mcDETECT: Decoding 3D Spatial Synaptic Transcriptomes with Subcellular-Resolution Spatial Transcriptomics

Chenyang Yuan[1,2], Krupa Patel[1], Hongshun Shi[1,3,4], Hsiao-Lin V. Wang[1,5], Feng Wang[1], Ronghua Li[1,5], Yangping Li[1†], Victor G. Corces[1,5], Hailing Shi[1,4,5], Sulagna Das[1,6], Jindan Yu[1,3,4], Peng Jin[1,5], Bing Yao[1 *] & Jian Hu[1,2 *]

1. Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA
2. Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA 30322, USA
3. Department of Urology, Emory University School of Medicine, Atlanta, GA 30322, USA
4. Winship Cancer Institute, Emory University School of Medicine, Atlanta, GA 30322, USA
5. Emory Center for Neurodegenerative Diseases, Emory University School of Medicine, Atlanta, GA 30322, USA
6. Department of Cell Biology, Emory University School of Medicine, Atlanta, GA 30322, USA

†: Deceased

*** Correspondence:**

Bing Yao, bing.yao@emory.edu

Jian Hu, jian.hu@emory.edu

Keywords:

spatial transcriptomics, synapse, subcellular mRNA, Alzheimer's disease, machine learning

Abstract

Spatial transcriptomics (ST) has shown great potential for unraveling the molecular mechanisms of neurodegenerative diseases. However, most existing analyses of ST data focus on bulk or single-cell resolution, overlooking subcellular compartments such as synapses, which are fundamental structures of the brain's neural network. Here we present mcDETECT, a novel framework that integrates machine learning algorithms and *in situ* ST (iST) with targeted gene panels to study synapses. mcDETECT identifies individual synapses based on the aggregation of synaptic mRNAs in three-dimensional (3D) space, allowing for the construction of single-synapse spatial transcriptome profiles. By benchmarking the synapse density measured by volume electron microscopy and genetic labeling, we demonstrate that mcDETECT can faithfully and accurately recover the spatial location of single synapses using iST data from multiple platforms, including Xenium, Xenium 5K, MERSCOPE, and CosMx. Based on the subsequent transcriptome profiling, we further stratify total synapses into various subtypes and explore their pathogenic dysregulation associated with Alzheimer's disease (AD) progression, which provides potential targets for synapse-specific therapies in AD progression.

Introduction

Synapses are subcellular structures that act as the basis of neural connections in the brain[1]. The discovery of thousands of mRNAs in synapses suggests that synapses gain autonomy by producing key proteins locally to maintain their plasticity and function[2]. Mounting studies have shown that synapse loss and dysfunction are hallmarks of many neurodegenerative diseases associated with cognitive decline[3,4]. For example, in Alzheimer's disease (AD), synapses are selectively lost prior to neuronal death and therefore are considered a more sensitive and reliable indicator for early diagnosis compared to neurons[1,5,6]. However, current treatments for most neurodegenerative diseases fail to effectively slow or stop synaptic loss and only provide limited symptomatic relief[7]. As synapses become an increasingly crucial target for therapeutic intervention, it is important to gain insight into their spatial distribution in brain and transcriptomic changes during disease progression.

Recent advancements in spatial transcriptomics (ST) have enabled gene expression profiling with location information and transformed our understanding of tissue functional organization[8-13]. In particular, *in situ* hybridization-based technologies (iST), such as STARmap[14], seqFISH[15-17], and MERFISH[18,19], offer gene expression measurement at subcellular resolution. Building on these techniques, methods such as Basyor[20], JSTA[21], FICTURE[22], and Bento[23] have recently been developed to analyze gene expression at the subcellular resolution. However, these methods primarily focus on mRNAs located within the somata, which represent only 60–70% of the total mRNA molecular content[20,24]. The remaining 30–40% of mRNAs located outside the somata, including those from extranuclear compartments like synapses, are often overlooked. These mRNAs play fundamental roles in neuronal functions, such as synaptic connections and communications. To fully harness the potential of iST for studying synapses, two major challenges must be addressed: (1) iST pipelines typically construct expression profile at single-

cell resolution by identifying cellular boundaries through nuclei and membrane staining. Consequently, synapses, which are not visible in these staining images, are not directly detectable. (2) iST experiments are limited to profiling genes from a predetermined panel, typically covering 300–500 genes, which often lacks adequate coverage of genes relevant to synaptic function. These technical bottlenecks in iST hinder the mapping of gene enrichment in individual synapses.

To overcome these aforementioned challenges, we developed mcDETECT (**m**ulti-**c**hannel **DE**nsity-based spatial **T**ranscriptom**E** **C**lus**T**ering), a machine learning framework designed to characterize the transcriptome profiles of individual synapses in a spatial context using iST data. mcDETECT is grounded in the fact that many synaptic mRNAs are localized at synapses for local protein synthesis. These local mRNAs form distinctive extranuclear aggregations at subcellular scale, demonstrated by both iST and fluorescent data generated from individually cultured neurons[25,26]. mcDETECT first employs density-based clustering to pinpoint the location of these aggregations as individual synapses in three-dimensional (3D) space. Subsequently, it assigns surrounding mRNA molecules to each synapse to reconstruct their transcriptome profiles. We have applied mcDETECT to both simulated data and real mouse brain data generated from three state-of-the-art iST platforms, and the synapse density identified by our method correlates well with the ground truth derived from volume electron microscopy and genetic labeling[27]. Moreover, mcDETECT can differentiate between various synapse subtypes based on their unique and signature transcriptome profiles, which exhibit distinct spatial distributions in the brain. In a case study, mcDETECT revealed variations in both synapse number and gene landscape between AD mouse models at two disease progression stages. To our knowledge, mcDETECT is the first method to enable spatial synapse transcriptome profiling at subcellular resolution. We believe mcDETECT offers a significant tool for the brain research

community, providing molecular insights and identifying potential targets for developing synapse-targeted therapies for neurodegenerative diseases.

Results

Overview of mcDETECT

The workflow of mcDETECT is illustrated in **Fig. 1**. It starts by examining the subcellular distribution of synaptic marker mRNAs in an iST sample. Each mRNA molecule is treated as a distinct point with its own 3D spatial coordinates considering the thickness of the sample. Unlike cell-type specific marker genes, which are typically found within nuclei, mRNAs of synaptic markers often form small aggregations outside the nuclei (**Fig. 1a**). mcDETECT uses a density-based clustering[28] approach detailed in **Fig. 1b** to identify these extranuclear aggregations. This involves calculating the Euclidean distance between mRNA points and defining the neighborhood of each point within a specified search radius. Points are then categorized into core points, border points, and noise points based on their reachability from neighboring points. mcDETECT recognizes each bundle of core and border points as a synaptic aggregation. To minimize false positives, it excludes aggregations that significantly overlap with nuclei identified by DAPI staining. Subsequently, mcDETECT repeats this process for multiple synaptic markers, merging aggregations from different markers with high overlaps (**Fig. 1c**). After encompassing all markers, an additional filtering step is performed to remove aggregations that contain mRNAs from negative control genes, which are known to be enriched only in nuclei. The remaining aggregations are considered individual synaptic aggregations. As shown in **Fig. 1d**, mcDETECT then uses the minimum enclosing sphere of each aggregation to gather all mRNA molecules and summarizes their counts for all measured genes to define the spatial transcriptome profile of individual synapses. mcDETECT is designed to study the transcriptomes of actively functioning synapses. Inactive synapses, which lack mRNA activity,

are beyond its detection range. In addition to synapses, mcDETECT also has the potential to identify other subcellular structures which contain localized mRNAs, such as dendrites, axons, and other subcellular components in tissues other than brain.

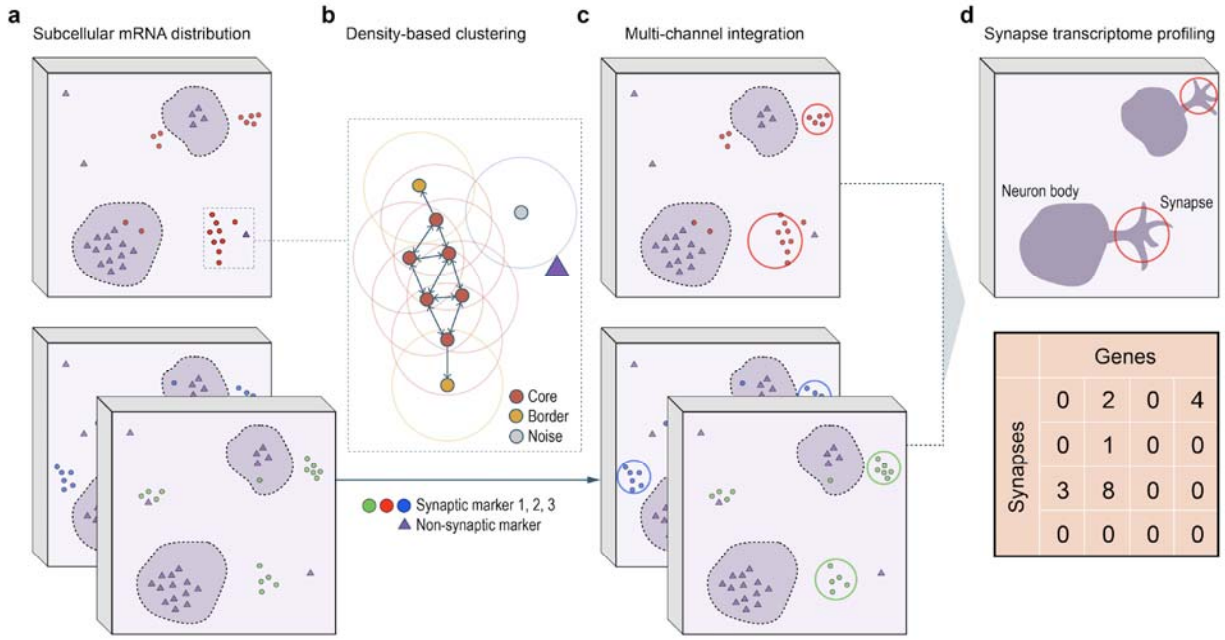

Fig. 1 | Workflow of mcDETECT. a. Schematic representation of subcellular spatial distribution of synaptic mRNAs (red, blue, and green dots) and non-synaptic mRNAs (purple triangles) within neurons. **b.** mcDETECT employs a density-based clustering approach to identify extranuclear mRNA aggregations for each synaptic marker. **c.** mcDETECT repeats this process for all synaptic markers, merges aggregations with significant spatial overlap, and pinpoints individual synapse locations. **d.** mcDETECT records all mRNAs within the minimal enclosing sphere of each aggregation to construct the transcriptome profile of individual synapses.

Selection of synaptic markers

mcDETECT requires a list of synaptic markers as input for accurate synapse detection. Many previous studies have reported on synaptic markers[2,26,29]. For example, Niu *et al.*[30] developed a droplet-based sequencing platform called Multiple-Annealing-and-Tailing-based Quantitative

Single-cell RNA-seq in Droplets (MATQ-Drop) to capture the transcriptome profiles of individual synapses, although their spatial information was not preserved. Other studies using *in situ* hybridization-based sequencing techniques[26], including ribosome sequencing (Ribo-seq)[2] and ribosome-bound mRNA mapping (RIBOmap)[31], have also identified certain mRNAs enriched or translated in synapses. Based on the genes reported in these studies, we selected a list of markers directly related to synapse structure and function, as detailed in **Supplementary Tables 2** and **3**. The dot plot in **Fig. 2a** compares the selected markers' expression in synapses and neuron nuclei based on the MATQ-Drop data. Considering the much smaller size of synapses compared to neuron nuclei, the patterns indicate that these genes are more enriched in synapses than in nuclei. To ensure that the gene expression patterns observed in droplet-based techniques can be accurately transformed to data generated from iST platforms, we next tested the correlation of gene expression between these two techniques. Existing iST experiments typically profile 300–500 genes from a predefined panel rather than covering the entire transcriptome. We selected the shared genes between the MATQ-Drop data and five iST datasets[32-36] from different platforms (**Supplementary Table 1**) and calculated the correlation of their expression pattern across five major cell types, including Glutamatergic neurons, GABAergic neurons, astrocytes, oligodendrocytes, and microglia. As illustrated in the box plots in **Fig. 2b**, we observed a high correlation between MATQ-Drop data and all iST datasets, confirming that mRNA measurements from these two categories of techniques are concordant. Therefore, the synaptic maker detected by MATQ-Drop can be transferred to iST data.

The iST data generated from individually cultured neurons in vitro demonstrates that mRNA molecules of synaptic markers aggregate in synapses[24]. To further confirm whether this subcellular distribution pattern exists in vivo for the selected synaptic markers, we analyzed two mouse brain iST datasets from the Xenium and CosMx platforms. As shown in the examples in **Fig. 2c**, two synaptic markers, *Dlg4* and *Syp*, consistently exhibit a clear extranuclear

aggregation pattern across the two datasets. In contrast, two neuronal markers, *Camk2b* and *Gad1*, are primarily enriched within the nuclei. This observation serves as a foundation for mcDETECT to accurately recover the spatial location of individual synapses from subcellular mRNA distribution.

Simulation study

To evaluate the effectiveness of mcDETECT in identifying synapses, we conducted a simulation study to benchmark its performance. We set up a $2000 \times 2000\ \mu m^2$ tissue area with a thickness $10.5\ \mu m$, which mimicked the Xenium sample setting. We considered three genes, denoted as A, B, and C, with varying mRNA densities based on the observed expression of the synaptic markers *Snap25*, *Slc17a7*, and *Dlg4* in a real Xenium mouse brain dataset. Based on iST data from individually cultured neurons, we simulated three patterns of the subcellular mRNA distribution for each marker, as depicted in **Fig. 2d**: (1) Complete spatial randomness (CSR), where mRNA points are randomly distributed without forming aggregations. (2) Aggregation within synapses, by simulating mRNA clusters with a mean radius of $1\ \mu m$, reflecting the reported synapse size in existing studies[37-39]. We randomly set 80% of the points in these clusters to be located outside the nuclei, with the remaining 20% distributed within the nuclei. (3) Aggregation within nuclei, by simulating mRNA clusters with a mean radius of $3.5\ \mu m$, and 80% of the points are located within the nuclei. For each gene, we conducted 200 simulation runs and applied mcDETECT to detect the synaptic aggregations. As illustrated in the bar plots in **Fig. 2e**, mcDETECT achieved overall accuracies of 97.7%, 96.4%, and 93.0% for markers A, B, and C, respectively. Notably, mcDETECT is not only sensitive to true synaptic aggregations—reflected by mean recalls of 94.1%, 93.8%, and 90.4%—but also effectively filtered out most intranuclear aggregations, as indicated by mean precisions of 97.7%, 96.5%, and 93.0% for markers A, B, and C, respectively.

The above simulation focused on synapse detection using a single marker. In a more realistic scenario with multiple synaptic markers, mcDETECT enhances statistical power by integrating aggregations from multiple markers while accounting for their colocalization patterns in synapses. To demonstrate this capability, we simulated a scenario with three synaptic markers, A, B, and C, whose mRNA molecules were visualized in red, green, and orange, respectively, and partially co-clustered within the same 3D space, as shown in **Fig. 2f**. Applying mcDETECT to this dataset across 200 simulation runs resulted in a mean accuracy of 96.1%, a mean precision of 96.0%, and a mean recall of 94.6%, as detailed in the boxplot in **Fig. 2g**. Notably, using only a single marker in this scenario recovered only a subset of the total synapses, resulting in a decreased recall of approximately 60% when using marker A, B, or C alone. This highlights that incorporating multiple synaptic markers enables mcDETECT to more effectively recover true positive synapses.

In current iST data analysis, since the x-y plane usually has a much larger scale ($10 \times 10\ mm^2$) than the z-axis ($10\ \mu m$), most existing analyses treat tissues as two-dimensional (2D) sections and ignore their thickness. By conducting the above simulation in a 2D setting, we also demonstrate that omitting the z-axis can result in inaccurate synapse detection. Compared to 2D, mcDETECT is more effective in 3D at accurately recovering true synapse locations and excluding nuclear aggregations and random debris (**Supplementary Note 1**).

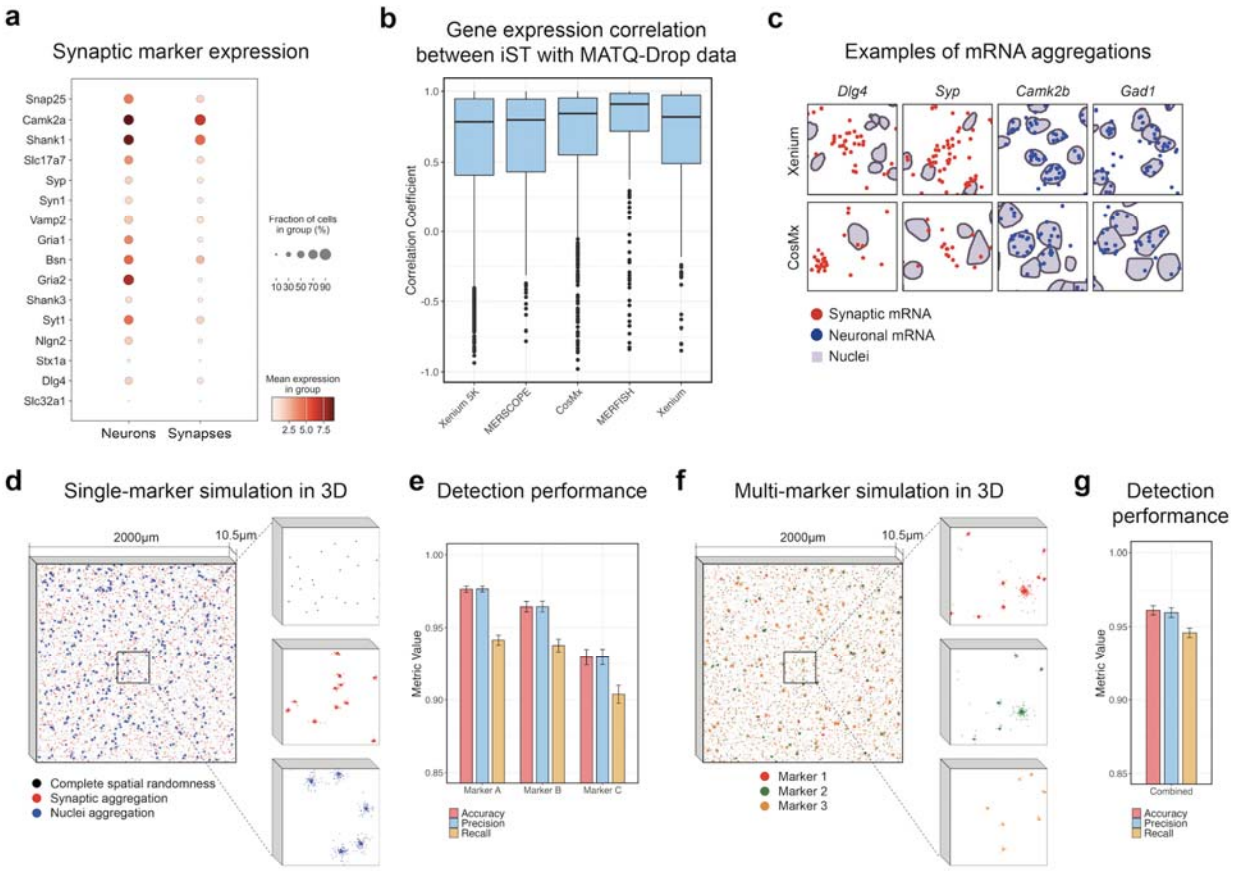

Fig. 2 | Synaptic marker selection and simulation study. a. Dot plot showing the expression levels of selected synaptic markers from MATQ-Drop data in neuronal nuclei and synapses. **b.** Box plot showing the gene expression correlation across five major cell types between the MATQ-Drop dataset and five iST datasets. The lower and upper hinges correspond to the first and third quartiles, and the center refers to the median value. The upper (lower) whiskers extend from the hinge to the largest (smallest) value no further (at most) than 1.5 × interquartile range from the hinge. Data beyond the end of the whiskers are plotted individually. **c.** Examples of subcellular mRNA aggregation patterns for synaptic markers (*Dlg4*, *Syp*) and neuronal markers (*Camk2b*, *Gad1*) in Xenium and CosMx datasets. Synaptic and neuronal mRNAs are shown as red and blue dots, respectively. Neuronal nuclei are identified from DAPI images and are highlighted in purple. **d.** Left: Simulation of a single synaptic marker in 3D space, where each dot represents a subcellular mRNA molecule. Black, red, and blue dots indicate complete

spatial randomness (CSR), synaptic aggregation, and nuclear aggregation, respectively. Right: Zoomed-in view of the three distribution patterns in a small region. **e.** Bar plot showing the detection accuracy, precision, and recall achieved by mcDETECT on the simulated 3D single-marker data across 200 simulation runs. Bar heights represent mean values, and error bars denote mean $\pm$ one standard deviation. **f.** Left: Simulation of three synaptic markers in 3D space, where each dot represents a subcellular mRNA molecule. Red, green, and orange dots indicate marker types A, B, and C, respectively. Right: Zoomed-in view of a small region stratified by marker type. **g.** Bar plot showing the detection accuracy, precision, and recall achieved by mcDETECT on the simulated 3D multi-marker data across 200 simulation runs. Bar heights represent mean values, and error bars denote mean $\pm$ one standard deviation.

Application to mouse brain 10x Xenium 5K data

To demonstrate the capabilities of mcDETECT in real data, we applied it to a nine-week-old wild-type (WT) mouse brain dataset generated from the 10x Xenium 5K platform[32]. This dataset has a broad gene coverage of 5,006 genes, including many of our selected synaptic markers. We began with cell type classification and spatial domain detection on the coronal slice, identifying 31,388 neurons from nine major brain regions (**Fig. 3a**). Using 12 synaptic markers available in this dataset, including *Snap25*, *Camk2a*, and *Slc17a7*, mcDETECT identified 123,098 synapses. The synapse distribution in **Fig. 3b** shows that these synapses are more concentrated in the isocortex layers, hippocampus (HPF), and olfactory areas (OLF) compared to other regions. Few synapses are observed in the fiber tracts (FT), which is consistent with the absence of neurons in that region[40].

Directly evaluating the accuracy of synapse detection can be challenging given the technical bottlenecks in obtaining ground truth on the same iST slide. Unlike cells, synapses are difficult to identify using synaptic protein staining due to issues with antibody penetration[41]. For instance,

we found that synapses remained unidentifiable when using stains like SAP102, a well-known postsynaptic protein as shown in **Supplementary Fig. 1**. Therefore, we assessed the accuracy of the synapses identified by mcDETECT in the following three aspects. First, we examined the sizes of the identified synapses. As illustrated in **Fig. 3c**, the radii of 90% of the identified synapses fall within the range of $0.56–1.67\ \mu m$ with a median of $0.94\ \mu m$, which is consistent with synapse sizes reported in previous studies[37-39]. Second, we compared the transcriptome profiles of the identified synapses with those of neurons. **Fig. 3d** presents a heatmap of 12 synaptic markers along with 10 selected neuronal markers[42-45] as negative controls. In the identified synapses, synaptic markers are highly enriched, whereas negative control mRNAs are rather sparsely located. Conversely, neurons show high enrichment of both types of markers. This distinct pattern of gene enrichment highlights that the identified synapses are not fragments of neuronal cell structures or random debris. Third, prior studies have measured synapse density in mouse brains using a combination of volume electron microscopy and genetic labeling[27]. These measurements, representing brain region-specific synapse densities in WT mice at postnatal day 56, are shown in **Fig. 3e (left)** and serve as the ground truth for evaluation. We first calculated the synapse densities obtained by mcDETECT within each $50 \times 50\ \mu m^2$ square spots, as shown in **Fig. 3e (right)**. The results recapitulate the same pattern as the ground truth. Next, we assessed the mean synapse density across nine major brain regions, as summarized in the line plot in **Fig. 3f**. Both figures highlight synapse density enrichment in the isocortex, HPF, and OLF, with a peak in the in the cornu ammonis (CA) area of the HPF and lower synapse densities in other regions. These findings are consistent with ground truth data, further supported by a Spearman correlation of 0.93, weighted by the area of each region.

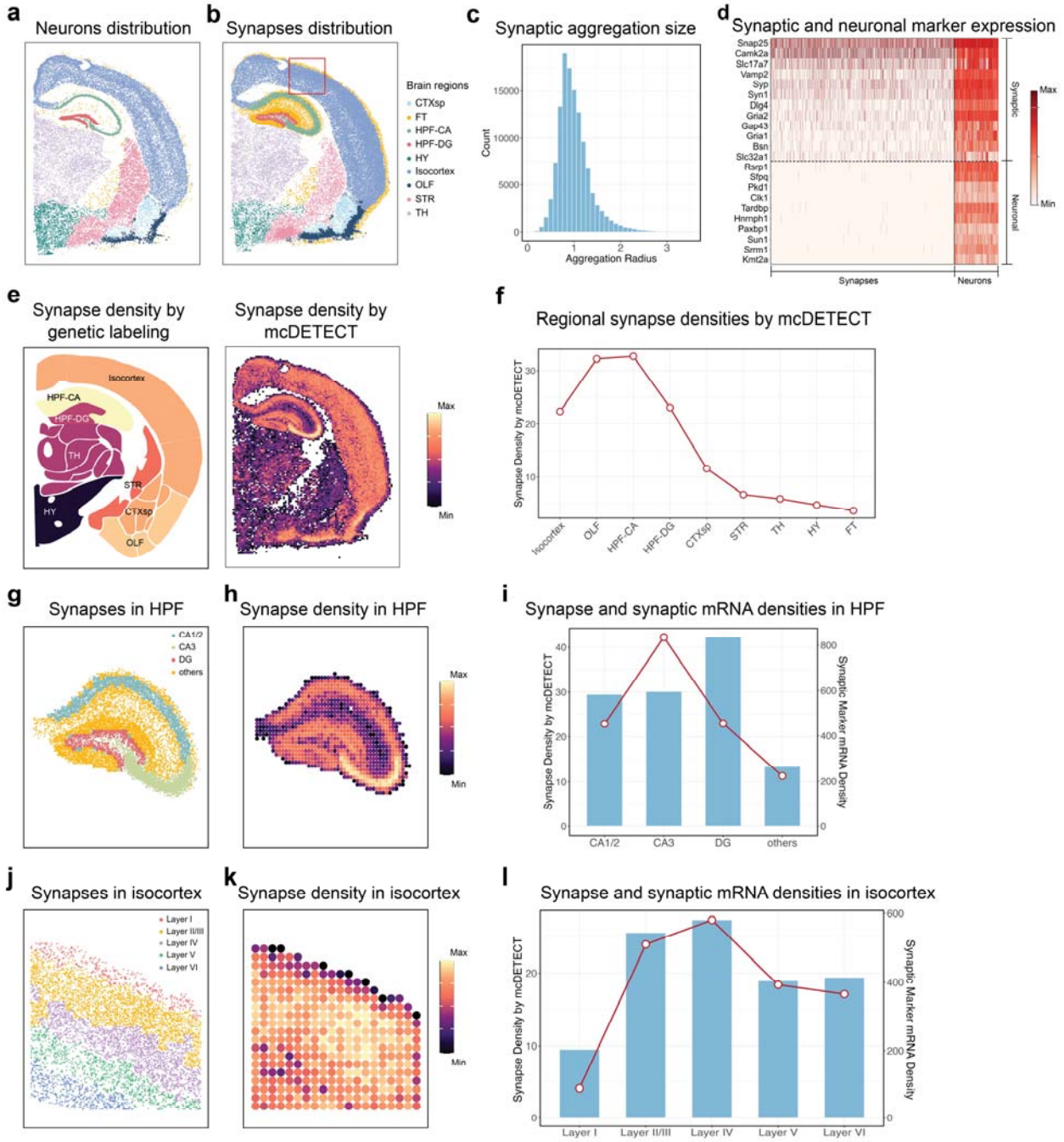

Fig. 3 | Synapse detection in the Xenium 5K dataset. a. Scatter plot of all neurons across nine major brain regions: cortical subplate (CTXsp), fiber tracts (FT), cornu ammonis of the hippocampus (HPF-CA), dendrite gyrus of the hippocampus (HPF-DG), hypothalamus (HY), isocortex, olfactory areas (OLF), striatum (STR), and thalamus (TH). **b.** Scatter plot of synapses identified by mcDETECT across the same nine major brain regions. **c.** Histogram showing the

distribution of radii for all identified synaptic aggregations. **d.** Heatmap showing the enrichment patterns of the 12 synaptic markers used for synapse detection and 10 neuronal markers in the identified synapses and neurons. **e.** Left: regional synapse densities estimated using volume electron microscopy and genetic labeling on a coronal slice corresponding to the approximate bregma position in **a**. Right: scatter plot showing the number of synapses identified by mcDETECT in $50 \times 50\ \mu m^2$ square spots spanning the tissue slice. **f.** Line plot displaying the regional synapse density trend inferred by mcDETECT. **g.** Scatter plot of synapses identified by mcDETECT across four HPF subregions. **h.** Scatter plot showing synapse counts identified by mcDETECT in $50 \times 50\ \mu m^2$ square spots spanning the HPF area. **i.** Line plot displaying synapse densities identified by mcDETECT across HPF subregions (left y-axis). Blue bars represent the number of synaptic mRNAs per spot in each subregion (right y-axis). **j.** Scatter plot of synapses identified by mcDETECT across five cortical layers within the selected isocortex region (red box in **b**). **k.** Scatter plot showing synapse counts identified by mcDETECT in $50 \times 50\ \mu m^2$ square spots spanning the selected isocortex region. **l.** Line plot displaying synapse densities identified by mcDETECT across isocortex layers (left y-axis). Blue bars represent the number of synaptic mRNAs per spot in each layer (right y-axis).

We next demonstrate that mcDETECT can accurately measure synapse density at a higher spatial resolution by focusing on the HPF region, which contains one of the highest concentrations of identified synapses. We subdivided the synapses in the HPF into four subregions, namely CA1/2, CA3, dentate gyrus (DG), and others, as depicted in **Fig. 3g**, and calculated their densities in each $50 \times 50\ \mu m^2$ square spot, as shown in **Fig. 3h**. Notably, the CA region exhibits a higher synapse density compared to the DG, recapitulating the trend in ground truth. However, we found that the synaptic mRNA density is higher in the DG compared to the CA, as indicated by the blue bars in **Fig. 3i**. This discrepancy can be attributed to the

developmental differences between neurons in the CA and DG regions. In DG, there is an immature neuronal population that arises from progenitor cells located in the subgranular zone[46]. These immature neurons lack fully formed synapses, resulting in synaptic markers being predominantly localized within nuclei. In contrast, neurons in the CA region are mostly mature, with synaptic markers localized within well-developed synapses[47], enabling mcDETECT to accurately recognize these extranuclear aggregations as true synapses. To further support this conclusion, we performed Differential Expression (DE) and Gene Ontology (GO) analysis to compare the neurons in the CA and DG regions. As summarized in **Supplementary Table 5** and **Supplementary Fig. 2**, genes and pathways enriched in the CA are primarily associated with synapse function and plasticity, indicating their maturity and active role in forming connections. In contrast, genes and pathways upregulated in the DG are linked to neuronal development, reflecting the immature neuron population in this region. This example emphasizes that the raw number of synaptic mRNAs is not a reliable indicator of synapse enrichment. Instead, mcDETECT leverages the subcellular distribution pattern of these mRNAs to recover the precise locations of individual synapses. Moreover, we examined the subregions within the CA and found that synapses were more densely concentrated in CA3 compared to CA1 and CA2. The higher synapse density in CA3 underscores its pivotal role in complex memory processing tasks, such as pattern separation and completion[48,49], while CA1 and CA2 primarily function as relay stations within the hippocampal circuit[50].

We next zoomed into a selected isocortex region, highlighted by the red box in **Fig. 3b**, and categorized the synapses according to their respective layers, as shown in **Fig. 3j**. Both the spot-level (**Fig. 3k**) and layer-level synapse density (**Fig. 3l**) exhibit an increasing trend from the external layer (Layer I) to the internal layers (Layer II/III), peaking in Layer IV before decreasing towards Layers V and VI. The dense synaptic architecture in Layer IV enables robust and

efficient processing of sensory information, distinguishing it from other layers–a distinction well-supported by existing literature[51].

After recovering synapse locations, mcDETECT enables the construction of transcriptome profiles for individual synapses, facilitating the study of their functions and subtype classification. Among the 123,098 identified synapses, we observed the presence of all 5,006 genes, with each synapse having a median mRNA count of 20. Synapses can be broadly categorized into two subtypes: the pre-synapse, which releases neurotransmitters into the synaptic cleft to communicate with the next cell, and the post-synapse, which serves as the receiving end, equipped with specialized receptors to detect and respond to the neurotransmitters. Using 20 selected markers for pre- and post-synapses in **Supplementary Table 3**, we performed K-Means clustering and identified 15 synapse clusters, as visualized in the t-SNE plot in **Fig. 4a**. Based on enrichment of these markers in **Fig. 4b**, we assigned clusters 1, 2, 3, 4, 5, 7, 9, 10, 12, 13 as pre-synapses and clusters 0, 8 as post-synapses. Next, we examined the spatial distribution of pre- and post-synapses. As shown in the pre- and post-synapse distribution plot in **Fig. 4c**, the thalamus (TH) has a higher concentration of pre-synapses while the HPF has denser post-synapses. The imbalance between pre- and post-synapses is an indicator of the directionality of signal transmission between brain regions. This observed pattern suggests that the TH primarily acts as a signal-exporting region[52,53], whereas the HPF predominantly function as a signal-receiving and processing region[54,55]. To validate this finding, we compared our results with the Allen Mouse Brain Connectivity Atlas (AMBCA)[56], which maps axonal projection strengths by tracking the trajectories of enhanced green fluorescent protein (EGFP)-expressing adeno-associated viral vectors. **Fig. 4d (left)** shows the projection strengths from source regions (y-axis) to target regions (x-axis). By summarizing the inward and outward projection for each region, we calculated a score to represent the relative signal export-to-import ratio, as shown in **Fig. 4d (right)**. The TH exhibits more exporting than importing projections and

primarily functions as an outputting unit, resulting in the formation of numerous pre-synaptic boutons. In contrast, the HPF primarily acts as an input-receiving unit and is enriched with dendritic spines, accounting for its high concentration of post-synapses.

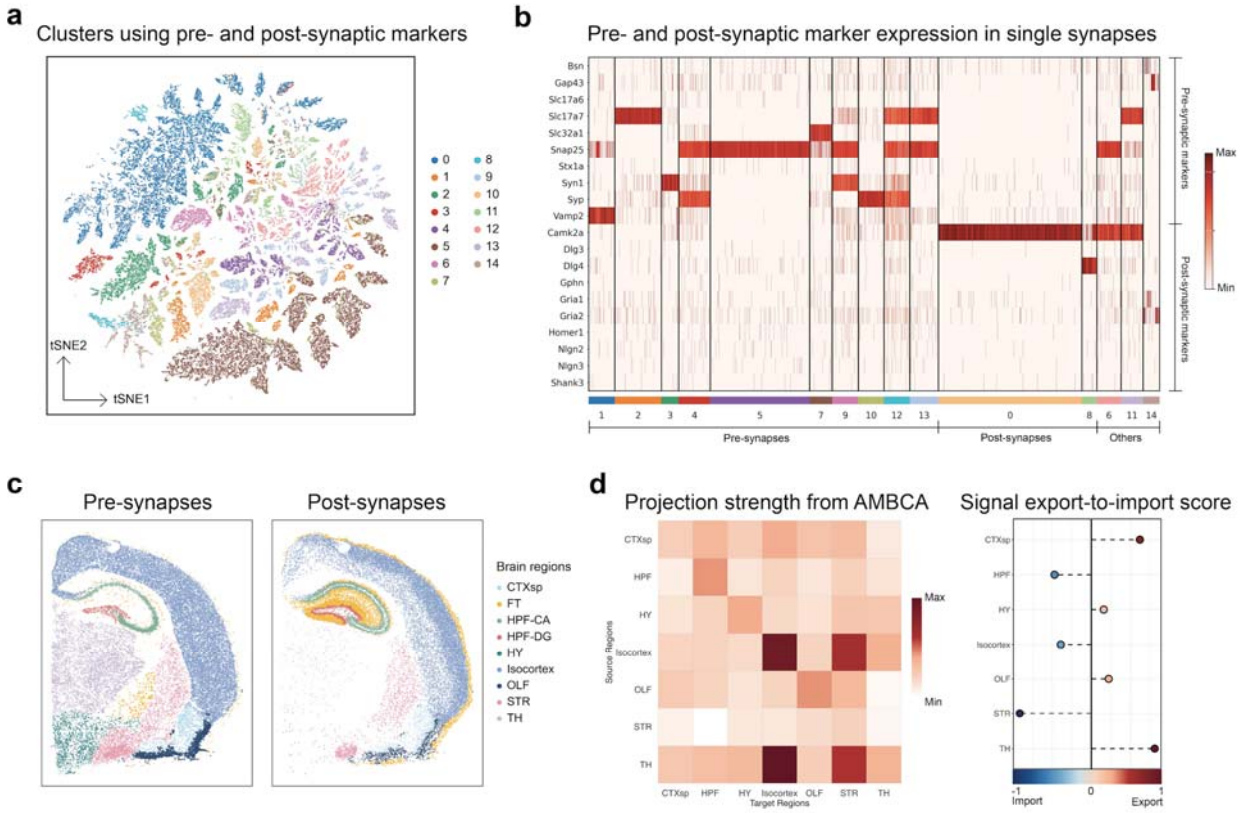

Fig. 4 | Synapse subtyping in the Xenium 5K dataset. a. t-SNE plot of all synapses using all genes, colored by 15 clusters from K-Means clustering based on 20 pre- and post-synaptic markers. **b.** Heatmap showing the enrichment patterns of the 20 pre- and post-synaptic markers across the 15 synapse clusters. **c.** Scatter plots showing the spatial distribution of identified pre-synapses (left) and post-synapses (right) across brain regions. **d.** Left: heatmap showing the pairwise axonal projection strength between source regions (y-axis) and target regions (x-axis) in the Allen Mouse Brain Connectivity Atlas. Right: dot plot summarizing the signal export-to-import score for each brain region, coloring them based on predominantly receiving signals (blue) or sending signals (red).

Application to in-house mouse brain MERSCOPE data with a specialized gene panel

The results from the Xenium 5K dataset demonstrate the capability of mcDETECT for identifying and profiling individual synapses. However, the investigation of synaptic transcriptome profiles and subtypes is limited by the dataset's limited coverage of synaptic genes. To overcome this limitation, we designed a specialized gene panel tailored for studying synapses detailed in **Supplementary Table 6**. This panel includes 292 genes, comprising 24 general synaptic markers and 15 subtype-specific markers reported by Niu *et al.* The remaining genes are included to mark cell types and brain regions, ensuring compatibility with standard analyses. Using this customized gene panel, we conducted a MERSCOPE experiment on a coronal slice from a WT mouse aged 6–8 weeks. **Fig. 5a** displays the neurons in different brain regions. Compared to the Xenium 5K sample, this slice is closer to the caudal end of the mouse brain and covers more of the midbrain, allowing for a holistic evaluation of our method. We applied mcDETECT to this dataset and identified 1,018,249 synapses across eight major brain regions, as presented in **Fig. 5b**. The synapse count is significantly elevated compared to the Xenium 5K dataset, due to differences in tissue area and platform variations in mRNA capture efficiency[57] (**Supplementary Note 2**). We then validated the identified synapses based on three aspects: size, transcriptome profile, and regional density. As shown in **Fig. 5c**, the radii of 90% of the aggregations fall within the range of 0.51–$1.84\ \mu m$ with a median of $0.94\ \mu m$, consistent with the reported size of mouse synapses. The heatmap in **Fig. 5d** shows that synaptic markers are highly expressed in the identified synapses compared to neuronal markers, highlighting the precision of synapse identification. Next, we compared the ground truth synapse density obtained through the genetic labeling approach in the same brain region (**Fig. 5e**, left) with the results from mcDETECT (**Fig. 5e**, right), which successfully replicated the same pattern. As summarized in the line plot in **Fig. 5f**, mcDETECT accurately captures synapse density peaks in the isocortex, HPF, and OLF, as well as the higher synapse density in the CA compared to the DG. The overall regional density trend also aligns with the ground truth, achieving a weighted

Spearman correlation of 0.99. Together, these findings underscore the reliability and accuracy of the synapses identified in this dataset.

Next, we investigated the synapse subtypes based on their transcriptome profiles. Among the identified synapses, we observed the presence of all 290 genes, with each synapse having a median mRNA count of 32. Similar to the Xenium 5K dataset, we separated the synapses into 15 clusters based on 24 pre- and post-synaptic markers, as visualized in the t-SNE plot in **Fig. 5g**. Based on the heatmap shown in **Fig. 5h**, we assigned clusters 2, 3, 5, 7, 8, 9, 10, 11, 12, and 14 as pre-synapses, and clusters 0, 1, 6, and 13 as post-synapses. The spatial distributions of these two subtypes are displayed in **Fig. 5i.** We observed a similar trend to the Xenium 5K dataset, where pre-synapses are more enriched in the TH but are less concentrated in the HPF compared to post-synapses. This imbalance reflects the directional flow of signals within these brain regions, with the TH primarily functioning as signal-sending hubs, while the HPF serves as a signal-receiving and processing center. This finding is further supported by the connectivity injection experiment detailed in **Fig. 4d**.

The inclusion of more synapse subtype markers in this customized panel enables the classification of synapses into additional subtypes beyond the established pre- and post-synaptic categories. Synapse subtyping remains a challenging and relatively unexplored field. Although some existing studies have identified synapse subtypes based on their morphology or associated neuron types, the understanding of synapse subtypes at the transcriptomic level is still limited. In Niu *et al.*, 12 synapse subtypes, named Syn1–12, were identified based on gene enrichment using MATQ-Drop, although their spatial distribution within the brain remains unknown. Using the subtype markers reported in their study, we conducted a clustering analysis and identified eight distinct clusters, as visualized in the t-SNE plot in **Fig. 5j**. By analyzing the enrichment of these markers, we uncovered correspondences between the synapse subtypes

identified by Niu *et al*. and those identified by mcDETECT, as shown in the heatmap in **Fig. 5k**. For example, subtype Syn12, reported in Niu *et al.*, corresponds to cluster 6 in mcDETECT, both characterized by the upregulation of *Ksr2*, a gene known to contribute to calcium-mediated extracellular signal-regulated kinase (ERK) signaling[58]. This subtype exhibits a relatively uniform distribution across the brain (**Fig. 5l**, top). Subtype Syn9 in Niu *et al.* aligns with cluster 7 in mcDETECT, with both showing upregulation of *Apc*, a key coordinator of presynaptic and postsynaptic maturation[59]. This subtype is more concentrated in the isocortex and HPF (**Fig. 5l**, bottom). These results together demonstrate that mcDETECT is a powerful and promising tool for studying synapse subtypes, their gene enrichment pattern, and spatial distributions.

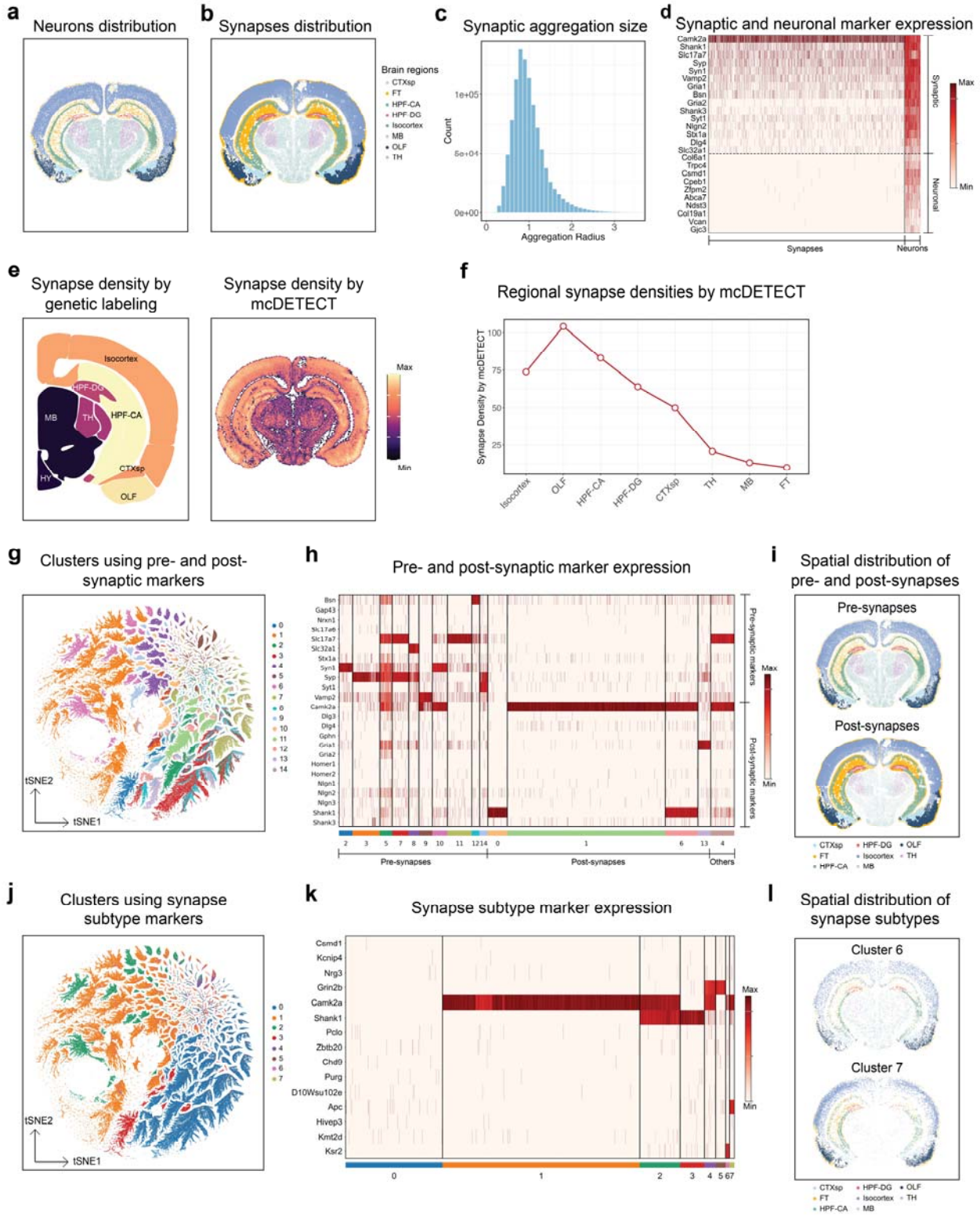

Fig. 5 | Synapse detection and subtyping in the in-house MERSCOPE dataset using a synaptic-focused gene panel. a. Scatter plot of all neurons across eight major brain regions:

cortical subplate (CTXsp), fiber tracts (FT), cornu ammonis of the hippocampus (HPF-CA), dendrite gyrus of the hippocampus (HPF-DG), isocortex, midbrain (MB), olfactory areas (OLF), and thalamus (TH). **b.** Scatter plot of synapses identified by mcDETECT across the same eight major brain regions. **c.** Histogram showing the distribution of radii for all identified synaptic aggregations. **d.** Heatmap showing the enrichment patterns of the 15 synaptic markers used for synapse detection and 10 neuronal markers in the identified synapses and neurons. **e.** Left: regional synapse densities estimated using volume electron microscopy and genetic labeling on a coronal slice corresponding to the approximate bregma position in **a**. Right: scatter plot showing the number of synapses identified by mcDETECT in $50 \times 50\,\mu m^2$ square spots spanning the tissue slice. **f.** Line plot displaying the regional synapse density trend inferred by mcDETECT. **g.** t-SNE plot of all synapses using all genes, colored by 15 clusters from K-Means clustering based on 24 pre- and post-synaptic markers. **h.** Heatmap showing the enrichment patterns of the 24 pre- and post-synaptic markers across the 15 synapse clusters. **i.** Scatter plots showing the spatial distribution of identified pre-synapses (top) and post-synapses (bottom) across brain regions. **j.** t-SNE plot of all synapses using all genes, colored by eight clusters from K-Means clustering based on 15 synapse subtype markers. **k.** Heatmap showing the enrichment patterns of the 15 synapse subtype markers across the eight synapse clusters. **l.** Scatter plots showing the spatial distribution of synapse cluster 6 (top) and cluster 7 (bottom) across brain regions.

Application to CosMx and MERFISH data with limited synaptic marker coverage

We have demonstrated that mcDETECT can accurately identify individual synapses in the Xenium 5K and MERSCOPE datasets, both of which provide substantial coverage of synaptic markers, either through a broad gene coverage or a specialized panel. However, most existing iST datasets contain only 300–500 genes and are not specifically designed to study synapses. To showcase the scalability of mcDETECT on such datasets with limited synaptic marker

coverage, we demonstrate that effective synapse detection does not require a large number of markers. Instead, an essential panel of synaptic genes would be sufficient to recover most synapses accurately. To validate this, we conducted benchmarking using the Xenium 5K dataset. We began with synapse detection for each gene individually and then combined the results, prioritizing by gene expression intensity and merging overlapping synapse detections. As illustrated in the line plot in **Fig. 6a**, starting with *Snap25*, the cumulative number of detected synapses increases from 53,000 to 141,000. After incorporating the top 10 genes, however, adding additional markers will not significantly elevate the number of detected synapses. Furthermore, benchmarking analyses reveal that even when strong markers like *Snap25* or *Slc17a7* are excluded, most synapses can still be recovered, with the resulting synapse density trends maintaining a high correlation with ground truth densities (**Fig. 6b**). This flexibility allows mcDETECT to be applied to existing iST datasets with limited synaptic markers or even with certain key markers missing, which enables a reanalysis of such datasets, offering deeper insights into brain organization and disease progression. Building on this foundation, we extended our analysis to two publicly available datasets generated using the CosMx[33] and MERFISH[34] platforms, both characterized by limited synaptic gene coverage. The results demonstrate that mcDETECT can accurately identify synapses in both datasets (**Supplementary Notes 3** and **4**).

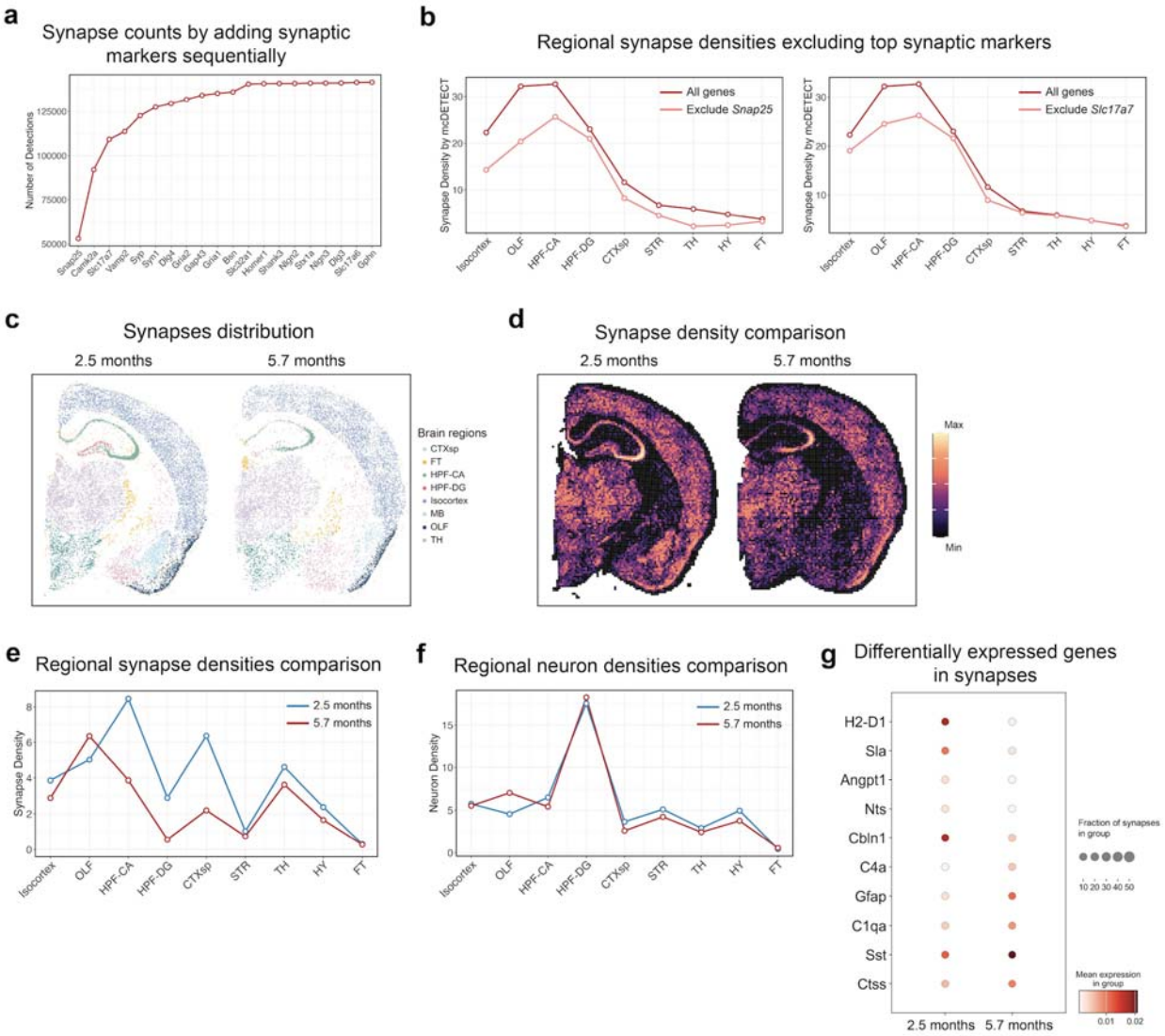

Fig. 6 | Benchmarking of mcDETECT on datasets with limited gene coverage and its application to AD mouse models across different disease stages. a. Line plot showing the identified synapse count in the Xenium 5K dataset by sequentially applying mcDETECT to all 20 synaptic markers and aggregating the results. **b.** Line plots comparing the regional synapse density trends using different synaptic marker lists in the Xenium 5K dataset. Left: comparison of synapse density trends when *Snap25* is excluded from the original gene list. Right: comparison when *Slc17a7* is excluded. **c.** Scatter plot of synapses identified by mcDETECT in the 2.5-month (left) and 5.7-month (right) TgCRND8 samples. The nine major brain regions identified on the associate coronal slice are the same as those in **Fig. 3a**. **d.** Scatter plot

showing the number of synapses identified by mcDETECT in $50 \times 50\ \mu m^2$ square spots in the 2.5-month (left) and 5.7-month (right) TgCRND8 samples. **e.** Line plot comparing the regional synapse density trends inferred by mcDETECT between the 2.5-month and 5.7-month TgCRND8 samples. **f.** Line plot comparing the regional neuron density trends between the 2.5-month and 5.7-month TgCRND8 samples. **g.** Dot plot showing the enrichment levels of five upregulated and five downregulated genes in the identified in the 5.7-month TgCRND8 sample compared to the 2.5-month sample.

Application to multi-sample 10x Xenium mouse brain data in AD progression

Studies have shown that synaptic loss and dysfunction are among the earliest pathological changes in neurodegenerative diseases, often occurring before neuronal death[5,60]. Investigating synapses offers critical insights into disease mechanisms, facilitating early diagnosis and the development of targeted therapies. However, the transcriptomic changes occurring in synapses during neurodegeneration remain poorly understood. In this study, we demonstrate how mcDETECT can fill this gap.

We applied mcDETECT to a 10x Xenium dataset derived from both WT and TgCRND8 mouse models[35], the latter being one of the most widely used models for AD[61]. This dataset includes six sections across three developmental stages. Our analysis focused on AD samples at two key ages: 2.5 months, representing a pre-symptomatic healthy stage, and 5.7 months, corresponding to a stage with observable symptoms[62]. We first applied mcDETECT to WT mice at these two ages as a control and found minimal changes in synapse density across nine major brain regions, confirming negligible batch effects in these samples since they were generated from the same study (**Supplementary Note 5**). Next, we compared the AD samples and observed a significant decline in synapse density as the disease progressed from 2.5 to 5.7 months, as shown in the synapse distribution and density scatter plots in **Fig. 6c** and **6d**. This

reduction is particularly pronounced in the CA, DG, and cortical subplate (CTXsp), as summarized in regional density plot in **Fig. 6e**. Notably, despite the significant decline in synapses, the number of neurons within each brain region remains constant (**Fig. 6f**). This discrepancy underscores synaptic loss as one of the earliest detectable signs of AD pathology, preceding observable neuronal loss[41].

Next, we compared the synapse transcriptome profiles obtained from mcDETECT in the two AD samples. DE analysis between synapses from the 2.5- and 5.7-month samples identifies genes significantly altered at the later AD stage, as detailed in **Fig. 6g**. Compared to the early stage, we observed the upregulation of *C4a*, a gene encoding a protein that is a key component of the complement system and plays a role in synaptic pruning[63], and downregulation of *Cbln1*, a gene encoding a synaptic organizer protein essential for maintaining excitatory synapses, whose reduction can destabilize synaptic connections[64]. Despite the limited gene coverage, these transcriptomic alterations align with the observed synaptic loss and dysfunction, further demonstrating mcDETECT's ability to track both spatial and molecular changes during disease progression.

Discussion

We have introduced mcDETECT, an algorithm designed to study synapses in iST data. While most existing iST brain studies focus on neurons and glial cells but overlook synapses, mcDETECT bridges this gap by reconstructing individual synapse locations and their transcriptome profiles from subcellular mRNA distributions in 3D—the first method of this kind. Our extensive analyses of seven publicly available iST mouse brain datasets from multiple platforms, including 10x Xenium, 10x Xenium 5K, CosMx, and MERSCOPE, not only demonstrate mcDETECT's ability to identify individual synapses but also characterize their

spatial transcriptome profiles and distinguish between pre- and post-synaptic structures. Moreover, using our specially designed synaptic-focused gene panel, we further demonstrate mcDETECT's ability to identify additional synapse subtypes based on their transcriptome profiles.

We also show that mcDETECT does not require an extensive list of synaptic markers, but fewer than 10 markers can identify most synapses. This capability allows for the re-analysis of existing datasets to provide deeper insights into neuronal network changes across various neurodegenerative diseases. For example, by comparing iST data from wild type and AD mouse models from an existing study, mcDETECT can detect changes in synapse numbers in specific brain regions and the underlying transcriptomic changes associated with AD progression.

mcDETECT identifies synapses based on mRNA presence, restricting its detection to actively functioning synapses while potentially missing inactive synapses that lack mRNA activity. This explains the discrepancy between the absolute number of synapses detected by mcDETECT and those identified using electron microscopy. However, since our goal is to study the transcriptome profile of synapses, excluding inactive synapses is not a limitation. Moreover, our method recovers the overall synapse density trend, indicating that the proportion of active synapses across different brain regions remains consistent. Although mcDETECT was originally developed and evaluated using mouse brain data due to its accessibility, it is a versatile method that can be applied to human iST data by using human-specific synaptic markers and adjusting synapse size filtering. As spatial omics technologies continue to advance and datasets with larger gene panels become available, we anticipate that mcDETECT will enable the identification of more synapses and their subtypes. This will play a crucial role in enhancing our understanding of subcellular neuronal connections and transmission, ultimately contributing to

early diagnosis and the development of synapse-targeted precision therapies for neurodegenerative diseases.

Acknowledgements

J.H. was partly supported by Emory AI Humanity Initiative and Georgia Clinical & Translational Science Alliance. B.Y. was supported by R01MH117122, R01AG062577, R01AG064786, R01AG078937, and R01NS118819.

Author Contributions

This study was conceived and led by J.H. and B.Y. C.Y. designed the model and algorithm with input from J.H., implemented the mcDETECT software, and led data analyses with evaluation input from V.G.C., Hailing S., S.D., J.Y., P.J., and B.Y. K.P., Hongshun S., H.-L.V.W., and R.L. generated the MERSCOPE mouse brain data. F.W. contributed to protein staining. Y.L. contributed to part of the data analysis. J.H. and C.Y. wrote the paper with feedback from all other co-authors.

Competing Interests

The authors declare no competing interests.

Methods

Input data

mcDETECT is designed to accommodate a wide range of iST data from various techniques. Most iST data provide a transcript file, which serves as the key input for identifying synapse locations. This transcript file contains the gene identity and (x, y, z) coordinates of each imaged mRNA molecule. Additionally, the input data should include a binary label indicating whether each mRNA molecule overlaps with a cell nucleus. If this in-nucleus label is not available, we can generate it by overlaying all mRNA transcripts onto the corresponding DAPI image of the tissue region. By matching the transcript coordinates with the image coordinates, we determine the relative position of each mRNA molecule with respect to cell nuclei. This process allows us to assign binary in-nucleus labels to the transcripts. In addition to the transcript file, mcDETECT requires a user-defined list of synaptic markers and negative control markers as input. In this study, we have summarized a set of synaptic markers in **Supplementary Table 3**. For negative control markers, we utilize the genes identified by Niu *et al*. as being enriched in neuronal nuclei (**Supplementary Table 7**) and retain those available in each iST dataset.

mcDETECT algorithm

mcDETECT is an end-to-end framework that processes transcript files from the iST platform along with user-defined synaptic markers, returning the spatial characteristics and transcriptome profiles of individual synapses. The complete mcDETECT algorithm consists of the following four steps:

Step 1: Density-based clustering on each single marker to identify extranuclear aggregations

The fundamental step of mcDETECT is to identify extranuclear aggregations of mRNAs in the input transcript file through an approach known as Density-Based Spatial Clustering with

Applications of Noise (DBSCAN)[65]. DBSCAN operates based on two predefined parameters: the search radius r and the minimum number of samples m. Denote T as the set of mRNAs from a single synaptic marker, where each mRNA $t_i \in T$ has 3D spatial coordinates (x_i, y_i, z_i). We first calculate the Euclidean distance between all mRNA pairs $t_i, t_j \in T$ by:

$$d(t_i, t_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}.$$

The neighbors of t_i are defined as the mRNAs falling within its neighborhood of radius r, and the number of its neighbors is recorded by:

$$N_i = \sum_{t_j \in T} I(d(t_i, t_j) < r),$$

where $I(\cdot)$ is the indicator function. Next, DBSCAN classifies the mRNAs within the set T as core points, border points, and noise points according to the following criteria:

1) a point t_i is a core point if the number of its neighbors exceeds or equals the minimum number of samples m, i.e., $N_i \geq m$;
2) a point t_j is directly reachable from a core point t_i if it is within the neighborhood of t_i, i.e., $d(t_i, t_j) < r$;
3) a point t_j is reachable from a core point t_i if there is a path $\{t_1, t_2, \cdots, t_n\}$ such that $t_i = t_1$ and $t_j = t_n$, where each point t_{i+1} is directly reachable from point t_i for $i \in \{1, \cdots, n-1\}$;
4) a point t_j is a border point if the number of its neighbors is less than the minimum number of samples m, i.e., $N_j < m$, but it is reachable from a core point t_i;
5) a point t_k is a noise point if it is not reachable from any other point.

A core point forms a cluster along with all points that are reachable from it, possibly consisting of other core points and border points, and each cluster represents a distinct aggregation of mRNAs. Based on this paradigm, DBSCAN effectively extracts the pattern where mRNAs aggregate while excluding the dissociated mRNAs that may not represent true biological structures.

The selection for the search radius r and the minimum number of samples m is highly influential on the DBSCAN output. In this study, we fix r at $1.5\,\mu m$, considering the size of synapses reported in prior studies. mcDETECT utilizes an automated procedure to determine the minimum number of samples m for different synaptic markers based on their transcript density. For a given synaptic marker, we first calculate its overall density λ as:

$$\lambda = \frac{\text{total number of mRNAs}}{\text{area of the tissue slice}}.$$

We then compare the observed mRNA distribution against a Complete Spatial Randomness (CSR) process to assess whether it exhibits an aggregation pattern. Under CSR, the number of mRNAs within a circle of radius r follows a Poisson distribution with mean $\lambda\pi r^2$. To distinguish true aggregations from random dispersion, we set the ideal minimum number of samples m^* such that:

$$\Pr(\text{Poisson}(\alpha\lambda\pi r^2) \leq m^*) = 0.95.$$

In other words, at significance level 0.95, the ideal m^* defined as such identifies a distribution pattern more clustered than a CSR process with density $\alpha\lambda$ within a circle of radius r. Notably, due to the typically low values of λ in real data, an additional scaling factor α is introduced to refine the criteria for selecting m^*, enhancing the accuracy of the identified aggregations. In this study, we fix $\alpha = 5$ across all iST datasets and find that the results are not sensitive to further increasing its value. Finally, in practice, we set the minimum number of samples m to be:

$$m = \lfloor max(m^*, 3) \rceil,$$

where a flooring threshold of 3 is set to ensure that an identified aggregation contains at least three mRNAs. We also round m to its nearest integer. In conclusion, this parameter selection procedure allows mcDETECT to automatically determine the best search radius r and minimum number of samples m to identify statistically significant aggregations for different input genes.

For each identified aggregation, mcDETECT constructs a minimum enclosing sphere to encompass all its mRNAs. The centroid and radius of this sphere respectively serve as proxies for the location and size of its associated aggregation. Although the aggregations are defined using a single synaptic marker, the enclosing spheres enable evaluating the presence of mRNAs from other synaptic markers within them. The location, size, and synaptic marker composition are then recorded as spatial characteristics of the identified aggregations and are used for downstream filtering. Aggregations with an excessively large radius ($4\,\mu m$) are removed, as they are likely to represent cell fragments rather than synapses. Additionally, mcDETECT uses the intracellular positions of mRNAs to assess the in-nucleus ratio of each aggregation. Aggregations are filtered out if the percentage of synaptic marker mRNAs located inside nuclei exceeds a predefined threshold τ. Specifically, let B denote the set of n mRNAs from all synaptic markers within a minimum enclosing sphere. mcDETECT discards its associated aggregation if

$$\frac{1}{n}\sum_{b_i \in B} I(b_i \text{ in nucleus}) > \tau.$$

The value of τ may vary across iST platforms due to differences in their pipelines for segmenting nuclei and determining the intracellular positions of mRNAs. In this study, we set $\tau = 0.2$ for iST datasets that include in-nucleus labels for each transcript. For datasets lacking such labels, where we generate them using DAPI images, we lower τ to 0.1 to be more conservative.

Step 2: Integrating aggregations from multiple markers

To reveal the overall distribution of synapses as informed by all synaptic markers, mcDETECT integrates aggregations from multiple markers while considering their colocalization pattern in synapses. To achieve this, mcDETECT incorporates an iterative procedure based on the relative positions of spheres in 3D space. Specifically, let s_1 and s_2 represent two spheres with

centroids c_1 and c_2 and radii r_1 and r_2, respectively. These spheres correspond to aggregations from two distinct synaptic markers. We first calculate the distance between c_1 and c_2:

$$d(c_1, c_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2},$$

where (x_1, y_1, z_1) and (x_2, y_2, z_2) are the 3D spatial coordinates of c_1 and c_2, respectively. Based on this inter-centroid distance, the relative positions of s_1 and s_2 can be classified and further proceed as follows:

1) Disjoint if $d(c_1, c_2) \geq r_1 + r_2$, and both s_1 and s_2 are retained.
2) Contained if $0 \leq d(c_1, c_2) \leq |r_2 - r_1|$, and the smaller sphere is merged into the larger one.
3) Intersecting if $|r_2 - r_1| < d(c_1, c_2) < r_1 + r_2$. If the overlap between s_1 and s_2 is significant, defined by $d(c_1, c_2) < \rho \cdot (r_1 + r_2)$, where ρ is a threshold parameter, mRNAs from s_1 and s_2 are merged, and a new enclosing sphere is generated. Otherwise, both s_1 and s_2 are retained. In this study, we set $\rho = 0.2$, indicating a heavy overlap, to prevent the merging of two closely located but distinct synapses.

In mcDETECT, mRNA aggregations from different synaptic markers are iteratively merged using the above criteria, ensuring efficient integration of multiple signals.

Step 3: Filtering based on negative control enrichment

Following the multi-marker integration step, mcDETECT performs another round of filtering based on the enrichment of negative control markers. In this study, we select these markers from a list of genes reported by Niu *et al.* as being more enriched in neuronal nuclei than in synapses. For each aggregation, we count the negative control mRNAs within its associated minimum enclosing sphere. An aggregation is filtered out if the ratio of negative control mRNAs to synaptic marker mRNAs exceeds 10%. Overall, the negative control filtering step reduces false positives caused by cell fragments or random mRNA debris. Aggregations that pass

screenings for size, in-nucleus ratio, and negative control enrichment are considered individual synapses.

Step 4: Spatial transcriptome profile construction for individual synapses

To construct the transcriptome profile of the identified synapses, mcDETECT maps the mRNA transcripts from all genes profiled in the iST dataset to the minimum enclosing spheres that represent individual synapses. By counting the number of mRNAs from each gene, mcDETECT constructs a gene count matrix for all identified synapses across all genes. Combining the count matrix with the spatial coordinates, we obtain a comprehensive spatial transcriptome profile of all individual synapses in the brain tissue. This profile can be used for further investigation of synapse heterogeneity and molecular alterations in neurodegenerative diseases.

Downstream analyses

Spatial domain assignment

To identify spatial domains in an iST dataset, we first generate a 2D grid over the tissue section, where each grid spot represents a $50 \times 50\ \mu m^2$ region. Next, we reconstruct spatial gene expression by mapping all mRNA transcripts onto these pseudo-spots. Using this spot-level gene expression data, we apply a spatial clustering method, SpaGCN[66], to identify distinct spatial domains. These domains are subsequently annotated as major brain regions based on anatomic maps from the Allen Mouse Brain Atlas. Finally, the assigned brain region labels for each pseudo-spot are propagated to all cells and synapses within that spot, facilitating region-specific spatial analysis at multiple biological scales.

Synapse density

The $50 \times 50\ \mu m^2$ pseudo-spots are used as the smallest units for summarizing synapse characteristics across the tissue region, such as density. The synapse density at the spot level

is calculated as the ratio of the number of synapses to the spot area. At the regional level, synapse density is determined by averaging the spot-level densities for all spots within a given brain region. To evaluate the similarity between regional synapse densities generated by mcDETECT and the genetic labeling approach, considered the ground truth, we utilize the Spearman correlation coefficient weighted by brain region areas. Specifically, let $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$ represent the synapse densities across n brain regions derived from the two approaches, respectively, and let $\{w_1, \ldots, w_n\}$ denote the corresponding areas of these brain regions. We begin by computing the ranks $R_A = \{R(a_1), \ldots, R(a_n)\}$ and $R_B = \{R(b_1), \ldots, R(b_n)\}$. Next, we calculate the weighted mean rank within group A by $\overline{R}_A^w = \sum_{i=1}^{n} w_i R(a_i) \,/\, \sum_{i=1}^{n} w_i$, and the calculation for $\overline{R}_B^w$ is likewise. Based on these weighted means, the weighted Spearman correlation is defined as:

$$\rho_S^w = \frac{\sum_{i=1}^{n} w_i (R(a_i) - \overline{R}_A^w)(R(b_i) - \overline{R}_B^w)}{\sqrt{\sum_{i=1}^{n} w_i (R(a_i) - \overline{R}_A^w)^2 \cdot \sum_{i=1}^{n} w_i (R(b_i) - \overline{R}_B^w)^2}}.$$

This weighted correlation coefficient provides a more accurate measure of similarity between regional densities by assigning appropriate significance to each brain region based on its area.

Synapse subtyping

To classify synapses into pre- and post-synapses, we use a predefined list of pre- and post-synaptic markers and apply K-Means clustering based on their enrichment levels in the identified synapses. Within each synapse cluster, we calculate the pre- and post-synaptic mRNA counts, normalize them by the number of corresponding markers used, and denote them as N_{pre} and N_{post}, respectively. A cluster is assigned as pre-synapse if $N_{pre}/N_{post} > 55\%$, as post-synapse if $N_{pre}/N_{post} < 45\%$, and as “neutral” otherwise. Repeating this procedure across all synapse clusters assigns each cluster a meaningful subtype label. Similarly, synapse subtyping based on the markers reported by Niu *et al*. involves unsupervised clustering of their

enrichment levels. The resulting synapse clusters are then matched to the reported subtypes by assessing the enrichment patterns of the subtype markers.

Simulation

Simulation of a single synaptic marker

The simulation is conducted over a $2000 \times 2000\ \mu m^2$ area with a thickness $10.5\ \mu m$, mimicking the scale of the Xenium platform. To model the subcellular mRNA distribution of a single synaptic marker, we assume it consists of three spatial patterns: complete spatial randomness (CSR), intranuclear aggregation, and extranuclear aggregation. The simulation procedures for generating the 3D coordinates (x, y, z) for each mRNA follow the methodology described by Diggle[67], as detailed below.

1) CSR. The CSR pattern is modeled using a homogeneous Poisson process, where the x, y, and z coordinates are randomly sampled from uniform distributions within the feasible region. This process exhibits properties identical to CSR.

2) Intranuclear aggregation. This aggregation pattern is modeled using a Poisson cluster process, where we first generate the locations of nucleus centroids using a homogeneous Poisson process. Next, a random number of mRNAs are assigned to the vicinity of each nucleus centroid, forming clusters around them. The number of mRNAs for each nucleus is randomly sampled from a Poisson distribution. The distances d between mRNAs and their respective nucleus centroids are independent and identically distributed (i.i.d.) exponential random variables. The directional angles of these distances are also i.i.d., determined by two uniform random variables: φ, the polar angle, ranging within $(0, \pi)$, and ϕ, the azimuthal

angle, ranging within $(0, 2\pi)$. Consequently, the 3D offsets from nucleus centroids to mRNAs, denoted by (x^*, y^*, z^*), are given as:

$$x^* = d \sin\varphi \cos\phi,$$

$$y^* = d \sin\varphi \sin\phi,$$

$$z^* = d \cos\varphi.$$

The coordinates of each mRNA (x, y, z) are then calculated by adding these offsets (x^*, y^*, z^*) to the coordinates of the corresponding nucleus centroid. Repeating this procedure for all nuclei produces a complete clustering pattern within the simulated area.

To characterize the intranuclear property, we additionally simulate a binary in-nucleus label for each mRNA. Within each cluster, these labels are randomly assigned based on a Bernoulli distribution with success probability p, representing the likelihood of an mRNA being located within a cell nucleus. This probability p, varying across clusters, is itself randomly sampled from a $\mathrm{Beta}(8, 2)$ distribution, ensuring an overall mean in-nucleus ratio of 0.8. Moreover, the distances d between mRNAs and their respective nucleus centroids are sampled to have a mean of $3.5\ \mu m$, reflecting typical nucleus sizes.

3) Extranuclear aggregation. The simulation for this aggregation pattern follows the same strategy as intranuclear aggregation, except that p is sampled from a $\mathrm{Beta}(2, 8)$ distribution, resulting in a mean in-nucleus ratio of 0.2. Additionally, d is sampled with a lower mean of $1\ \mu m$, reflecting synapse sizes reported in previous studies.

The final step overlays the CSR points, intranuclear aggregations, and extranuclear aggregations within the simulated area to create the complete spatial pattern. A similar simulation can be performed in a 2D plane by omitting the z-axis from all resulting mRNAs.

Simulation of multiple synaptic markers

We consider three synaptic markers in the multi-marker scenario. Similar to the single-marker setting, each marker's subcellular mRNA distribution follows one of three spatial patterns: CSR, intranuclear aggregation, or extranuclear aggregation, with the latter two distinguished by differences in in-nucleus ratios and aggregation sizes. Additionally, we also account for the colocalization pattern of multiple synaptic markers within nuclei and synapses. The simulation procedures for this scenario are detailed below:

1) CSR. The CSR patterns are simulated independently for each synaptic marker.

2) Intranuclear aggregation. This aggregation pattern is modeled using a multi-sample Poisson cluster process. We begin by generating nucleus centroid locations using a homogeneous Poisson process. Next, we determine which of the three markers are expressed in each nucleus. Specifically, let K be a categorical random variable representing the number of markers expressed in a nucleus, drawn from the following probability mass function (*pmf*):

$$p(K = 1) = 0.8,$$

$$p(K = 2) = 0.1,$$

$$p(K = 3) = 0.1.$$

Given the assigned value of K for each nucleus, we randomly select the corresponding markers. Subsequently, for each expressed marker, we independently simulate its mRNA cluster following the procedures described in the single-marker setting. As a result, the generated clusters exhibit diverse combinations of marker types.

3) Extranuclear aggregation. The simulation for this aggregation pattern follows the same strategy as intranuclear aggregation. In addition to differences in in-nucleus ratios and aggregation sizes, we modify the *pmf* of K to:

$$p(K = 1) = 0.4,$$

$$p(K = 2) = 0.3,$$

$$p(K = 3) = 0.3,$$

which reflects the higher degree of colocalization of synaptic markers within synapses.

For both single-marker and multi-marker settings, we apply mcDETECT to the simulate mRNAs to recover the centroids of extranuclear aggregations, which serve as ground truth. To evaluate the detection performance, we calculate the accuracy, precision, and recall based on whether such centroids are encompassed within the spheres generated by mcDETECT. The detailed setup for the simulation study, including marker density, the number of aggregations, and the parameters governing all distributions, is summarized in **Supplementary Table 4**.

Protein staining

Animals

Wild-type mice were obtained from the Jackson Laboratory (Wild Type: C57BL/6J 000664). Whole brains were collected at 3 months of age. All animal procedures and protocols were approved by the Emory University Institutional Animal Care and Use Committee (Protocol ID: PROTO201800040).

Immunofluorescence

Three-month-old mice were anesthetized with isoflurane and perfused with 25 mL of cold phosphate-buffered saline (PBS) and 25 mL of 4% paraformaldehyde (PFA). After perfusion, the whole brain was collected and post-fixed with 4% PFA overnight in a tube at 4 °C. The

brains were then transferred to 30% sucrose for 24–48 hours, ensuring that the brain completely sank to the bottom of the tube. The brain was then sectioned into 12 to 20 μm slices using a cryostat (Leica CM1850) at -20 °C. The brain slices were frozen at -80 °C for storage. Slices were taken from -80 °C and dried at room temperature until no water remained on the slides. The brain slices were washed three times for 5 minutes each with PBS and then blocked with 5% normal goat serum (NGS) and 0.2% Triton X-100 in PBS at room temperature for 30 minutes. Primary antibodies (1:500, SAP102, PA5-29116) were incubated with 5% NGS in PBS at 4 °C overnight. On the next day, the brain slices were washed three times for 5 minutes each and then incubated with the secondary antibody in 5% NGS in PBS at room temperature for 1 hour in darkness. The brain slices were washed again three times for 5 minutes each, followed by mounting with DAPI (1:1000, Thermo Scientific, 62248). Images were captured using a Nikon Eclipse Ti-E microscope.

MERSCOPE data generation

Tissue preparation and hybridization for MERSCOPE

A 10-μm thick cryosection of a fresh frozen brain region from a 6–8-week-old female C57BL/6J mouse was generously provided by Vizgen. The tissue section was mounted on the MERSCOPE slide and stored in 70% ethanol at 4 °C.

Following the manufacturer's protocol manual (Vizgen, MERSCOPE Fresh and Fixed Frozen Tissue Sample Preparation User Guide, 91600002, Rev F, 2023), the section was prepared for MERSCOPE probe hybridization. The MERSCOPE slide was removed from 70% ethanol, transferred into a small sterile petri dish with the tissue side facing up, and briefly rinsed once with 5 mL of Sample Prep Wash Buffer. Then, we aspirated the Sample Prep Wash Buffer and incubated the section in 5 mL of Formamide Wash Buffer at 37 °C for 30 minutes in a humidified incubator. After 30 minutes, the Formamide Wash Buffer was carefully aspirated from all around

the section, and we added 50 mL of our predesigned 300-gene panel probe directly onto the tissue section. Then, we used tweezers to peel off the $2 \times 2\ cm^2$ parafilm to carefully cover and seal the probe on the tissue section. We then sealed the petri dish and incubated the section at 37 °C for 36–48 hours in a humidified incubator.

After probe hybridization, the section was incubated twice in 5 mL of Formamide Wash Buffer at 47 °C for 30 minutes each. While incubating, we prepared the Gel Embedding Solution containing Gel Embedding Premix, 10% w/v ammonium persulfate solution, and N,N,N',N',-tetramethylethylethylenediamine (5 mL per sample) and placed it on ice until ready to use. Then, we prepared the Gel Coverslip by applying 100 mL of Gel Slick Solution onto the Gel Coverslip for at least 10 minutes at room temperature. After the second Formamide Wash Buffer incubation, we briefly washed the section with 5 mL of Sample Prep Wash Buffer. From the 5 mL of freshly prepared Gel Embedding Solution, we retained 100 mL and added the remaining solution into the petri dish containing our section, incubating it for 1 minute at room temperature. Then, we transferred the Gel Embedding Solution into a waste 50 mL tube to monitor gel formation and applied 50 mL of the retained Gel Embedding Solution directly onto the tissue section. We gently wiped the coverslip to remove any residue, then lifted the coverslip using tweezers, ensuring that the Gel Slick-treated side faced downward toward the tissue. Using another tweezer to guide and angle the coverslip, we slowly mounted it onto the tissue section without trapping air bubbles. We adjusted the position of the coverslip so that it was centered on the MERSCOPE slide. Excess Gel Embedding Solution was removed by gently tapping on the coverslip and carefully aspirating the excess solution from the sides. Without disturbing the petri dish, we allowed 1.5 hours of incubation at room temperature for gel formation. After that, we used the sharpened tip of a hobby blade to carefully lift the coverslip and proceeded to the next step of clearing for non-resistant fresh frozen tissue.

We used prewarmed Clearing Premix at 37 °C for 30 minutes and prepared a 5 mL Clearing Solution containing the Clearing Premix and 50 μL of Proteinase K. The Clearing Solution was then added to the petri dish containing the section, after which we sealed the petri dish and placed it in a humidified 37 °C cell culture incubator for 24 hours or until the tissue section became transparent.

MERSCOPE imaging

To prepare our tissue section for imaging, we first stained it with DAPI and PolyT stain. We began by warming a water bath to 37 °C, ensuring the water level was adequate but did not cover the valve of the Imaging Cartridge. The Imaging Cartridge was placed in the water bath for 60 minutes to allow it to thaw. Additionally, the DAPI and PolyT Staining reagent were prewarmed for 10 minutes in a 37 °C water bath. Once the tissue was cleared and became transparent, we aspirated the Clearing Solution and washed the section twice with 5 mL of Sample Prep Wash Buffer. Then, we added 3 mL of prewarmed DAPI and PolyT Staining reagent and allowed it to incubate for 15 minutes on a rocker. This was followed by a wash with 5 mL of Formamide Wash Buffer, with a 10-minute incubation. The section was washed again with 5 mL of Sample Prep Wash Buffer and left in the buffer until the MERSCOPE instrument was ready for imaging.

For MERSCOPE instrument operation and data acquisition, we followed the manufacturer's protocol manual (Vizgen, MERSCOPE Instrument User Guide, 91600001, Rev H, 2023). Before starting the MERSCOPE run, we first washed the instrument according to the manual's instructions. In this experiment, we did not perform the MERSCOPE Cell Boundary Stain or any additional antibody staining. We ensured that the Imaging Cartridge was completely thawed, then slowly inverted the Cartridge ten times to mix the reagents inside. The Cartridge was activated by piercing the foil at the Cartridge Activation Port with a clean pipette tip. Using a 1-

mL pipette set to 1 mL, we carefully added all of the Imaging Activation Mix into the bottom of the Cartridge via the Activation Port and plunged the pipette ten times at a moderate speed to thoroughly mix the solution inside. Then, using a serological pipette, we carefully layered 15 mL of mineral oil over the liquid inside the Imaging Cartridge. The Imaging Cartridge was inserted into the MERSCOPE instrument, with the barcode side facing the front of the instrument.

Next, we assembled the MERSCOPE Flow Chamber parts along with the MERSCOPE Slide containing the tissue section in the correct orientation. The Flow Chamber was locked in place, and image acquisition proceeded according to the manual's instructions. The total MERSCOPE imaging time for this experiment was approximately 24–36 hours, followed by data acquisition.

Regional connectivity map

Oh *et al*. reported a whole-brain connectivity matrix based on axonal projection strengths measured using enhanced green fluorescent protein (EGFP)-expressing adeno-associated viral vectors. From their connectivity matrix, we select the subset of brain regions available in our iST datasets, resulting in a square matrix where each element represents the connectivity strength from a source region (rows) to a target region (columns). We then calculate $\mathrm{ex}_{\mathrm{total}}$ and $\mathrm{im}_{\mathrm{total}}$, representing the total export and import strengths of each region, by summing the corresponding rows and columns. Based on these values, we defined an export-to-import score (E/I score) for each region as:

$$\mathrm{E/I\ score} = \frac{\mathrm{ex}_{\mathrm{total}} - \mathrm{im}_{\mathrm{total}}}{\mathrm{ex}_{\mathrm{total}} + \mathrm{im}_{\mathrm{total}}}.$$

This score, ranging from -1 to 1, represents the relative role of a brain region in terms of sending or receiving axonal projections. An E/I score closer to 1 indicates that the region primarily functions as an outputting unit, whereas a score closer to -1 suggests it predominantly serves as an input-receiving unit.

Data availability

We analyzed the following publicly available iST datasets: (1) Xenium 5K mouse brain data (https://www.10xgenomics.com/datasets/xenium-prime-fresh-frozen-mouse-brain); (2) CosMx mouse brain data (https://nanostring.com/products/cosmx-spatial-molecular-imager/ffpe-dataset/cosmx-smi-mouse-brain-ffpe-dataset/); (3) MERFISH mouse brain data (https://info.vizgen.com/mouse-brain-map); (4) 10x Xenium mouse brain data (https://www.10xgenomics.com/cn/datasets/xenium-in-situ-analysis-of-alzheimers-disease-mouse-model-brain-coronal-sections-from-one-hemisphere-over-a-time-course-1-standard). We also analyzed an in-house MERSCOPE mouse brain dataset. It has been uploaded to GEO with restricted access and will be publicly accessible upon the acceptance of the manuscript. The accession number can be requested from the corresponding authors. For temporary access, we have uploaded the data to Google Drive: https://drive.google.com/drive/folders/1ydX0_n4htW1CECLYMgsKfsFzYBFhLXH7?usp=sharing. Details of the datasets analyzed in this paper are described in **Supplementary Table 1**.

Software availability

An open-source implementation of the mcDETECT algorithm can be downloaded from GitHub: https://github.com/chen-yang-yuan/mcDETECT.

References

1 Sheng, M., Sabatini, B. L. & Südhof, T. C. Synapses and Alzheimer's disease. *Cold Spring Harbor perspectives in biology* **4**, a005777 (2012).

2 Glock, C. *et al.* The translatome of neuronal cell bodies, dendrites, and axons. *Proceedings of the National Academy of Sciences* **118**, e2113929118 (2021).

3 Terry, R. D. *et al.* Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* **30**, 572-580 (1991).

4 Glenner, G. G. & Wong, C. W. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochemical and biophysical research communications* **120**, 885-890 (1984).

5 Arendt, T. Synaptic degeneration in Alzheimer's disease. *Acta neuropathologica* **118**, 167-179 (2009).

6 D Skaper, S., Facci, L., Zusso, M. & Giusti, P. Synaptic plasticity, dementia and Alzheimer disease. *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders)* **16**, 220-233 (2017).

7 Picciotto, M. R., Higley, M. J. & Mineur, Y. S. Acetylcholine as a neuromodulator: cholinergic signaling shapes nervous system function and behavior. *Neuron* **76**, 116-129 (2012).

8 Liao, J., Lu, X., Shao, X., Zhu, L. & Fan, X. Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics. *Trends Biotechnol* **39**, 43-58, doi:10.1016/j.tibtech.2020.05.006 (2021).

9 Waylen, L. N., Nim, H. T., Martelotto, L. G. & Ramialison, M. From whole-mount to single-cell spatial assessment of gene expression in 3D. *Commun Biol* **3**, 602, doi:10.1038/s42003-020-01341-1 (2020).

10 Burgess, D. J. Spatial transcriptomics coming of age. *Nat Rev Genet* **20**, 317, doi:10.1038/s41576-019-0129-z (2019).

11 Asp, M., Bergenstrahle, J. & Lundeberg, J. Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. *Bioessays* **42**, e1900221, doi:10.1002/bies.201900221 (2020).

12 Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat Rev Genet* **16**, 57-66, doi:10.1038/nrg3832 (2015).

13 Moor, A. E. & Itzkovitz, S. Spatial transcriptomics: paving the way for tissue-level systems biology. *Curr Opin Biotechnol* **46**, 126-133, doi:10.1016/j.copbio.2017.02.004 (2017).

14 Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).

15 Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* **11**, 360-361, doi:10.1038/nmeth.2892 (2014).

16 Shah, S., Lubeck, E., Zhou, W. & Cai, L. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* **92**, 342-357, doi:10.1016/j.neuron.2016.10.001 (2016).

17 Eng, C. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235-239, doi:10.1038/s41586-019-1049-y (2019).

18 Moffitt, J. R. *et al.* Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, doi:10.1126/science.aau5324 (2018).

19 Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090, doi:10.1126/science.aaa6090 (2015).

20 Petukhov, V. *et al.* Cell segmentation in imaging-based spatial transcriptomics. *Nature biotechnology* **40**, 345-354 (2022).

21 Littman, R. *et al.* Joint cell segmentation and cell type annotation for spatial transcriptomics. *Molecular systems biology* **17**, e10108 (2021).

22 Si, Y. *et al.* FICTURE: scalable segmentation-free analysis of submicron-resolution spatial transcriptomics. *Nature Methods* **21**, 1843-1854 (2024).

23 Mah, C., Ahmed, N. & Lam, D. Bento: A toolkit for subcellular analysis of spatial transcriptomics data. bioRxiv. *Published online June* **13**, 2022.2006 (2022).

24 Shi, H. *et al.* Spatial atlas of the mouse central nervous system at molecular resolution. *Nature* **622**, 552-561 (2023).

25 Wang, G. *et al.* Spatial organization of the transcriptome in individual neurons. *BioRxiv*, 2020.2012. 2007.414060 (2020).

26 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).

27 Santuy, A. *et al.* Estimation of the number of synapses in the hippocampus and brain-wide by volume electron microscopy and genetic labeling. *Scientific Reports* **10**, 14014 (2020).

28 Khan, K., Rehman, S. U., Aziz, K., Fong, S. & Sarasvady, S. in *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014).* 232-238 (IEEE).

29 Cizeron, M. *et al.* A brainwide atlas of synapses across the mouse life span. *Science* **369**, 270-275 (2020).

30 Niu, M. *et al.* Droplet-based transcriptome profiling of individual synapses. *Nature Biotechnology*, 1-13 (2023).

31 Zeng, H. *et al.* Spatially resolved single-cell translatomics at molecular resolution. *Science* **380**, eadd3067 (2023).

32 Dataset. Mouse brain hemisphere (coronal). *10x Genomics, Xenium 5K,* https://www.10xgenomics.com/datasets/xenium-prime-fresh-frozen-mouse-brain (2024).

33 Dataset. Mouse brain hemisphere (coronal). *nanoString, CoxMx,* https://nanostring.com/products/cosmx-spatial-molecular-imager/ffpe-dataset/cosmx-smi-mouse-brain-ffpe-dataset/ (2024).

34 Dataset. Mouse brain (coronal). *vizgen, MERFISH,* https://info.vizgen.com/mouse-brain-map (2021).

35 Dataset. Mouse brain hemisphere (coronal). *10x Genomics, Xenium,* https://www.10xgenomics.com/cn/datasets/xenium-in-situ-analysis-of-alzheimers-disease-mouse-model-brain-coronal-sections-from-one-hemisphere-over-a-time-course-1-standard (2023).

36 Dataset. Mouse brain (coronal). *in-house, MERSCOPE.* (2024).

37 Biesemann, C. *et al.* Proteomic screening of glutamatergic mouse brain synaptosomes isolated by fluorescence activated sorting. *The EMBO journal* **33**, 157-170 (2014).

38 Dunkley, P. R., Jarvie, P. E. & Robinson, P. J. A rapid Percoll gradient procedure for preparation of synaptosomes. *Nature protocols* **3**, 1718-1728 (2008).

39 Kumar, S. Synaptosome microRNAs: emerging synapse players in aging and Alzheimer's disease. *Neural Regeneration Research* **18**, 1275-1276 (2023).

40 Keller, D., Erö, C. & Markram, H. Cell densities in the mouse brain: a systematic review. *Frontiers in neuroanatomy* **12**, 83 (2018).

41 Bensussen, S. *et al.* A viral toolbox of genetically encoded fluorescent synaptic tags. *Iscience* **23** (2020).

42 Cosker, K. E., Fenstermacher, S. J., Pazyra-Murphy, M. F., Elliott, H. L. & Segal, R. A. The RNA-binding protein SFPQ orchestrates an RNA regulon to promote axon viability. *Nature neuroscience* **19**, 690-696 (2016).

43 Jakovcevski, M. *et al.* Neuronal Kmt2a/Mll1 histone methyltransferase is essential for prefrontal synaptic plasticity and working memory. *Journal of Neuroscience* **35**, 5097-5108 (2015).

44 Jo, M. *et al.* The role of TDP-43 propagation in neurodegenerative diseases: integrating insights from clinical and experimental studies. *Experimental & molecular medicine* **52**, 1652-1662 (2020).

45 Wang, J.-Y. *et al.* Sun1 deficiency leads to cerebellar ataxia in mice. *Disease models & mechanisms* **8**, 957-967 (2015).

46 Kempermann, G., Song, H. & Gage, F. H. Neurogenesis in the adult hippocampus. *Cold Spring Harbor perspectives in biology* **7**, a018812 (2015).

47 Denoth-Lippuner, A. & Jessberger, S. Formation and integration of new neurons in the adult hippocampus. *Nature Reviews Neuroscience* **22**, 223-236 (2021).

48 Le Duigou, C., Simonnet, J., Teleñczuk, M. T., Fricker, D. & Miles, R. Recurrent synapses and circuits in the CA3 region of the hippocampus: an associative network. *Frontiers in cellular neuroscience* **7**, 262 (2014).

49 Song, D. *et al.* The lateralization of left hippocampal CA3 during the retrieval of spatial working memory. *Nature communications* **11**, 2901 (2020).

50 Dong, C., Madar, A.D. & Sheffield, M.E.J. Distinct place cell dynamics in CA1 and CA3 encode experience in new environments. *Nature communications* **12**, 2977 (2021).

51 Harris, K. D. & Shepherd, G. M. The neocortical circuit: themes and variations. *Nature neuroscience* **18**, 170-181 (2015).

52 Mitchell, A. S. *et al.* Advances in understanding mechanisms of thalamic relays in cognition and behavior. *Journal of Neuroscience* **34**, 15340-15346 (2014).

53 Phillips, J. W. *et al.* A repeated molecular architecture across thalamic pathways. *Nature neuroscience* **22**, 1925-1935 (2019).

54 Gómez-Ocádiz, R. *et al.* A synaptic signal for novelty processing in the hippocampus. *Nature Communications* **13**, 4122 (2022).

55 Zemla, R. & Basu, J. Hippocampal function in rodents. *Current opinion in neurobiology* **43**, 187-197 (2017).

56 Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207-214 (2014).

57 Ren, P. *et al.* Systematic Benchmarking of High-Throughput Subcellular Spatial Transcriptomics Platforms. *bioRxiv*, 2024.2012. 2023.630033 (2024).

58 Dougherty, M. K. *et al.* KSR2 is a calcineurin substrate that promotes ERK cascade activation in response to calcium signals. *Molecular cell* **34**, 652-662 (2009).

59 Rosenberg, M. M., Yang, F., Mohn, J. L., Storer, E. K. & Jacob, M. H. The postsynaptic adenomatous polyposis coli (APC) multiprotein complex is required for localizing neuroligin and neurexin to neuronal nicotinic synapses in vivo. *Journal of Neuroscience* **30**, 11073-11085 (2010).

60 Zeng, H. *et al.* Integrative in situ mapping of single-cell transcriptional states and tissue histopathology in a mouse model of Alzheimer's disease. *Nature Neuroscience* **26**, 430-446 (2023).

61 Sasaguri, H. *et al.* Recent advances in the modeling of Alzheimer's disease. *Frontiers in Neuroscience* **16**, 807473 (2022).

62 Jacobsen, J. S. *et al.* Early-onset behavioral and synaptic deficits in a mouse model of Alzheimer's disease. *Proceedings of the National Academy of Sciences* **103**, 5161-5166 (2006).

63 Presumey, J., Bialas, A. R. & Carroll, M. C. Complement system in neural synapse elimination in development and disease. *Advances in immunology* **135**, 53-79 (2017).

64 Matsuda, K. & Yuzaki, M. Cbln family proteins promote synapse formation by regulating distinct neurexin signaling pathways in various brain regions. *European Journal of Neuroscience* **33**, 1447-1461 (2011).

65 Ester, M., Kriegel, H. P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD* **96**, 226-231 (1996).

66 Hu, J. et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature Methods* **18**, 1342-1351 (2021).

67 Diggle, P. J. *Statistical Analysis of Spatial Point Patterns. 2nd edn.* (Hodder Education, 2003).