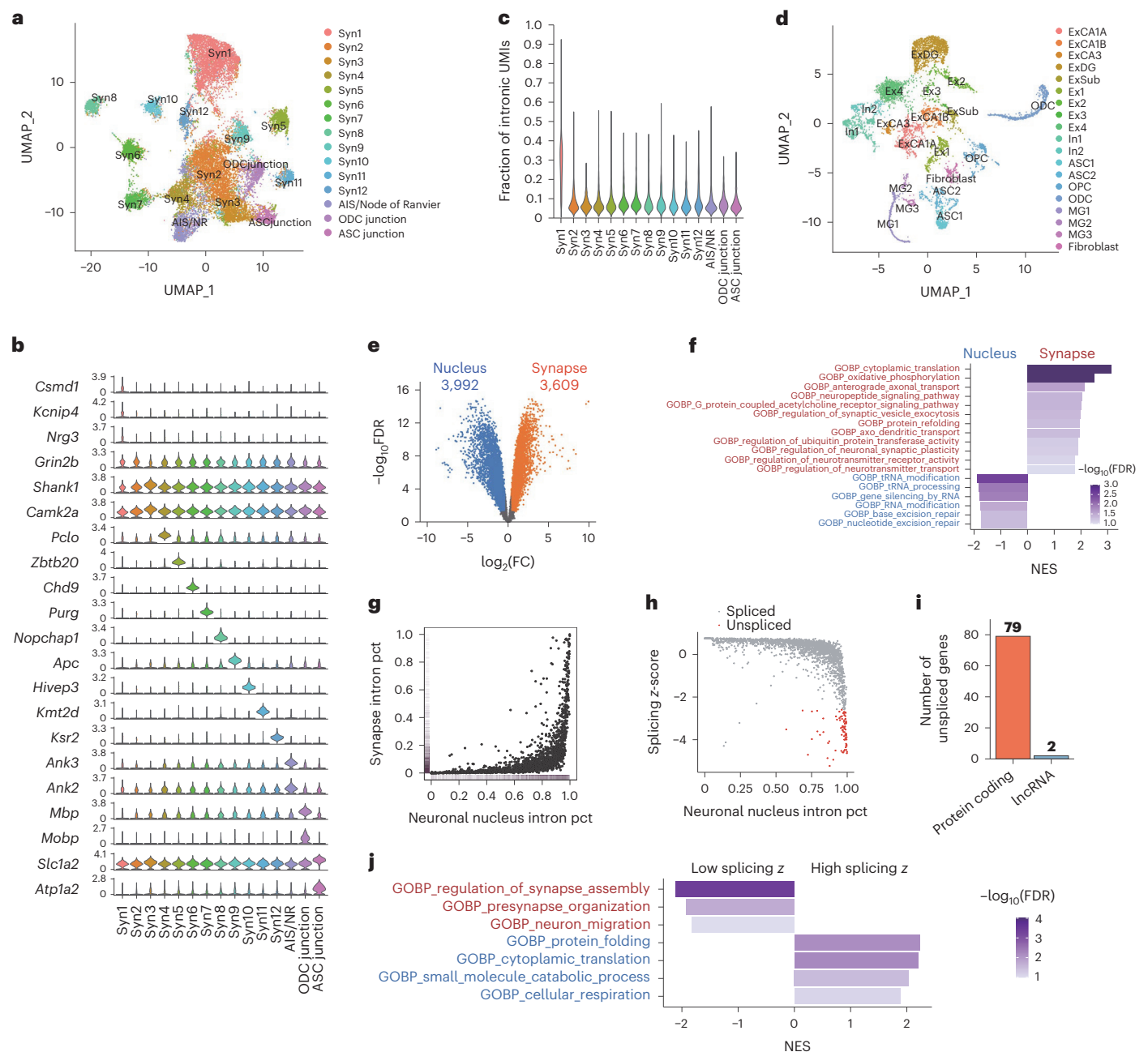


Fig. 2 | The mouse hippocampus cell atlas and synaptome atlas. a, UMAP visualization of synaptosome and neuron-glia junction subtypes of the mouse hippocampus. **b**, Violin plots showing the expression of subcellular-type-enriched markers in different clusters (Syn1: *Csmd1*, *Kcnp4* and *Nrg3*; Syn2: *Grin2b*; Syn3: *Shank1* and *Camk2a*; Syn4: *Pclo*; Syn5: *Zbtb20*; Syn6: *Chd9*; Syn7: *Purg*; Syn8: *Nopchap1*; Syn9: *Apc*; Syn10: *Hivep3*; Syn11: *Kmt2d*; Syn12: *Ksr*; AIS/NR: *Ank3* and *Ank2*; ODC junctions: *Mbp* and *Mobp*; ASC junctions: *Slc1a2* and *Atp1a2*). **c**, Fraction of intronic UMIs for each synaptosome and neuron-glia junction cluster in the mouse hippocampus. **d**, UMAP visualization of 19 cell populations identified in the primary clustering analysis. ExCA1A-B, CA1 excitatory neuron A-B; ExCA3, CA3 excitatory neuron; ExDG, DG excitatory

neuron; ExSub, subiculum excitatory neuron; Ex1-4, other excitatory neurons 1-4; In1-2, inhibitory neurons 1-2; ASC1-2, astrocytes 1-2; MG1-3, microglia 1-3. **e**, Volcano plots showing the exon-based DEGs between synapses and neuronal nuclei. **f**, Pathways enriched in the synapses and nuclei, identified through GSEA. NES, Normalized enrichment score. **g**, The average intronic UMI fraction in synapses versus neuronal nuclei, with the marginal rug plot indicating density. **h**, Identification of the unspliced synaptic genes in neurons. **i**, Number of the unspliced synaptic genes grouped by gene type. **j**, Pathways enriched in unspliced and fully spliced genes, identified through preranked GSEA based on splicing score.

their marker genes, mutations in *Zbtb20* have been shown to affect the synaptic structures by altering ZBTB20 protein localization in subneuronal compartments¹⁷; *Purg* (detected in Syn7) was reported to display strong and early upregulation during synaptogenesis in primary mouse hippocampal neurons¹⁸; *Ksr2* (detected in Syn12) contributes to calcium-mediated ERK (extracellular signal-regulated

kinase) signaling¹⁹. In addition to synapse-associated clusters, we also observed axon initial segments and nodes of Ranvier (AIS/NR cluster) and neuron-glia junctions including neuron-oligodendrocyte junctions (ODC junctions) and neuron-astrocyte junctions (ASC junctions) (Fig. 2a,b, Supplementary Fig. 4a,b and Supplementary Tables 1 and 2).

Mouse synaptome based on density gradient centrifugation

Next, we used the freshly prepared mouse samples as the control to compare the effects of different synaptosome isolation procedures on synaptome profiling. We performed synaptome profiling using the synaptosomes isolated from the standard sucrose density gradient-based ultracentrifugation protocol. In comparison to the direct sorting-based procedure, we observed 53% fewer genes detected per synaptosome (median 146 genes versus 306 genes; Supplementary Fig. 6a), which is probably due to RNA decay and the leakage during the extensive processing time without PFA fixation. Although we still observed some evidence of the regional distribution for a few clusters including Syn1 (*Kcnp4*), Syn6 (*Chd9*) and Syn8 (*Nopchp1*) on the uniform manifold approximation and projection (UMAP), the overall clustering result has low resolution with certain ambiguity (Supplementary Fig. 6b,c).

Connect the mouse synapse clusters to neuron subtypes

To compare the synaptome with the nucleus transcriptome, next we performed single-nucleus transcriptome profiling for the same mouse hippocampus. Based on the nascent RNA expression matrix, we identified nine subtypes of excitatory neurons from different subregions, two subtypes of inhibitory neurons, astrocytes (ASCs), oligodendrocyte progenitor cells (OPCs), oligodendrocytes (ODCs), microglia (MGs) and fibroblasts (Fig. 2d, Supplementary Fig. 7a–d, and Supplementary Tables 3 and 4), and an average of 83% of UMIs detected can be attributed to introns (Supplementary Fig. 7e).

Next, we compared the single-nucleus transcriptome with the synaptome to identify the connection between synapse clusters and neuron subtypes. It is interesting that we did not identify statistically significant connections. This result supports the capture of different synapse transcriptomic states with the synaptosomes prepared from fresh mouse brain samples^{20,21}. Furthermore, we would like to investigate, among different synaptic states (synapse clusters), whether we can identify the subclusters associated with neuron subtypes. To do so, we used the highly variable genes across neuronal nuclei (Methods) as the coordinates for supervised clustering analysis. As shown in Supplementary Fig. 8a–f, we did observe the evident association between the distribution of synaptosomes and different neuronal subtypes. But the subclusters are less separated, probably because they share the features of the same synaptic states. Based on the mouse synaptome data, we conclude that there are two layers of synapse heterogeneities: the first layer is associated with synaptic states and the second with neuron subtypes.

Differential expression between synapses and nuclei in mice

Next, we performed differentially expressed gene (DEG) analysis between synapses and nuclei. Considering the dominance of matured mRNA in the synapses, in the present study we used the exon-based gene expression matrix for DEG analysis between synapses and nuclei to avoid potential bias by gene length. As a result, we identified 3,609 synapse-enriched genes and 3,992 nucleus-enriched genes (Fig. 2e and Supplementary Table 5) (DEGs defined as $(\text{abs}(\log_2(\text{fold-change (FC)})) > \log_2(1.3))$, false discovery rate (FDR) < 0.05)). As expected, synapse-enriched genes were overrepresented with synaptic signaling and protein synthesis pathways. In contrast, the nucleus-enriched genes were overrepresented with gene regulation, RNA processing and DNA repair pathways (Fig. 2f).

Local splicing landscape in mouse hippocampal synapses

Studies have shown that the genes with retained introns are crucial for the intraneuronal transport of the transcript²². Furthermore, synaptic alternative splicing is also vital for quick modulation of synaptic functions^{23–26}. Next, based on nascent RNA detection in MATQ-Drop data, we characterized the synaptic transcripts with intron retention (Methods). We observed only a small percentage of unspliced synaptic transcripts (81 out of 2015, 4%), including 79 protein-coding genes and

2 lncRNAs (Fig. 2g–i and Supplementary Tables 6 and 7). When we performed gene set enrichment analysis (GSEA) for the genes preranked by splicing z-score (Methods), on one end of the enrichment, the spliced transcripts were enriched for basic cellular activities such as protein synthesis and metabolism; on the other end of the enrichment, the unspliced transcripts were enriched for synapse assembly, organization and neuron migration pathways, suggesting the important role of local splicing in synaptogenesis (Fig. 2j).

Synaptome profiling of the frozen human brain samples

To test whether the single-synaptosome RNA-seq pipeline can be adopted to construct the human synaptome, next we processed the frozen human brain hippocampus and prefrontal cortex (PFC) samples from two individuals. It is worth noting that we specifically requested the dentate gyrus regions of the hippocampus samples. For two human hippocampus samples (Supplementary Table 8), we generated the transcriptome of 10,428 single subcellular structures (Fig. 3a and Supplementary Fig. 9a) and we observed 11 major clusters corresponding to different types of neurite structures. As shown in Supplementary Fig. 9b, the batch effects between samples were undetectable. In Fig. 3b,c, we annotated these clusters as subtypes of synapses and neuron–glia junctions based on the well-known molecular markers enriched in those subcellular structures (Supplementary Table 9 and 10). In total, we assigned six synapse-associated clusters: four synapse clusters with high RNA abundance (we denote them as HI-synapses), one synapse cluster with lower RNA abundance (we denote this as LO-synapse) (Fig. 3d,e) and another synapse cluster containing relatively higher nascent transcripts (we denote this as N-synapse) (Fig. 3a,h). In contrast, 15 synapse-associated clusters were observed in mouse hippocampus samples. The potential reason for this discrepancy is that mouse brain samples were freshly prepared right after the sacrifice of the mice, whereas the human brain samples often had long postmortem intervals (PMIs, 12 and 13 h, respectively, for the two brain samples we sequenced), which could lead to the decay of transcripts and distort the synapse clusters. Therefore, synapse states were probably better preserved in the mouse samples.

Despite the potential distortion of the transcriptome caused by the PMI, when we compared the synaptome profile with single-nucleus transcriptome profiles described below, the four HI-synapse clusters could be associated with excitatory neurons in CA1, CA3 and DG regions and inhibitory neurons, respectively. It is worth pointing out that the inhibitory HI-synapse cluster (Synapse_In in Fig. 3a) can be further classified into three subtypes by additional subclustering analysis (Supplementary Fig. 10a–c). When we profiled the synaptome of two human PFC samples, similar clusters of HI-synapses, LO-synapses and N-synapses were observed (Supplementary Fig. 11a–c). The HI-synapses can also be subclustered into excitatory and inhibitory subtypes. The expression of marker genes in each cluster is shown in Supplementary Fig. 11d,e. The detection sensitivity is shown in Supplementary Fig. 11f,g.

Additional validation of synaptosome isolation procedure

In the present study, we further validated the synaptosome isolation procedure using frozen human brain samples (Methods). First, we confirmed the enrichment of synaptic proteins synaptophysin and synapsin-1 in the Hoechst-negative subcellular structures using western blotting (Supplementary Fig. 12a). In addition, we also performed immunostaining for the Hoechst-negative particles using presynapse marker synaptophysin and postsynapse marker PSD95 (Supplementary Fig. 12b–e). Using flow cytometry analysis, we observed that 60.1% of Hoechst-negative particles were synaptophysin positive and 38.1% were PSD95 positive. Next, we sorted out double-positive particles (34.6%) and performed transcriptome profiling. When we combined its transcriptome data with the transcriptome data of the total Hoechst-negative particles of the same sample, we observed a complete overlap between Hoechst-negative particles and double-positive

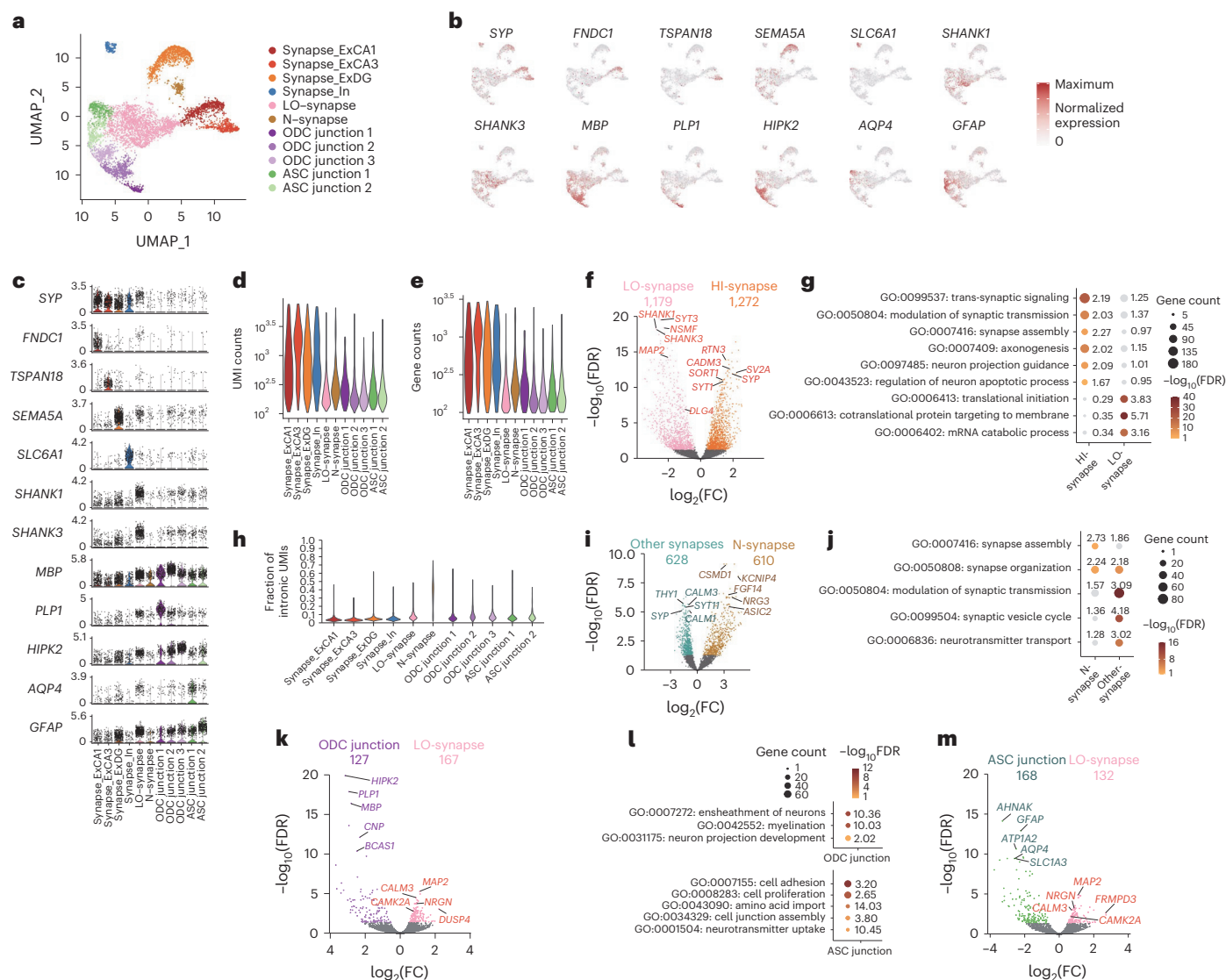


Fig. 3 | The human hippocampus synaptome atlas. a, UMAP visualization of synaptosome and neuron-glia junction subtypes of the human hippocampus. **b,c**, UMAP feature plots (**b**) and violin plots (**c**) showing the expression of subcellular-type-enriched markers in different clusters (HI-synapse: *SYP*; CA1 excitatory HI-synapse (Synapse_ExcA1): *FND1*; CA3 excitatory HI-synapse (Synapse_ExcA3): *TSPAN18*; DG excitatory HI-synapse (Synapse_ExcDG): *SEMA5A*; inhibitory HI-synapse (Synapse_In): *SLC6A1*; LO-synapses: *SHANK1* and *SHANK3*; ODC junctions: *MBP*, *PLP1* and *HIPK2*; ASC junctions: *AQP4* and *GFAP*). **d,e**, Detection sensitivity for each cluster in UMI counts (**d**) and gene counts (**e**). **f**, Volcano plot showing the DEGs between hippocampus HI-synapses and LO-

synapses. **g**, Pathway enrichment of hippocampus HI-synapse-enriched and LO-synapse-enriched genes. Fold enrichment is labeled next to the dots. **h**, Fraction of intronic UMIs for each synaptosome and neuron-glia junction cluster in the human hippocampus. **i**, Volcano plot showing the DEGs between N-synapses and other synapses in the hippocampus. **j**, Pathway enrichment of hippocampus N-synapse-enriched and other-synapse-enriched genes. Fold enrichment is labeled next to the dots. **k**, Volcano plot showing the DEGs between ODC junctions and LO-synapses. **l**, Pathway enrichment of ODC junction-enriched and ASC junction-enriched genes. Fold enrichment is labeled next to the dots. **m**, Volcano plot showing the DEGs between ASC junctions and LO-synapses.

particles (Supplementary Fig. 12f,g), indicating that the vast majority of Hoechst-negative particles are synaptosomes and neuron-glia junctions. It is worth noting that neuron-glia junctions were reported to express synaptic proteins²⁷, therefore they were enriched in the synaptophysin and PSD95 double-positive population.

We also sorted out the double-negative particles (36.4%) and performed transcriptome profiling. As a result, we observed that the corresponding transcriptome had extremely low RNA abundance per particle, equivalent to 4% of RNA yield compared with the double-positive population (Supplementary Fig. 12h). Hence, when we profile the transcriptome of all Hoechst-negative particles, the double-negative particles will be effectively filtered out by RNA abundance cutoff and do not contribute to the synaptome. Therefore, we

conclude that the unbiased profiling of the Hoechst-negative population authentically represented the transcriptome of synaptosomes and neuron-glia junctions.

Differential expression between human synaptosome subtypes

Next, we performed DEG analysis to identify transcriptomic differences between the HI-synapses and the LO-synapses for the hippocampus synaptome (Fig. 3f and Supplementary Table 11) and the PFC synaptome (Supplementary Fig. 11h and Supplementary Table 11). We identified 1,272 and 807 HI-synapse-enriched genes ($\text{abs}(\log_2(\text{FC})) > \log_2(1.3)$, $\text{FDR} < 0.05$) in the hippocampus and PFC, respectively, both including well-established synaptic vesicle genes (*SYT1*, *SYP*, *SV2A* and *SORT1*)²⁸.

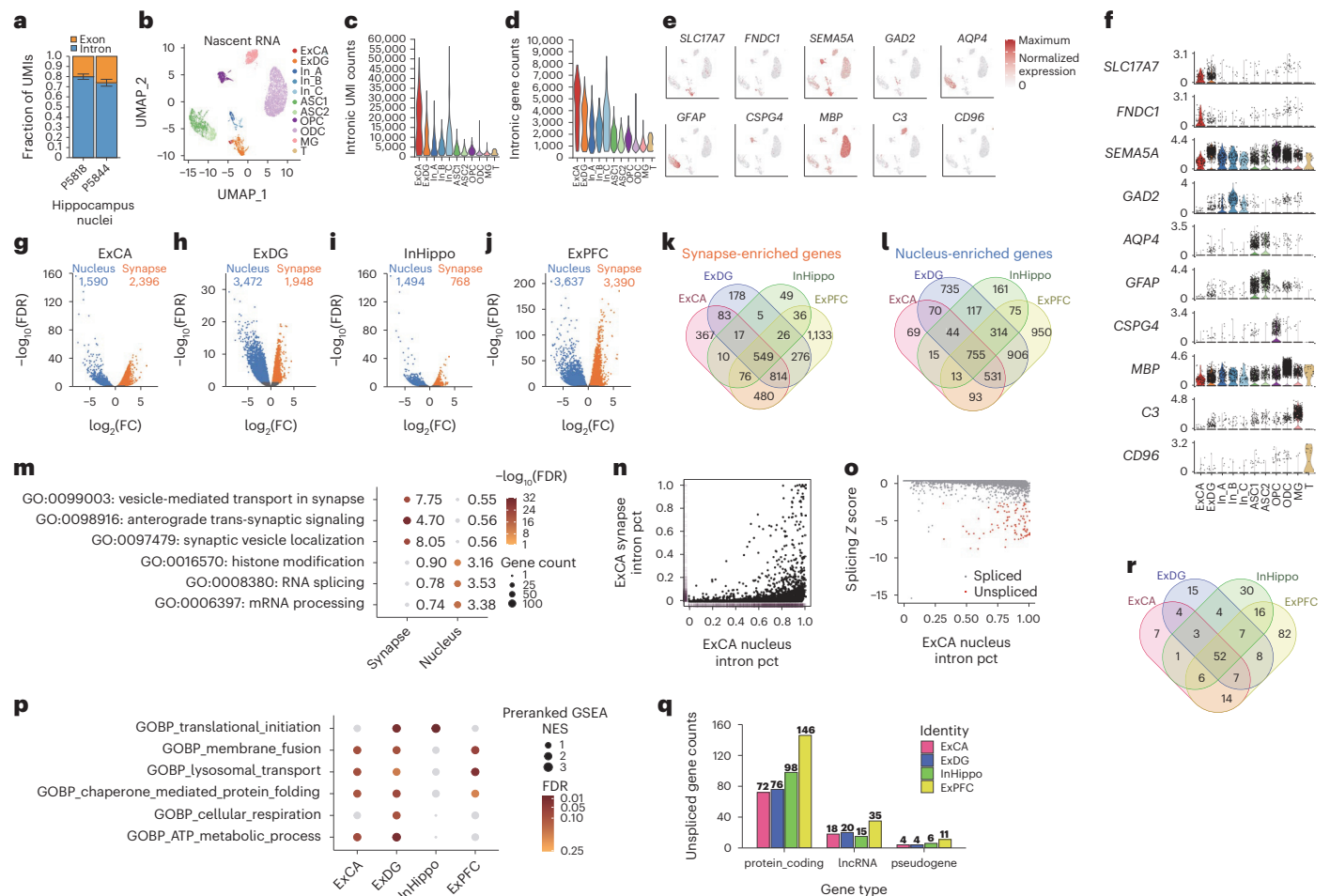


Fig. 4 | Nascent RNA-based cell atlas of the human hippocampus. **a**, Fractions of UMIs in exons and introns (mean $\pm$ s.d.; P5818: $n = 4,127$ single nuclei; P5844: $n = 3,985$ single nuclei). **b**, UMAP visualization of 11 cell populations identified in the primary clustering analysis. ExCA, CA excitatory neuron; ExDG, DG excitatory neuron; In_A-C, inhibitory neuron A-C; ASC1–2, astrocyte 1–2; T, T cells. **c,d**, Detection sensitivity for each hippocampus cluster in UMI counts (**c**) and gene counts (**d**). **e**, UMAP feature plots showing the log(normalized expression) of the well-established, cell-type-specific markers in different clusters (excitatory neuron: *SLC17A7*; CA neuron: *FND1*; DG neuron: *SEMA5A*; inhibitory neuron: *GAD2*; ASCs: *AQP4* and *GFAP*; OPC: *CSPG4*; ODC: *MBP*; MG: *C3*; T: *CD96*). **f**, Violin plots showing the marker gene expression level in different clusters. **g**, UMAP visualization of clustering results using only lncRNA expression matrix, colored by nascent RNA-based annotation. **h**, UMAP visualization of clustering results using only lncRNA expression matrix, colored by nascent RNA-based annotation. **i–j**, Volcano plots showing the exon-based DEGs between HI-synapses and nuclei for four neuronal subtypes. ExCA, CA

excitatory neurons (**g**); ExDG, DG excitatory neurons (**h**); InHippo, hippocampal inhibitory neurons (**i**); ExPFC, prefrontal cortex excitatory neurons (**j**). **k,l**, Venn diagram showing the overlap of synapse-enriched (**k**) or nucleus-enriched (**l**) genes among four neuronal subtypes. **m**, Pathway enrichment of shared synapse-enriched and nucleus-enriched genes. Fold enrichment is labeled next to the dots. **n**, The average intronic UMI fraction in HI-synapses versus nuclei of CA excitatory neurons, with the marginal rug plot indicating density. **o**, Identification of the unspliced synaptic genes in CA excitatory neurons. **p**, Pathways enriched in fully spliced genes, identified through preranked GSEA based on splicing score. NES, normalized enrichment score. **q**, Number of the unspliced synaptic genes grouped by gene type. **r**, Venn diagram of the unspliced synaptic genes across four neuronal subtypes and the list of 41 protein-coding genes detected in all four neuronal subtypes.

Next, we identified 1,179 and 855 LO-synapse-enriched genes in the hippocampus and PFC, respectively. Among the enriched genes in the LO-synapses, we noticed the dendrite marker gene *MAP2* (ref. 29), the well-known postsynaptic scaffold genes *SHANK1*, *SHANK3* and *DLG4* (ref. 30) and the postsynaptic gene *SYT3* (ref. 31). The differential marker genes show the enrichment of presynaptic transcriptomic features in the HI-synapses (which resembles mouse Syn2 and Syn4 clusters) and the enrichment of postsynaptic transcriptomic features in the LO-synapses (which resemble the mouse Syn3 cluster). In the enriched functionals and pathways of the HI-synapse cluster, we observed synaptic signaling and axonogenesis in both the hippocampus (Fig. 3g and Supplementary Table 12) and the PFC (Supplementary Fig. 11i and Supplementary Table 12). For the LO-synapse cluster, the protein synthesis and mRNA catabolism-related pathways are enriched (Fig. 3g, Supplementary Fig. 11i and Supplementary Table 13),

suggesting that high protein synthesis activities and turnover rates exist in the postsynapses. The enrichment of the mRNA catabolism pathway in the LO-synapses also explains the low RNA abundance observed in these synapses in comparison to HI-synapses.

It is also worth pointing out that, in contrast to the clear difference in RNA abundance between HI-synapses and LO-synapses detected in the human brain, we did not observe such a discrepancy in mouse brains (Supplementary Fig. 4c,d). This difference requires future investigation. It could be caused by species differences between humans and mice or by different RNA decay rates between presynapses and postsynapses during the postmortem time. If the postsynapses have a much higher RNA decay rate than the presynapses, then, with the long PMIs for the human samples we used, a substantial portion of RNA in postsynapses might have been decayed before they could be captured by MATQ-Drop.

It is interesting that, although most synapses displayed a low intron fraction (7.85%, on average), we did notice that one cluster (N-synapses) exhibited a significantly higher intron fraction (30.79% on average (Fig. 3h) and PFC data in Supplementary Fig. 13a), which is the human counterpart of the mouse Syn1 cluster. Through the DEG analysis between N-synapses and the rest of the synapses (Fig. 3i and Supplementary Table 14), we observed that the genes enriched in N-synapses were overrepresented in synapse assembly and synapse organization gene sets (Fig. 3j). In contrast, the genes involved in synaptic signaling were overrepresented in other clusters of synapses (Fig. 3j and Supplementary Fig. 13b,c). The results indicate that N-synapses represent the immature synapses that are in the process of assembly and maturation. The significantly higher percentage of intronic reads in the N-synapses also buttresses the important roles of unspliced nascent RNA and the related local splicing in the synaptic assembly and maturation process.

Transcriptome profiling of human neuron–glia junctions

Besides the clusters of synaptosomes, we also identified two major cell–cell junctions formed between neurons and glial cells: ODC junctions and ASC junctions in the hippocampus (Fig. 3a) and PFC (Supplementary Fig. 11a–e). Both the noncompact myelin gene, *CNP*, and compact myelin genes, *PLP1* and *MBP*, were highly expressed in ODC junctions (Fig. 3k and Supplementary Table 15). It is worth noting that the upregulated genes in the ODC junctions are enriched in the myelination process (Fig. 3l), which is consistent with the well-known axon–ODC signaling related to the myelination process^{32,33}. More importantly, the detection of transcripts in ODC junctions in our data suggests the importance of local translation at the ODC junctions during myelination. This indication of local translation is also consistent with the recent study by Wake et al.³⁴.

In the ASC junctions, we observed the local enrichment of ASC-specific genes, for example, *GFAP*, *ATPIA2*, *AQP4* and *SLC1A3* (Fig. 3m and Supplementary Table 15). These upregulated genes are enriched in cell adhesion, proliferation and neurotransmitter uptake pathways (Fig. 3l). Consistent with our observation of transcripts enriched in the ASC junctions, local translation has also been recently observed in astrocyte peripheral processes³⁵. Overall, the transcriptome profiling of neuron–glia junctions allows the comprehensive identification of locally translated genes in the cell–cell junctions between neurons and glial cells. The functional roles of these genes are worth future investigation.

Effective cell atlas construction using only nascent RNAs

To identify the connection between different subtypes of synaptosomes and different subtypes of neurons, we next applied MATQ-Drop to profile the total-RNA-based transcriptome for 8,112 single nuclei isolated from two dissected frozen human hippocampi. First, in the single-nucleus transcriptome data, we noticed that the portion of reads that represented nascent RNAs in the brain samples was significantly higher than that in the cell-line samples. We observed that 78% of the UMIs were mapped to intronic regions in the brain samples (Fig. 4a), in contrast to 63% of intronic reads in the cell lines (Fig. 1e).

Next, we calculated the gene expression matrix based only on the unspliced transcript sequence with the reads mapped to the intron regions. Using this nascent RNA-based gene expression matrix, we evaluated its performance in constructing a cell atlas for the human hippocampus samples that we profiled. In Fig. 4b and Supplementary Table 16, using the standard Seurat v4 unsupervised clustering pipeline, we successfully identified the following ten primary clusters in the hippocampal nuclei: two excitatory neuronal subtypes from the Cornu Ammonis region (ExCA) and dentate gyrus (ExDG), respectively; three inhibitory neuronal subtypes (In_A, In_B, In_C); four glial cell types, including two subtypes of ASCs (ASC1–2), OPCs, ODCs and MGs. As shown in Supplementary Fig. 14, no batch-to-batch variations

were observed. In terms of detection sensitivity, the UMI and gene detection are shown in Fig. 4c,d. The markers of each cluster were also consistent with well-established cell-type-specific markers (Fig. 4e,f and Supplementary Table 17), suggesting robust cell typing using a nascent-transcript-based gene expression matrix.

Similar to the hippocampus, we also successfully constructed the cell atlas for the human PFC sample of the same individuals (Supplementary Fig. 15). With the profiling of 939 single nuclei, we successfully identified 15 primary clusters with high confidence, which included 6 excitatory neuronal subtypes (Ex1–6), 4 inhibitory neuronal subtypes (In1–4), 4 glial cell types (including ASCs, OPCs, ODCs and MGs) and endothelial cells (ENDs) (Supplementary Fig. 15a–c and Supplementary Table 16). The markers of each cluster were also consistent with the standard cell-type-specific markers (Supplementary Fig. 15d,e and Supplementary Table 17). Based on the expression of the previously reported layer-specific markers³⁶, the six excitatory neuron subtypes were assigned to different cortical layers (Supplementary Fig. 15h,i). Among the inhibitory neurons from both regions, we identified eight subtypes with additional subclustering analysis (Supplementary Fig. 16a,b and Supplementary Table 18). The unique combinations of marker genes were detected in the subtypes of inhibitory neurons (Supplementary Fig. 16c,d and Supplementary Table 19).

Differential expression between synapses and nuclei in human

Now with the single-nucleus transcriptome data from the same tissues of the synaptome, we were able to connect the subclusters in the synaptome to different neuronal nucleus types based on the shared marker genes (Fig. 3a and Supplementary Tables 10 and 17). We have identified that three HI-synapse clusters could be connected to the excitatory neurons in hippocampal CA1, CA3 and DG regions, and another HI-synapse cluster could be connected to inhibitory neurons (Fig. 3a). Next, we investigated the differential patterns of RNA expression between the synaptic clusters and the associated nuclei in a neuron-type-specific manner.

In total, we identified an average of 2,126 synapse-enriched genes and 2,548 nucleus-enriched genes (Fig. 4g–j and Supplementary Table 20). In Fig. 4k,l, we examined the overlapped genes between different neuronal subtypes. We identified that there were a total of 4,099 synapse-enriched genes and a total of 4,848 nucleus-enriched genes; 549 synapse-enriched genes and 755 nucleus-enriched genes were shared by all 4 neuronal subtypes, respectively (Fig. 4k,l). Next, in the functional analysis, we observed that the 549 shared synapse-enriched genes were overrepresented in the pathways directly related to synaptic signaling (Fig. 4m and Supplementary Table 21). In contrast, the 755 nucleus-enriched genes were overrepresented in the epigenetic regulation and RNA-processing pathways (Fig. 4m and Supplementary Table 22).

Local splicing landscape in different subtype of synapses

Similar to the mouse data, for all four clusters in HI-synapses, we observed a long tail of outliers with clear evidence of intron retention (Fig. 4n, Supplementary Fig. 17a–c and Supplementary Table 23). It is worth noting that 84.5% of synaptic transcripts were already spliced (intron percentage <5%). Next, we compared the intron percentage between synapses and nuclei for each gene and calculated the splicing z-score (Fig. 4o, Supplementary Fig. 17d–f and Supplementary Table 24). We then performed GSEA with the genes preranked by the splicing z-score and we observed that the essential cellular functions such as translation, protein folding and metabolism were significantly enriched in the fully spliced genes (Fig. 4p). This result confirms that the fully spliced genes detected in synapses are mainly responsible for basic cellular functions. Based on the splicing z-score, we also identified the unspliced genes with statistical significance. In total, we detected 256 genes from different HI-synapse subtypes, including 49 lncRNAs, 11 pseudogenes and 196 protein-coding genes

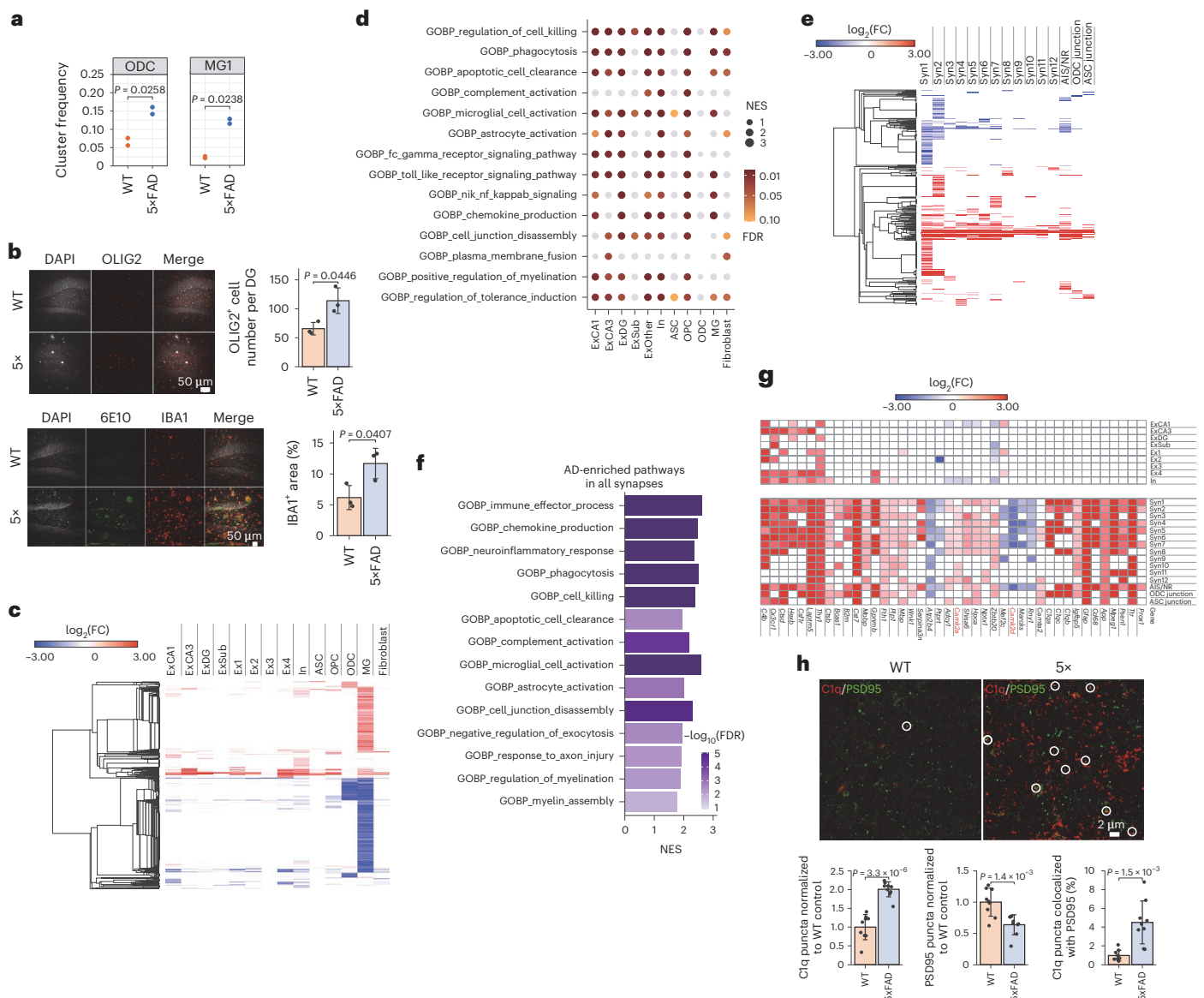


Fig. 5 | AD-associated synaptopathy. a, The comparison of cell frequency for ODC and MG1 cells between 5x FAD mice and WT mice. Two-sided Student's *t*-test was performed and *P* values were indicated on the plot. **b**, Immunofluorescence staining and quantification of ODC marker OLIG2 and MG marker IBA1 in 5x FAD and WT mouse hippocampus brain sections (three mice per genotype, mean $\pm$ s.d., Student's *t*-test). **c**, Heatmap showing the fold-change of intron-based DEGs ($\text{abs}(\log_2(\text{FC})) > \log_2(1.3)$, $\text{FDR} < 0.05$) between 5x FAD and WT mice in different types of nuclei. **d**, Functional enrichment of the DEGs between the single-nucleus transcriptome of 5x FAD and WT for different cell types. **e**,

Heatmap showing the fold-change of DEGs ($\text{abs}(\log_2(\text{FC})) > \log_2(1.3)$, $\text{FDR} < 0.05$) between 5x FAD and WT mice in different synaptosome and neuron–glia junction subtypes. **f**, Pathways enriched in 5x FAD synapses compared with WT mice. **g**, Heatmap showing the fold-change of synapse–AD–DEGs shared by at least six subtypes and their intron-based expression fold-changes in the AD nuclei. **h**, Immunofluorescence staining and quantification of C1q-tagged synapses in 5x FAD and WT mouse hippocampus brain sections (three mice for each genotype and three sections for each mouse, mean $\pm$ s.d.; two-sided Student's *t*-test was performed).

(Fig. 4q). It is of interest that only 41 unspliced protein-coding genes were shared by all 4 synapse subclusters (Fig. 4r). This result indicates that a significant portion of local synaptic splicing is uniquely associated with specific synapse types.

Characterization of synaptopathy in AD synaptome

As a hallmark of AD, β -amyloid plaques are known to impair synaptic function and induce synaptopathy. It has been shown that β -amyloid plaques can induce an inflammatory response that activates MGs to prune synapses^{37,38} and block postsynaptic *N*-methyl-D-aspartic acid receptors and, therefore, suppress trans-synaptic signaling³⁹. Current profiling of transcriptomic changes associated with AD

has been done only with snRNA-seq^{40,41}. In the present study, we applied the MATQ-Drop to characterize the synaptome changes in the 5x FAD mouse model and examine whether different synapse subtypes have different responses to β -amyloid plaques. In total, we profiled the transcriptome of 6,989 single nuclei and 20,456 single Hoechst-negative particles isolated from two wild-type (WT) and two 5x FAD mice.

From single-nucleus transcriptome data, we first observed a 2.3-fold overrepresentation of ODCs compared with WT mice in terms of cell-type composition (Fig. 5a). This result is probably due to the response to axon demyelination. We also observed a 5.4-fold increase of the major MG subtype (MG1) in the 5x FAD mice, indicating

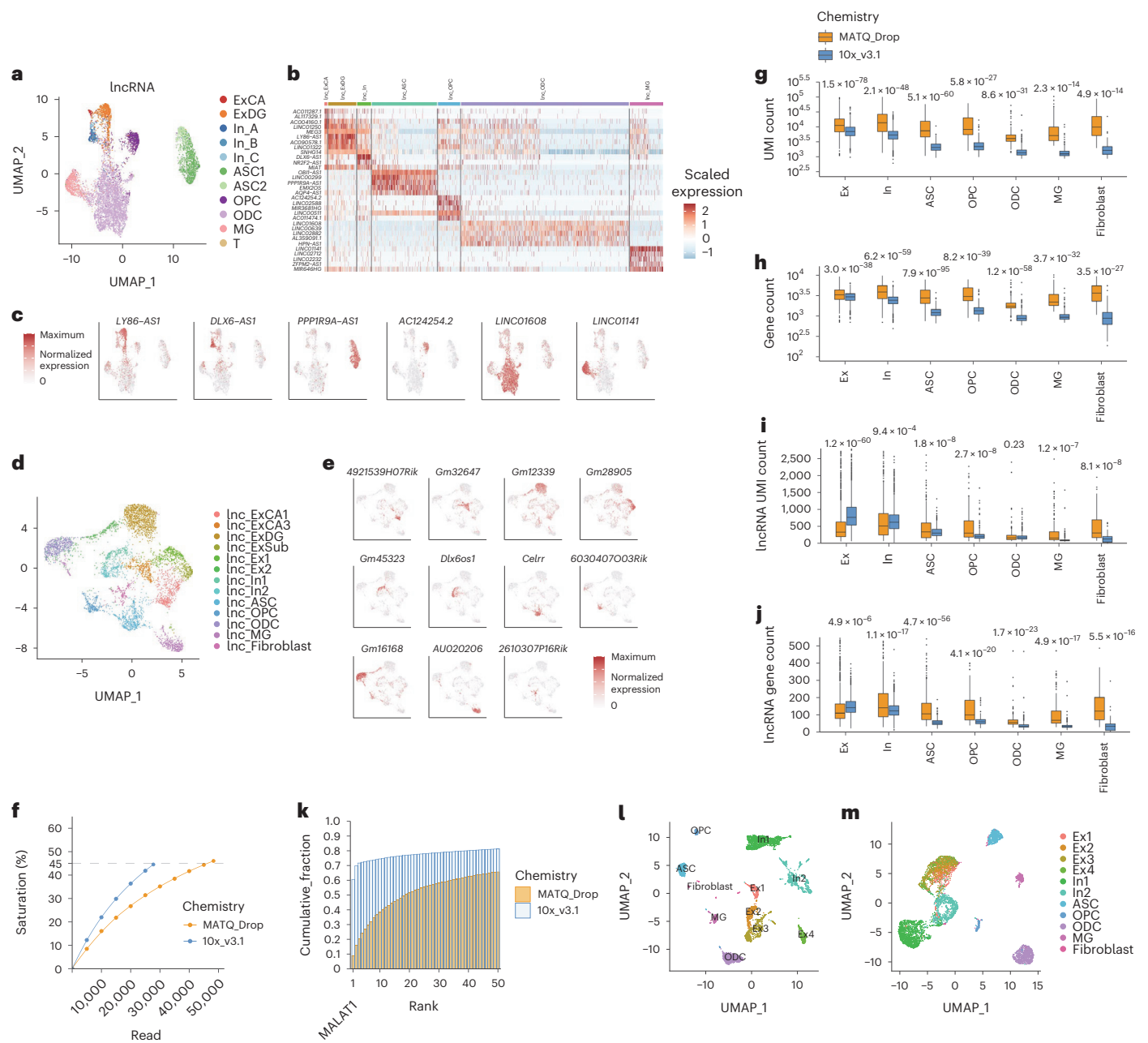


Fig. 6 | The cell typing using only lncRNA species and the detection sensitivity comparison between 10x Chromium and MATQ-Drop. a, UMAP visualization of clustering results using only lncRNA gene expression matrix from the single-nucleus transcriptome of the human hippocampus. **b**, Heatmap showing the scaled expression levels of cell-type-specific lncRNA genes. **c**, UMAP feature plot showing the log(normalized expression level) of cell-type-specific lncRNA genes (excitatory neuron: *LY86-AS1*; inhibitory neuron: *DLX6-AS1*; ASC: *PPP1R9A-AS1*; OPC: *AC124254.2*; ODC: *LINC01608*; MG: *LINC01141*). **d**, UMAP visualization of clustering results using only lncRNA expression matrix from the single-nucleus transcriptome of mouse hippocampus. **e**, UMAP feature plot showing the log(normalized expression level) of cell-type-specific lncRNAs (CA1 excitatory neuron: *4921539H07Rik*; CA3 excitatory neuron: *Gm32647*; DG excitatory neuron: *Gm12339*; subiculum excitatory neuron: *Gm28905*; inhibitory neurons: *Gm45323* and *Dlx6s1*; ASC: *Celrr*; OPC: *60304070O3Rik*; ODC: *Gm16168*; MG: *AU020206*; fibroblast: *2610307P16Rik*). **f**, Determination of the sequencing depth to allow benchmark comparison at the same sequencing saturation. **g, h**, Transcript-based

detection sensitivity (UMI count (**g**) and gene count (**h**)) compared with 10x Chromium⁴¹. **i, j**, lncRNA detection sensitivity (UMI count (**i**) and gene count (**j**)) compared with 10x Chromium⁴¹. **g–j**, The data were derived from two MATQ-Drop WT mouse samples (1,520 single excitatory neurons, 479 inhibitory neurons, 440 ASCs, 180 OPCs, 284 ODCs, 164 MGs and 108 fibroblasts) and one 10x Chromium v3.1 sample (1,711 single excitatory neurons, 2,344 inhibitory neurons, 493 ASCs, 162 OPCs, 798 ODCs, 259 MGs and 76 fibroblasts). **g–j**, Boxplots represent the center line and median, box limits the upper and lower quartiles, whiskers the 1.5× IQR and points the outliers. Two-tailed Student's *t*-test was performed and the Benjamini–Hochberg P_{adj} values were labeled in the plot. **k**, Accumulated fraction of UMIs on the axis of the ranked lncRNA genes. **l**, UMAP showing the transcript-based unsupervised clustering results with 10x Chromium v3.1 mouse brain single-nucleus transcriptome dataset. **m**, UMAP showing the lncRNA-based unsupervised clustering results with 10x Chromium v3.1 data, cell types and subtypes colored based on the transcript-based clustering in **l**.

an activated inflammatory response (Fig. 5a). The increased proportions of ODCs and MGs in 5xFAD models are consistent with previous studies^{40,41} and are further confirmed through the immunofluorescence staining on the brain sections (Fig. 5b). Next, for each neuronal subtype and glial cell type, we identified the DEGs associated with AD based on nascent RNA (Fig. 5c and Supplementary Table 25) and mature RNA, respectively (Supplementary Fig. 18 and Supplementary Table 26). In particular, we observed that MGs consistently displayed the highest numbers of DEGs, suggesting more sensitive roles of these cells in disease response compared with other cell types (Fig. 5c and Supplementary Fig. 18), which are also consistent with the previous study⁴². When we performed GSEA, myelination and multiple inflammatory response pathways, including cell killing, complement activation and chemokine production, were enriched in AD across various cell types (Fig. 5d). It is worth emphasizing that, although similar pathways were enriched in GSEA for different cell types (Fig. 5d), the DEGs are not identical for different cell types (Fig. 5c), suggesting that there are different response mechanisms to the amyloid pathology among different cell types.

Next, we identified DEGs of each cluster of synapses and neuron–glia junctions between the 5xFAD and WT mice in the hippocampal synaptome (Fig. 5e). In total, 410 genes with significant DEGs ($\text{abs}(\log_2(\text{FC})) > \log_2(1.3)$, $\text{FDR} < 0.05$) were identified among different clusters, among which 42 genes were shared by more than half of synapse clusters and 246 genes were unique to single clusters (Fig. 5e and Supplementary Table 27). In line with the single-nucleus results, neuroinflammatory response, complement activation and myelination pathways were significantly enriched in the AD synaptosomes (Fig. 5f), indicating the general inflammatory stress associated with β -amyloid plaques. In addition, we also observed the enrichment of cell junction disassembly and negative regulation of exocytosis pathways, indicating synapse loss and decreased synaptic function. The subtype-specific pathway enrichments for different synapse clusters are summarized in Supplementary Fig. 19.

For the 42 AD DEGs shared by all synapse subtypes, we plotted the corresponding gene expression changes in nuclei in Fig. 5g (nascent RNA-based DEGs; top: nuclei, bottom: synapses). It is worth noting that 24 synapse AD DEGs cannot be detected from the nucleus transcriptome data. Furthermore, eight genes exhibited opposite dysregulation directions from the DEG changes based on the nucleus transcriptome data. It is interesting that we also observed that three complement component genes, *C1qa*, *C1qb* and *C1qc*, were significantly upregulated in the synapses but not significantly in the nuclei, indicating a potential role of local translation of these components in complement-mediated synapse pruning. It is desirable to unveil how these complement component transcripts are transported to the abnormal synapses that require pruning. Consistent with the transcriptome results, in the immunofluorescence staining, we also observed a significant increase of C1q puncta and a decrease of PSD95 puncta in the 5xFAD hippocampus compared with WT mice, suggesting increased complement-related inflammation response and synapse loss. More importantly, the percentage of C1q-tagged synapses increased by approximately fourfold in the 5xFAD hippocampus, which validated the upregulated, complement-mediated synapse pruning in AD (Fig. 5h). Compared with the overall hippocampus area, the C1q puncta numbers demonstrated more dramatic fold-changes in the synapses, which highlighted the synapse-related pathology in AD.

We also observed the bifurcated expression of two calcium/calmodulin-dependent protein kinase II (CaMKII) genes: *Camk2a* and *Camk2d* (Fig. 5g, red-labeled genes), which suggest a switch of CaMKII isoforms in AD that potentially impacted synaptic plasticity⁴³. Other unique DEGs observed in the synapses are also worth detailed investigation for their functional roles in the future.

Construction of cell atlas using only lncRNA species

It is worth pointing out that, different from mature RNA-based droplet platforms, the total RNA-based chemistry of MATQ-Drop allows the efficient detection of lncRNAs. Next, we examined whether we could successfully identify the cell types using only the lncRNA expression matrix. The successful construction of a cell atlas using only lncRNA species will indicate that cell-type-specific lncRNA species or cell-type-specific composition of lncRNA species exist. As shown in Fig. 6a, by unsupervised clustering, we achieved robust construction of the cell atlas for the human hippocampus at the cell-type and -subtype resolution. The clustering result is also consistent with nascent RNA-based clustering (Supplementary Table 28). The lncRNA-based cell atlas of human PFC was also successfully constructed (Supplementary Fig. 20a–h and Supplementary Text 2). Our results show that cell-type-specific lncRNA markers can be systematically identified by MATQ-Drop (Fig. 6b,c and Supplementary Table 29).

For the mouse hippocampus, we also successfully constructed the cell atlas using only lncRNA species (Fig. 6d). The clustering result is highly consistent between lncRNA-based and nascent RNA-based clustering (Fig. 2d and Supplementary Table 30). As a result, cell-type-specific lncRNA markers were systematically identified (Fig. 6e, Supplementary Fig. 21a,b and Supplementary Table 31). It is worth noting that lncRNAs with poly(A) tails can also be detected using SMARTer chemistry on the Fluidigm platform⁴⁴. However, MATQ-Drop chemistry allows the detection of the complete spectrum of lncRNAs, including those with poly(A) tails and those without poly(A) tails. Furthermore, the droplet platform offers higher throughput than the Fluidigm platform in identifying cell-type-specific lncRNA species.

Benchmark comparison between MATQ-Drop and 10x Chromium

Using the MATQ-Drop-based, single-nucleus transcriptome data of mouse hippocampus, we next performed a benchmark comparison against the 10x Chromium 3' v3.1 platform with the mouse brain single-nucleus transcriptome dataset by 3' v3.1 chemistry (10x Genomics online data; <https://www.10xgenomics.com/resources/datasets/5k-adult-mouse-brain-nuclei-isolated-with-chromium-nuclei-isolation-kit-3-1-standard>). For equal footing comparison, we downsampled the sequencing depth to allow the samples to reach the same sequencing saturation level of 45% (Fig. 6f). When counted based on transcripts, MATQ-Drop detected a median of 11,148 UMIs and 3,392 genes for single neuronal nuclei, and 5,634 UMIs and 2,346 genes for glial nuclei. Both are significantly higher than the 10x Chromium data (median 5,836 UMIs and 2,618 genes for single neuronal nuclei or 1,513 UMIs and 994 genes for single glial nuclei; Fig. 6g,h). Overall, MATQ-Drop demonstrated up to 135% (neurons 30%, ASCs 128%, OPCs 121%, ODCs 99%, MGs 135%) sensitivity improvement compared with 10x Chromium v3.1 platform.

In Fig. 6i (UMI count) and Fig. 6j (gene count), we compared the sensitivity of the MATQ-Drop and 10x Chromium v3.1 platform in the detection of lncRNA. Although statistically significant, both methods demonstrated essentially comparable lncRNA gene detection in neurons, whereas MATQ-Drop demonstrated an approximately onefold improvement of lncRNA detection in glial cell types compared with 10x Chromium. However, when we examined the 10x Chromium data in detail, we noticed that a single lncRNA gene, *Malat1*, contributed to 60% of the total UMI count (Fig. 6k). This biased detection is probably due to the large portion of the AT-rich sequence in this gene, therefore allowing more efficient internal hybridization by oligo(dT) primers than other genes. In contrast, we did not observe a substantial contribution by one gene in MATQ-Drop.

This biased lncRNA detection in 10x Chromium could preclude it from generating a high-resolution, lncRNA-based cell atlas. In the unsupervised clustering based on only lncRNA with 10x Chromium v3.1 data (Fig. 6l,m), we observed that, although different neuron

types and glial cell types were well separated, the excitatory neuronal subtypes (Ex1–4) cannot be well distinguished. Overall, in terms of lncRNA gene detection, MATQ-Drop shows significantly less bias than 10x Chromium and, as a result, MATQ-Drop performs better in identifying subtypes of cells.

Discussion

In the present study, we developed a total RNA detection-based, high-throughput scRNA-seq assay. Supported by the chemistry and sensitivity of MATQ-Drop, we profiled the transcriptome of individual synapses in high throughput. We successfully detected different subtypes of synaptosomes and other types of junctions between neurons and nonneuronal cells. The enrichment of different functional pathways between synaptosome subtypes was also observed, supporting the existence of phenotypical heterogeneity between different synaptosomes. We also showed that different synaptosome subtypes could be connected to different types of neurons.

It is worth pointing out that, during the process of synaptosome isolation, we cannot completely rule out potential contamination by the RNA molecules released from the cytoplasm and nuclei. However, we reason that the contamination is random and infrequent and, therefore, will not skew the gene expression profile of individual synaptosomes. This reasoning is supported by our robust detection of synaptosome clusters across different brain samples. It is also worth pointing out that another potential source of contamination could be the condensates dissociated from the nucleus or cytoplasm. These condensates are probably enriched with a large number of RNA transcripts that may affect our clustering analysis. In the present study, we reason three possible scenarios for condensate contamination. The first scenario is that the synaptosomes could be contaminated by a common type of condensate. To examine this, we identified the common genes detected in >50% of synaptosomes. Under this criterion, only 17 genes were detected using human hippocampus synaptome data. From the gene ontology (GO)-CC (cellular component) functional enrichment with $P_{\text{adj}} < 0.01$, we observed that all enriched GO terms are synapse related. Therefore, we can exclude the possibility of contamination of common condensate from the cytoplasm or nucleus. The second scenario is that the synaptosomes are contaminated by heterogeneous condensates, which then lead to the different clusters we observed in the data. However, we have observed clear synapse-related marker genes for different clusters in both mouse and human synaptome data, and the functional enrichment between clusters is also related to the synapses. The third scenario is that the synaptosomes are contaminated by both common and heterogeneous condensates. But the rationale described above for ruling out the first two scenarios can be used to refute this scenario. Overall, we would like to point out the potential sources of contamination that could affect the specificity of our approach. On the other hand, based on the reasoning described above, we expect these condensate contaminations to be infrequent and random; therefore, they will not affect our clustering analysis and functional analysis that are based on hundreds of individual synaptosomes in each cluster.

Besides synaptome profiling, MATQ-Drop can also be used to construct a cell atlas. More importantly, we showed that we could successfully construct a cell atlas using only lncRNA species. Overall, the MATQ-Drop platform permits the efficient characterization of synaptic heterogeneity and large-scale cell atlas construction. In the future, MATQ-Drop can be readily applied to other neurological and neurodegenerative diseases and shed new insights into understanding synaptic biology. It could also be used as a new tool to construct the brain connectome.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information,

acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-022-01635-1>.

References

- Gupta, A., Wang, Y. & Markram, H. Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science* **287**, 273–278 (2000).
- Husi, H., Ward, M. A., Choudhary, J. S., Blackstock, W. P. & Grant, S. G. Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. *Nat. Neurosci.* **3**, 661–669 (2000).
- Ibanez-Sandoval, O. et al. Electrophysiological and morphological characteristics and synaptic connectivity of tyrosine hydroxylase-expressing neurons in adult mouse striatum. *J. Neurosci.* **30**, 6999–7016 (2010).
- Cizeron, M. et al. A brainwide atlas of synapses across the mouse life span. *Science* **369**, 270–275 (2020).
- Zhu, F. et al. Architecture of the mouse brain synaptome. *Neuron* **99**, 781–799 e710 (2018).
- Sheng, K., Cao, W., Niu, Y., Deng, Q. & Zong, C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat. Methods* **14**, 267–270 (2017).
- Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
- Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
- Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
- Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
- Ramskold, D. et al. Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
- Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622–1626 (2012).
- Hu, P. et al. Dissecting cell-type composition and activity-dependent transcriptional state in mammalian brains by massively parallel single-nucleus RNA-Seq. *Mol. Cell* **68**, 1006–1015.e1007 (2017).
- Hafner, A. S., Donlin-Asp, P. G., Leitch, B., Herzog, E. & Schuman, E. M. Local protein synthesis is a ubiquitous feature of neuronal pre- and postsynaptic compartments. *Science* **364**, eaau3644 (2019).
- Cajigas, I. J. et al. The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. *Neuron* **74**, 453–466 (2012).
- Stuart, T. et al. Comprehensive Integration of single-cell data. *Cell* **177**, 1888–1902 e1821 (2019).
- Jones, K. A. et al. Neurodevelopmental disorder-associated ZBTB20 gene variants affect dendritic and synaptic structure. *PLoS ONE* **13**, e0203760 (2018).
- Frese, C. K. et al. Quantitative map of proteome dynamics during neuronal differentiation. *Cell Rep.* **18**, 1527–1542 (2017).
- Dougherty, M. K. et al. KSR2 is a calcineurin substrate that promotes ERK cascade activation in response to calcium signals. *Mol. Cell* **34**, 652–662 (2009).
- Montgomery, J. M. & Madison, D. V. Discrete synaptic states define a major mechanism of synapse plasticity. *Trends Neurosci.* **27**, 744–750 (2004).
- Montgomery, J. M. & Madison, D. V. State-dependent heterogeneity in synaptic depression between pyramidal cell pairs. *Neuron* **33**, 765–777 (2002).

22. Buckley, P. T. et al. Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons. *Neuron* **69**, 877–884 (2011).
23. Glanzer, J. et al. RNA splicing capability of live neuronal dendrites. *Proc. Natl Acad. Sci. USA* **102**, 16859–16864 (2005).
24. Bell, T. J. et al. Cytoplasmic BK(Ca) channel intron-containing mRNAs contribute to the intrinsic excitability of hippocampal neurons. *Proc. Natl Acad. Sci. USA* **105**, 1901–1906 (2008).
25. Bell, T. J. et al. Intron retention facilitates splice variant diversity in calcium-activated big potassium channel populations. *Proc. Natl Acad. Sci. USA* **107**, 21152–21157 (2010).
26. Aoto, J., Martinelli, D. C., Malenka, R. C., Tabuchi, K. & Sudhof, T. C. Presynaptic neurexin-3 alternative splicing trans-synaptically controls postsynaptic AMPA receptor trafficking. *Cell* **154**, 75–88 (2013).
27. Hughes, A. N. & Appel, B. Oligodendrocytes express synaptic proteins that modulate myelin sheath formation. *Nat. Commun.* **10**, 4125 (2019).
28. Koopmans, F. et al. SynGO: an evidence-based, expert-curated knowledge base for the synapse. *Neuron* **103**, 217–234.e214 (2019).
29. Caceres, A., Banker, G., Steward, O., Binder, L. & Payne, M. MAP2 is localized to the dendrites of hippocampal neurons which develop in culture. *Brain Res.* **315**, 314–318 (1984).
30. Naisbitt, S. et al. Shank, a novel family of postsynaptic density proteins that binds to the NMDA receptor/PSD-95/GKAP complex and cortactin. *Neuron* **23**, 569–582 (1999).
31. Awasthi, A. et al. Synaptotagmin-3 drives AMPA receptor endocytosis, depression of synapse strength, and forgetting. *Science* **363**, eaav1483 (2019).
32. Hines, J. H., Ravanelli, A. M., Schwindt, R., Scott, E. K. & Appel, B. Neuronal activity biases axon selection for myelination in vivo. *Nat. Neurosci.* **18**, 683–689 (2015).
33. Mensch, S. et al. Synaptic vesicle release regulates myelin sheath number of individual oligodendrocytes in vivo. *Nat. Neurosci.* **18**, 628–630 (2015).
34. Wake, H. et al. Nonsynaptic junctions on myelinating glia promote preferential myelination of electrically active axons. *Nat. Commun.* **6**, 7844 (2015).
35. Sakers, K. et al. Astrocytes locally translate transcripts in their peripheral processes. *Proc. Natl Acad. Sci. USA* **114**, E3830–E3838 (2017).
36. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
37. Hong, S. et al. Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* **352**, 712–716 (2016).
38. Roy, E. R. et al. Type I interferon response drives neuroinflammation and synapse loss in Alzheimer disease. *J. Clin. Invest.* **130**, 1912–1930 (2020).
39. Shankar, G. M. et al. Natural oligomers of the Alzheimer amyloid-beta protein induce reversible synapse loss by modulating an NMDA-type glutamate receptor-dependent signaling pathway. *J. Neurosci.* **27**, 2866–2875 (2007).
40. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
41. Habib, N. et al. Disease-associated astrocytes in Alzheimer's disease and aging. *Nat. Neurosci.* **23**, 701–706 (2020).
42. Zhou, Y. et al. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat. Med.* **26**, 131–142 (2020).
43. Zalcman, G., Federman, N. & Romano, A. CaMKII isoforms in learning and memory: localization and function. *Front. Mol. Neurosci.* **11**, 445 (2018).
44. Liu, S. J. et al. Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.* **17**, 67 (2016).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

Methods

Microfluidic device design and fabrication

The design and fabrication of the hydrogel bead generation device and the cell encapsulation device have been previously described⁴⁵.

Barcoded bead synthesis

The hydrogel bead production and barcode synthesis procedures were based on the work by Zilionis et al.⁴⁵. Two modifications were introduced in hydrogel bead production. First, the Acrydite-modified oligonucleotide (sequence in Supplementary Table 32) was designed to contain a dU base instead of a photocleavable moiety. Therefore, the primers can be released by the USER enzyme (New England Biolabs (NEB)) instead of UV exposure. The dim illumination step is eliminated. Second, the concentration of the Acrydite-modified DNA primer was reduced to 40 μM in the acrylamide–primer mix.

After hydrogel bead production, two rounds of split and pool were performed for barcode synthesis. In each round, the hydrogel beads were split into 144 wells; each well contained primers with a unique barcode as the template (sequences in Supplementary Table 32). Bst 2.0 warm-start DNA polymerase was used for barcode extension. The reaction was set at 55 °C for 3 h for the first round of split and pool, and 52 °C for 3 h for the second round. After each extension step, the reaction was quenched with a 1.5 volume of 25 mM EDTA and leftover template oligonucleotides were denatured by alkaline and washed away after the protocol. Exonuclease I digestion was performed to remove primers with failed barcode extension.

Cell culture

HEK293T and NIH/3T3 cells were grown in Dulbecco's modified Eagle's medium/High Glucose medium (Gibco) with 10% fetal bovine serum (FBS, Gibco). Cell culture was passaged every 2–3 d.

Mice

The C57BL/6 WT and 5xFAD mice were obtained from the Jackson Laboratory. Mice were housed four per cage in a pathogen-free mouse facility with free access to food and water on a 12-h light:dark cycle at the ambient temperature of 20.0–22.2 °C and humidity of 30–70%. Female mice were used for all experiments. All procedures were performed following the National Institutes of Health (NIH) guidelines and approval of the Baylor College of Medicine Institutional Animal Care and Use Committee.

Immunofluorescence staining

Animals were perfused transcardially with 4% PFA in 0.1 M phosphate-buffered saline (PBS), pH 7.4, under ketamine (300 mg kg⁻¹) and xylazine (30 mg kg⁻¹) anesthesia. Brains were harvested, postfixed in the same fixative overnight at 4 °C, dehydrated with 30% sucrose in PBS and serially sectioned at 30- μm thickness on a sliding microtome (Leica). For immunofluorescence, sections were permeabilized in PBS/0.1% Triton X-100 for 30 min and blocked with 4% normal donkey serum in PBS/0.1% Triton X-100 for 1 h at room temperature. Sections were then incubated with primary antibodies in 2% serum in PBS/0.1% Triton X-100 overnight at 4 °C: OLIG2 (EMD Millipore, catalog no. AB9610), IBA1 (Wako, catalog no. 019–19741), β -amyloid (clone 6E10, BioLegend, catalog no. 803001), C1q (clone 4.8, Abcam, catalog no. ab182451) and PSD95 (clone 6G6-1C9, Millipore, catalog no. MAB1596). Sections were then washed and incubated with donkey-anti-mouse Alexa Fluor-488-conjugated secondary antibody (Invitrogen, catalog no. A-21202) or donkey-anti-rabbit Alexa Fluor-555-conjugated secondary antibody (Invitrogen, catalog no. A-31572) for 1 h at room temperature. After washing with PBS, sections were incubated with DAPI to stain the nucleus. Images were captured using a Laser-Scanning Confocal Microscopy (Leica) and quantified with ImageJ. Three animals in each group and three sections from each animal were analyzed.

Mouse hippocampus dissection

Mice aged about 9 months were deeply anesthetized with ketamine (300 mg kg⁻¹) and xylazine (30 mg kg⁻¹) solution, intraperitoneally, and perfused with saline. The brains were removed from the skull and adult mouse brain hemispheres were separated in half; the hippocampus was isolated from each hemisphere and immediately frozen in liquid nitrogen.

Cell-line nucleus preparation

Cells were trypsinized and washed twice with PBS. An equal number of HEK293T cells and NIH/3T3 cells were mixed, and then lysed into nuclei by incubating with the ice-cold lysis buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP-40 and 0.1% Tween-20) on ice for 5 min. Before fixation, the nuclei were washed with 3 $\times$ wash buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl₂, and 0.1% Tween-20). For each wash, the nuclei were first centrifuged at 500g and 4 °C for 3 min, the supernatant was aspirated and the nuclei pellet was resuspended in the wash buffer. After the third wash, we resuspended the nuclei in the fixation buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.2% Tween-20 and 3% PFA) and incubated at room temperature for 10 min on an end-over-end rotator to fix the nuclei. Fixation was quenched by mixing with 3:20 volume of 2.5 M glycine. The fixed nuclei were washed twice with the wash buffer and then passed through a 40- μm cell strainer.

Human brain nucleus preparation from frozen samples

Frozen human brain tissues were requested from NIH NeuroBioBank, with the sample information summarized in Supplementary Table 8. We followed the protocol developed by Krishnaswami et al.⁴⁶ to isolate the nuclei from the frozen brain samples. Briefly, the tissues were homogenized with a Dounce homogenizer and 0.1% Triton X-100, followed by 3% PFA fixation at room temperature for 10 min. After quenching and washing away residual PFA, the homogenate was stained with Hoechst. Fluorescence-activated nucleus sorting was performed to unbiasedly collect the Hoechst-positive single nuclei.

Human brain synaptosome preparation from frozen samples

The method for synaptosome preparation is similar to nucleus preparation, but with two major differences: (1) Triton X-100 was omitted in the homogenization buffer; (2) the Hoechst-negative population with a diameter <5 μm was sorted by FACS. The detailed procedure is described as follows. First, an ~2-mm³ section of frozen brain tissues was chopped and rinsed in the homogenization buffer (250 mM sucrose, 25 mM KCl, 5 mM MgCl₂, 10 mM Tris-HCl, pH 8.0, 1 μM dithiothreitol (DTT), 1 $\times$ Halt protease inhibitor cocktail (Thermo Fisher Scientific) and 0.2 U μl^{-1} of RNase in ribonuclease inhibitor (Promega)). The tissue was then transferred to the Dounce homogenizer (Wheaton) and homogenized using five strokes with the loose pestle and ten strokes with the tight pestle. The homogenate was passed through a 40- μm cell strainer and centrifuged at 1,500g for 10 min at 4 °C. The pellet was immediately resuspended in 25 ml of fixation buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl₂ and 3% PFA) and incubated at room temperature for 10 min. Fixation was quenched by mixing with a 3:20 volume of 2.5 M glycine. The fixed subneuronal structures were washed with wash buffer (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl₂ and 0.1% Tween-20) once, passed through another 40- μm cell strainer and stained with Hoechst. FACS was then performed on a BD FACS Aria instrument to enrich the Hoechst-negative synaptosome population <5 μm in diameter, calibrated using standard beads.

Immunostaining of the brain synaptosomes

The fixed subneuronal structures were permeabilized with 0.2% Triton X-100 in PBS for 10 min on ice and then pelleted by 3,000g centrifugation at 4 °C for 5 min. Blocking of nonspecific binding was performed by incubating the samples with 5% bovine serum albumin (BSA) in PBS at

room temperature for 30 min with rotation. The following primary antibodies were used for immunostaining: rabbit-anti-synaptophysin (clone SP11, Invitrogen, catalog no. MA5-14532, 1:60) and mouse-anti-PSD95 (clone 6G6-1C9, Invitrogen, catalog no. MA1-045, 1:400). Primary antibody binding was performed by a 80-min incubation with 0.5% BSA in PBS on an end-over-end rotor at room temperature. The samples were washed 3× with 1 ml of PBS and 0.5% BSA. Secondary antibody binding was performed by a 40-min incubation with 0.5% BSA in PBS on an end-over-end rotor at room temperature, with the following secondary antibodies: goat-anti-rabbit Alexa Fluor-647 (Invitrogen, catalog no. A21244, 1:1,667) and goat-anti-mouse-Cy3 (Invitrogen, catalog no. A10521, 1:1,667). The subneuronal structures were washed 3×, stained with Hoechst 33342 and then submitted to flow cytometry on a BD FACS Aria instrument. The flow cytometry data were analyzed using BD FACSDiva v.8.0.1 software.

Western blotting

To recover protein from fixed samples, we resuspended the samples in the fixation lysis buffer (500 mM Tris-HCl, pH 7.4, 2% sodium dodecylsulfate, 25 mM EDTA, 100 mM NaCl, 1% Triton X-100, 1% NP-40 and 1× Halt protease inhibitor cocktail) and heated at 90 °C for 2 h. Protein concentration was quantified by BioRad DC Protein Assay and 0.5 µg of total protein was loaded for each western blot using the standard protocol. The following primary antibodies were used in the present study: synaptophysin (clone SP11, Invitrogen, catalog no. MA5-14532, 1:200), synapsin-I (clone D12G5, Cell Signaling Technology, catalog no. 5297, 1:1,000), CNPase (clone 11-5B, Millipore, catalog no. MAB326R, 1:500), glial fibrillary acidic protein (GFAP; clone GA5, Millipore, catalog no. MAB360, 1:1,000) and β-actin (clone AC-15, Sigma-Aldrich, catalog no. A1978, 1:2,000).

Permeabilization

Permeabilization of the PFA-fixed subcellular structures is required for efficient primer hybridization. To permeabilize the subcellular structures, we resuspended them in ice-cold PBS with 1% Triton X-100 and incubated them on ice for 5 min. The permeabilized subcellular structures were washed twice with ice-cold PBS containing 0.2% Triton X-100, and then adjusted to the concentration of ~2,300 subcellular structures per µl before proceeding with reverse transcription.

MATQ-Drop procedure

In situ reverse transcription. For ~25,000 subcellular structures, we prepared the following in situ reverse transcription mix: 4 µl of 5× first-strand buffer (Invitrogen), 1 µl of 0.1 M DTT, 1 µl of 1.8% Triton X-100, 0.5 µl of 10 mM dNTP, 0.5 µl of RNaseOUT (Invitrogen), 2 µl of 11.5 µM MALBAC primer mix, 1 µl of Superscript III reverse transcriptase (Invitrogen) and 11 µl of fixed subcellular structures resuspended in PBS. Ten cycles of multiple annealing ramping from 8 °C to 50 °C were performed for efficient primer hybridization and reverse transcription.

In situ poly(A) tailing. The residual primers and any primer dimers were first washed away and the subcellular structures were resuspended in 14.5 µl of PBS with 0.2% Triton X-100. Next, 1 µl of 1 mM dATP (mixed with 3 µM ddATP), 2 µl of 10× terminal transferase buffer (NEB), 2 µl of 2.5 mM CoCl₂ and 0.5 µl of terminal transferase (NEB) were subsequently added to the subcellular structure suspension. The in situ poly(A) tailing reaction was incubated at 37 °C for 4 h and quenched with 1.6 µl of 0.5 M EDTA. In the reaction, we spiked in 1:333 of ddATP to prevent the poly(A) tail from growing too long, at the cost of losing 1 – (332/333)²⁰ = 6% of the amplicons with a poly(A) tail that was too short (<20) for efficient second-strand synthesis.

Barcoded second-strand synthesis. The fixed subcellular structures carrying poly(A)-tailed cDNA were washed and individual subcellular structures were encapsulated with barcoded dT20 hydrogel beads

and the 2× reaction mix using the microfluidic platform as previously described⁴⁵. After droplet encapsulation, the reaction was first incubated at 37 °C for 45 min to release the primers from the beads by USER enzyme (NEB); meanwhile, cDNA was released from RNA templates due to RNA digestion by RNase H (NEB) and RNase I_r (NEB) digestion. Next, a 3-h incubation at 72 °C is performed to allow cDNA to diffuse out of the nucleus. We performed ten cycles of (48 °C, 2 min and 72 °C, 1 min) to allow the barcoded dT20 primers to hybridize to the poly(A) tail of the released cDNA and Deep Vent (exo-) DNA polymerase (NEB) will initiate extension from the barcoded dT20 primers and accomplish second-strand synthesis. It is worth noting that this procedure does not involve a melting step, so each amplicon can only be converted to one double-strand DNA fragment.

Post-barcoding cleanup. After the barcoded second-strand synthesis was completed, the droplet emulsion was broken by mixing the emulsion with 1H,1H,2H,2H-perfluoro-1-octanol (Sigma-Aldrich) in the presence of EDTA, which immediately quenches polymerase activity on droplet breakage and therefore prevents barcode crosstalk. The remaining hydrogel beads in the aqueous phase were removed by centrifugation and the supernatant was purified with 1× AMPure XP beads (Beckman) and eluted in 37.5 µl of nucleus-free water.

The ddTTP sealing of unused bead primers. To minimize barcode crosstalk in the amplification step, it is critical to quench the residual barcoded bead primers by ddTTP. We prepared the following ddTTP sealing mix: 37.5 µl of purified product, 0.5 µl of 10 mM ddTTP, 5 µl of 10× terminal transferase buffer, 5 µl of 2.5 mM CoCl₂ and 1 µl of terminal transferase, and incubated at 37 °C for 3 h. The product was purified with 1× AMPure XP beads (Beckman) and eluted in 41 µl nucleus-free water.

Library amplification. PCR was performed to amplify 41 µl of the purified product by adding 5 µl of 10× ThermoPol Buffer (NEB), 2.5 µl of 10 µM GAT27 primer (GTG AGT GAT GGT TGA GGA TGT GTG GAG), 1 µl of 10 mM dNTP and 0.5 µl of Deep Vent (exo-) DNA polymerase. The following PCR program was used: 95 °C 2 min, 16–18 cycles of (95 °C, 20 s; 63 °C, 20 s; 72 °C 2 min) and 72 °C, 3 min. The amplified product was purified with 0.9× AMPure XP beads (Beckman) and the yield quantified by Qubit (Invitrogen).

Sequencing of MATQ-Drop library

Sequencing library preparation. The transposase approach was used for sequencing library construction (Supplementary Fig. 1b). For each MATQ-Drop library, 10 ng of the amplified product was mixed with 5 µl of tagmentation DNA buffer (Illumina) and 0.6 µl of tagmentation DNA enzyme 2 (TDE2, Illumina), and the volume was brought up to 10 µl by adding nuclease-free water. The transposition mix was incubated at 55 °C for 15 min. Next, the reaction was quenched by adding 0.4 µl of 0.5 M EDTA and the transposase was released by 50 °C heating for 30 min.

To introduce the i5 index, the following 38.25 µl of reaction mix was prepared and added to each tube: 4 µl of 10× ThermoPol Buffer (NEB), 2 µl of 0.1 M MgSO₄, 1 µl of 10 mM dNTP, 1.75 µl of 10 µM Illumina Nextera N5XX indexed primer (AAT GAT ACG GCG ACC ACC GATG ATC TAC AC (i5 index) TCG TCG GCA GCG TC), 1.75 µl of 10 µM MATQ-P700 primer (ACG TGT GCT CTT CCG ATC TCG CCG AAG ATG GTT GAG GA T GTG TGG AGA TA), 0.7 µl of Deep Vent (exo-) DNA polymerase and 28.8 µl of nuclease-free water. The reaction was set on a thermal cycler with the following program: 65 °C, 1 min; 72 °C, 4 min; 95 °C, 2 min; and 7 cycles of (95 °C, 20 s; 57 °C, 30 s; and 72 °C, 1 min) and 72 °C, 2 min. The product was purified with 0.9× AMPure XP beads and eluted in 16 µl of nuclease-free water.

To introduce the i7 index, we prepared the following PCR reaction: 16 µl of preamplified product, 2 µl of 10× ThermoPol Buffer, 0.5 µl of

10 μ M P5-22b primer (AAT GAT ACG GCG ACC ACC GAG A), 0.5 μ l of 10 μ M P7-i7-MATQ indexed primer (CAA GCA GAA GAC GGC ATA CGA GAT (i7 index) GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T), 0.4 μ l of 10 mM dNTP and 0.3 μ l of Deep Vent (exo-) DNA polymerase. The reaction was set on a preheated thermal cycler with the following program: 95 °C, 2 min; 5 cycles of (95 °C, 20 s; 61 °C, 20 s; and 72 °C, 1 min) and 72 °C, 2 min. The product was purified with 0.85 $\times$ AMPure XP beads (Beckman) and eluted in 20 μ l of nuclease-free water.

Sequencing. Libraries were pooled and quantified following the Illumina manual and the pooled libraries were sequenced on the Illumina Nextseq 500 platform with 150 cycle sequencing kit. Customized read 2 primer (CGC CGA AGA TGG TTG AGG ATG TGT GGA GAT A) was used following the Illumina manual. The sequencing cycles were either: read 1: 110 cycles; index 1: 6 cycles; index 2: 6 cycles; and read 2: 45 cycles; or read 1: 76 cycles; index 1: 8 cycles; index 2: 8 cycles; and read 2: 45 cycles. The library information is summarized in Supplementary Tables 33 and 34.

MATQ-Drop raw data processing

Raw sequencing fastq files were generated using the standard Illumina bcl2fastq (v.2.20) software. The 3'-poly(A) tail of cDNA on read 1 was trimmed with cutadapt⁴⁷ v.3.1 paired-read mode, with the read-length filtering criteria: --minimum-length=30 --pair-filter=any. Next, a customized Python script was used to assign the read 2 cell barcode sequences to the predefined combination of barcode1 and barcode2 sequences (Supplementary Table 32) with a maximum two mismatches allowed for each segment of the barcode. Umi_tools⁴⁸ (v.1.1.1) 'extract' command was used to extract the reads with successfully assigned cell barcodes. Extracted read 1 was mapped to the hg19 genome (or a combined genome of hg19 and mm10) with STAR⁴⁹ v.2.5.3a and the uniquely mapped reads with mapping scores no smaller than 250 were used for downstream analysis. The filtered reads were assigned to genes by featureCounts⁵⁰ v.2.0.1 with Gencode annotation gtf files (human: v.19, mouse: v.M10) and the assignment was based on transcript feature (-t transcript) with strandness (-s 2). For the reads with unambiguously assigned gene features, the umi_tools 'count' command was used to generate the transcript-based digital gene expression matrix (parameter: --per-gene --gene-tag=XT --per-cell --method=directional).

To determine the cell barcodes that represent true nuclei instead of background crosstalk, we plotted out the (UMI counts) versus (barcode rank by UMI) plot, and the knee point was determined as the threshold for true nuclei (exemplified in Fig. 1b). Next, the cell barcodes representing true cells were used to generate the transcript-based gene expression matrix for true nuclei.

To generate the exon-based gene expression matrix, we first filtered out the reads with unambiguously assigned transcript-based gene features. We then reran featureCounts assignment with exon feature only (-t exon) and strandness (-s 2), followed by umi_tools count. The intron-based gene expression matrix was derived by subtracting the exon-based gene expression matrix from the transcript-based gene expression matrix.

Benchmark analysis raw data processing

Benchmark analysis was performed against 10x Chromium 3' v.3.1 platform with the following dataset: 5k Adult Mouse Brain Nuclei Isolated with Chromium Nuclei Isolation Kit.

We followed the same criteria and references in the raw data processing as MATQ-Drop. In detail, the cell barcode and UMI were extracted using the Umi_tools v.1.1.1 following the user manual and the cell number threshold was automatically determined by the umi_tools software. Next, the extracted reads were mapped to the mm10 genome with Gencode v.M10 gtf as gene model annotation reference using STAR v.2.5.3a. The uniquely mapped reads with mapping scores no smaller than 250 were used for downstream analysis. The filtered reads were assigned to genes

by featureCounts v.2.0.1 with the Gencode v.M10 gtf file and the assignment was based on transcript feature (including both exon and intron, or exon only) with strandness. For the reads with unambiguously assigned gene features, the umi_tools 'count' command was used to generate the transcript-based digital gene expression matrix (parameter: --per-gene --gene-tag=XT --per-cell --method=directional). Sensitivity comparison was performed by subsampling datasets to the same sequencing saturation, which is defined as $1 - (n_{\text{deduped_reads}} / n_{\text{reads}})$.

Clustering analysis

Data filtering. Nuclei with mitochondrial UMI percentages >5% were excluded for downstream analysis. In synaptome data, synapses with mitochondrial UMI percentages <5% were excluded for downstream analysis. Then, mitochondrial and ribosomal genes were removed from the gene expression matrix. Low-quality nuclei with <200 intronic genes were excluded and the nuclei with UMIs in the top 0.5% quantile were removed. Low-quality, Hoechst-negative subneuronal structures with <100 intronic genes were excluded and those with UMIs in the top 0.5% quantile were also removed.

Unsupervised clustering. Standard Seurat4 integration pipeline with SCTransform normalization was used for clustering analysis^{51,52}. Briefly, the intron-based (for nuclei) or the transcript-based (for synapses) gene expression matrix was normalized based on regularized negative binomial regression. Doublets were identified by the R package DoubletFinder⁵³ v.2.0 with a stringently estimated doublet rate (5%). Next, datasets of different biological samples were integrated following the Seurat scRNA-seq integration vignette. Principal component analysis and graph-based clustering were performed with the integrated data slot. Visualization of the clustering was accomplished with UMAP. Markers for each cluster were identified by the MAST⁵⁴ algorithm embedded in the Seurat package with the following parameters: only.pos=TRUE, min.pc=0.25, logfc.threshold=0.5 for nuclei or logfc.threshold=0.25 for synapses. Cell types were empirically assigned based on the overlap between cluster markers and canonical cell-type-specific markers. The above pipeline also applies to subclustering and lncRNA-based clustering analyses, except that the doublet identification and removal step were skipped because we used only the nuclei passing the 'singlet' filter described above.

Doublet removal. Doublets were identified and removed by the R package DoubletFinder⁵³ v.2.0 with a stringently estimated doublet rate (5%).

Markers for each cluster were identified by the MAST⁵⁴ algorithm embedded in the Seurat package with the following parameters: only.pos=TRUE, min.pct=0.25 and logfc.threshold=0.5. Cell types were empirically assigned based on the cluster markers and the expression of canonical cell-type-specific markers.

The same pipeline applies to subclustering and lncRNA-based clustering analyses, except that the doublet identification and removal steps were skipped because we used only the nuclei passing the 'singlet' filter described above. For lncRNA-based clustering, only the top 1,000 variable features were used for PCA.

DEG analysis

For the cluster populations of interest, a pseudobulk count matrix was assembled for each biological sample by summarizing the total UMI counts. Next, bulk DEGs were identified with edgeR⁵⁵ v.3.16. A gene is defined as 'differentially expressed' if $\text{abs}(\log_2(\text{FC})) > \log_2(1.3)$ and Benjamini-Hochberg FDR < 0.05. It is worth noting that, compared with the single-cell approach, the pseudobulk approach yields robust fold-change calculation when the two datasets show large differences in UMI detection, for example, nuclei versus synapses. The transcript-based gene expression matrix was used for DEG analysis among different subneuronal structures, whereas the exon-based

gene expression matrix was used for DEG analysis between synapses and nuclei. GO enrichment analysis of the DEGs was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID), and we used the shared expressed genes (counts per million (c.p.m.) > 2) as the background list. GSEA was performed on the $\log_2(\text{c.p.m.} + 1)$ matrix with the pseudobulk.

Identification of unspliced genes

For each type of subcellular structure, a gene is defined as ‘expressed’ if detected in at least 5% of the subcellular structures. For each neuronal type, only the expressed genes shared by presynapses and nuclei were kept for analysis. The average intron percentages of the transcripts in presynapses ($\text{pct_intron}_{\text{syn}}$) and nuclei ($\text{pct_intron}_{\text{nucleus}}$) were computed, respectively, and the splicing score (SS) at the synapse is defined as:

$$\text{SS} = \begin{cases} \min\left(\frac{\text{pct_intron}_{\text{syn}} - \text{pct_intron}_{\text{nucleus}}}{\text{pct_intron}_{\text{nucleus}}}, 0\right), & \text{if } \text{pct_intron}_{\text{nucleus}} \neq 0 \\ 1, & \text{if } \text{pct_intron}_{\text{nucleus}} = 0 \end{cases}$$

For a transcript that is fully unspliced at the synapse, $\text{SS} = 0$, whereas for a transcript that is fully spliced at the synapse, $\text{SS} = 1$. For each neuronal type, the distribution shows a peak at 1, with a long tail toward 0. Therefore, we transform the SS into z-scores and a gene is considered unspliced if splicing z-score < -2.58 (equivalent to $P < 0.01$), and $\text{pct_intron}_{\text{nucleus}} > 0.25$. The SS metrics were used in preranked GSEA.

Reporting summary

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

Data availability

The raw sequencing files are available in Gene Expression Omnibus (GEO) database under accession no. [GSE199346](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE199346).

The following public datasets were used in the present study for benchmark comparison: (1) DroNc-seq and Drop-seq on 3T3 cell line (https://singlecell.broadinstitute.org/single_cell/study/SCP128/dronc-seq-and-drop-seq-on-3t3-cell-line#study-download), (2) GEO accession no. [GSE106678](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106678) and (3) 5k Adult Mouse Brain Nuclei Isolated with Chromium Nuclei Isolation Kit (10x Genomics, <https://www.10xgenomics.com/resources/datasets/5k-adult-mouse-brain-nuclei-isolated-with-chromium-nuclei-isolation-kit-3-1-standard>). The mm10 genome can be accessed at https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.20 and the Gencode gene annotation file at https://www.gencodegenes.org/mouse/release_M10.html. The hg19 genome can be accessed at https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001405.25 and the Gencode gene annotation file at https://www.gencodegenes.org/human/release_19.html.

Code availability

The analysis code customized for MATQ_Drop sequencing data is available at https://github.com/zonglab/MATQ_Drop.

References

- Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017).
- Krishnaswami, S. R. et al. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. *Nat. Protoc.* **11**, 499–524 (2016).
- Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 3 (2011).
- Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

- Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
- Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 e3529 (2021).
- Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
- McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 e324 (2019).
- Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

Acknowledgements

We thank the McNair family for their support. We also thank NIH NeuroBioBank for providing brain tissue samples. We thank H. Dierick and M. Xue for their helpful discussion. We thank S. Eshghjoo for providing the mouse samples. We thank other Zong lab members for their help in this project. FACS experiments were performed in the Cytometry and Cell Sorting Core at Baylor College of Medicine with funding from the CPRIT Core Facility Support Award (no. CPRIT-RP180672), the NIH (grant nos. P30 CA125123, S10 RR024574 and S10 OD025251) and the assistance of J. M. Sederstrom. C.Z. is supported by a McNair Scholarship, NIH Director’s New Innovator Award (no. 1DP2EB020399) and the Behavioral Plasticity Research Institute (grant no. DBI-2021795). H.Z. is supported by the NIH (grant nos. RF1 AG020670, RF1 NS093652 and P01 AG066606).

Author contributions

C.Z. and D.A.W. designed the project. C.Z. and M.N. wrote the manuscript. M.N. and W.C. contributed to the MATQ-Drop chemistry development and hydrogel bead design and synthesis, and performed the major experiments. W.C., Y.W., Q.Z. and M.N. contributed to the microfluidics operation and microfluidics fabrication. J.L. performed the FACS and western blotting experiments. H.Z. and B.W. contributed mouse brain samples and AD mouse models. B.W. performed immunofluorescence staining experiments and related quantification. M.N. and C.Z. performed the bioinformatic analysis.

Competing interests

M.N., D.A.W. and C.Z. are cofounders and equity holders of Pioneer Genomics Inc. Baylor College of Medicine have submitted a patent application related to MATQ-Drop (US Provisional Patent Application serial no. 63/240339, 2 September 2021). The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-022-01635-1>.

Correspondence and requests for materials should be addressed to David A. Weitz or Chenghang Zong.

Peer review information *Nature Biotechnology* thanks Tibor Harkany and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Raw sequencing data were generated using the standard Illumina bcl2fastq (v2.20) pipeline. Flow cytometry data was acquired using BD FACSAriaII (P6950002) instrument.
Data analysis	The sequencing data analysis code is available online: https://github.com/zonglab/MATQ_Drop . In detail, The 3' polyA tail of cDNA on Read 1 was trimmed with cutadapt48 v3.1 paired read mode, with the read length filtering criteria: --minimum-length=30 --pair-filter=any. Next, a custom Python script was used to assign the Read 2 cell barcode sequences to the pre-defined combination of Barcode1 and Barcode2 sequences with maximal two mismatches allowed for each segment of the barcode. Umi_tools (v1.1.1) "extract" command was used to extract the reads with successfully assigned cell barcodes. Extracted Read 1 was mapped to the hg19 genome (or a combined genome of hg19 and mm10) with STAR v2.5.3a, and the uniquely mapped reads with mapping scores no smaller than 250 were used for downstream analysis. The filtered reads were assigned to genes by featureCounts v2.0.1 with Gencode annotation gtf files (human: v19, mouse: vM10), and the assignment was based on transcript feature (-t transcript) with strandness (-s 2). For the reads with unambiguously assigned gene features, the umi_tools "count" command was used to generate the transcript-based digital gene expression matrix (parameter: --per-gene --gene-tag=XT --per-cell --method=directional). Standard Seurat v4 integration pipeline with SCTransform normalization was used for clustering analysis, and the R package DoubletFinder v2.0 was used to identify potential doublets. MAST algorithm embedded in Seurat v4 package and edgeR (v3.16) was used for differential analysis. The flow cytometry data was analyzed using BD FACSDiva Version 8.0.1 software.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The sequencing data and processed files have been deposited in GEO under the accession code GEO: GSE199346.

The following public datasets were used in this study for benchmark comparison: (1) DroNc-seq and Drop-seq on 3T3 cell line (https://singlecell.broadinstitute.org/single_cell/study/SCP128/dronc-seq-and-drop-seq-on-3t3-cell-line#study-download), (2) GEO: GSE106678, and (3) 5k Adult Mouse Brain Nuclei Isolated with Chromium Nuclei Isolation Kit (10X Genomics, <https://www.10xgenomics.com/resources/datasets/5k-adult-mouse-brain-nuclei-isolated-with-chromium-nuclei-isolation-kit-3-1-standard>).

The mm10 genome can be accessed at https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.20/, and the Gencode gene annotation file can be accessed at https://www.encodegenes.org/mouse/release_M10.html. The hg19 genome can be accessed at https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_000001405.25/, and the Gencode gene annotation file can be accessed at https://www.encodegenes.org/human/release_19.html.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	The frozen postmortem human brain tissues were obtained from NIH NeuroBioBank and they are IRB exempt. The samples were dissected from one biological male and one biological female.
Population characteristics	NA. The frozen postmortem human brain tissues were obtained from NIH NeuroBioBank for synaptosome and nucleus isolation and are IRB exempt.
Recruitment	NA. The frozen postmortem human brain tissues were obtained from NIH NeuroBioBank for synaptosome and nucleus isolation and are IRB exempt.
Ethics oversight	NA. The frozen postmortem human brain tissues were obtained from NIH NeuroBioBank for synaptosome and nucleus isolation and are IRB exempt.

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

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sampling size was pre-determined for accommodate robust clustering.
Data exclusions	Nuclei with mitochondrial UMI percentages higher than 5% were excluded for downstream analysis. In synaptome data, synapses with mitochondrial UMI percentages lower than 5% were excluded for downstream analysis. Then, mitochondrial and ribosomal genes were removed from the gene expression matrix. Low-quality nuclei with fewer than 200 intronic genes were excluded, and the nuclei with UMIs in the top 0.5% quantile were also removed. Low-quality Hoechst-negative subneuron structures with fewer than 100 intronic genes were excluded, and those with UMIs in the top 0.5% quantile were also removed. Doublets were identified and removed by the R package DoubletFinder with a stringently estimated doublet rate (5%).
Replication	The clustering results and cluster marker genes are reproducible, as demonstrated by the consistency among biological and technical replicates. In the immunofluorescence staining experiments, three animals were used for each genotype, and each animal had three slides as technical replicates.
Randomization	Pseudobulk data were assembled by randomly assigning single cells into three groups. Randomization was accomplished using the "random" library in python3.
Blinding	Not applicable. No experiments need blinding for analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used

The following primary antibodies were used in western blot: synaptophysin (clone SP11, Invitrogen, MA5-14532), synapsin-I (clone D12G5, Cell Signaling Technology, 5297), CNPase (clone 11-5B, Millipore, MAB326R), GFAP (clone GA5, Millipore, MAB360), and β -actin (clone AC-15, Sigma-Aldrich, A1978). The following primary antibodies were used for immunostaining of synaptosomes: synaptophysin (clone SP11, Invitrogen, MA5-14532), and PSD95 (clone 6G6-1C9, MA1-045). The following primary antibodies were used for immunofluorescence staining on brain sections: Olig2 (Millipore, AB9610), Iba1 (Wako, 019-19741), beta-amyloid (clone 6E10, Biolegend 803001), C1q (clone 4.8, Abcam, ab182451), and PSD95 (clone 6G6-1C9, Millipore, MAB1596). The following secondary antibodies were used for immunofluorescence staining: donkey-anti-mouse Alexa Fluor 488 (Invitrogen A-21202), donkey-anti-rabbit Alexa Fluor 555 (Invitrogen A-31572), goat-anti-rabbit-Alexa Fluor 647 (Invitrogen, A21244) and goat-anti-mouse-Cy3 (Invitrogen, A10521).

Validation

The synaptophysin (Invitrogen, MA5-14532) antibody was verified by Knockout by the manufacturer to ensure that the antibody binds to the antigen stated in human SH-SY5Y cells. The C1q antibody (Abcam, ab182451) were verified by Knockout to ensure that the antibody binds to the antigen in mouse samples stated by the manufacturer. Synapsin-I antibody (Cell Signaling Technology, 5297), CNPase antibody (Millipore, MAB326R), GFAP antibody (Millipore, MAB360), β -actin antibody (Sigma-Aldrich, A1978), Olig2 antibody (Millipore, AB9610) Iba1 antibody (Wako, 019-19741), 6E10 antibody (Biolegend, 803001), and PSD95 antibody (Millipore, MAB1596) were verified by western blot using mouse and human samples with expected target size by the manufacturer. PSD95 antibody (Invitrogen, MA1-045) was verified by western blot and relative expression to ensure that the antibody binds to the antigen stated by the manufacturer.

Eukaryotic cell lines

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

Cell line source(s)

HEK293T was acquired from MD Anderson cell line core facility. 3T3 cell line was acquired from Dr. Yi Li's lab at Baylor College of Medicine.

Authentication

Authentication of two cell lines have been performed by MD Anderson and BCM cell line cores based on morphology and karyotyping.

Mycoplasma contamination

Both cell lines were tested negative for mycoplasma contamination.

Commonly misidentified lines (See [ICLAC](#) register)

No commonly misidentified cell lines were used in this study.

Animals and other research organisms

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

Laboratory animals

The C57BL/6 WT and 5xFAD mice were obtained from the Jackson Laboratory (Bar Harbor, ME). Mice were housed four per cage in a pathogen-free mouse facility with ad libitum access to food and water on a 12-hour light/dark cycle at the ambient temperature of 20.0-22.2° and the humidity of 30%-70%. Nine-month-old female mice were used for all experiments.

Wild animals

No wild animals were used in this study.

Reporting on sex

Only female mice were used for AD-related synaptome study.

Field-collected samples

No field-collected samples were used in this study.

Ethics oversight

All procedures were performed following National Institutes of Health (NIH) guidelines and approval of the Baylor College of Medicine Institutional Animal Care and Use Committee.

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

Flow Cytometry

Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

The frozen brain samples are homogenized with Dounce Homogenizer, PFA-fixed, and then stained with Hoechst 33342.

Instrument

BD FACSAriaII (P6950002)

Software

BD FACSDiva Version 8.0.1

Cell population abundance

Post-sort is performed to guarantee that the purity of sorted single nuclei or Hoechst-negative sub-neuronal structures exceeds 99.9%.

Gating strategy

UV450-A was used to identify single nucleus originated from diploid cells (therefore removing doublets generated during PFA fixation). The UV450-A low population with the FSC-A smaller than 5-um beads is considered as Hoechst-negative sub-neuronal structures.

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.