Rapid and Signal Crowdedness-Robust In-Situ Sequencing through Hybrid Block Coding

Tianyi Chang,[1,2,3,#] Wuji Han,[1,4,#] Mengcheng Jiang,[1,5,#] Jizhou Li,[6,#] Jie Shen,[7] Zitian Chen,[1] Peng Fei,[8] Xianwen Ren,[9] Yuhong Pang,[1] Guanbo Wang,[1,10] Jianbin Wang,[6,*] and Yanyi Huang[1,2,5,10,11,*]

[1] Biomedical Pioneering Innovation Center (BIOPIC), Peking University, Beijing 100871, China
[2] Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China
[3] Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China
[4] Yuanpei College, Peking University, Beijing 100871, China
[5] Materials Science and Engineering, College of Engineering, Peking University, Beijing 100871, China
[6] School of Life Sciences, Tsinghua University, Beijing 100084, China
[7] School of Basic Medical Sciences, Beijing Key Laboratory of Neural Regeneration and Repair, Capital Medical University, Beijing 100069, China
[8] School of Optical and Electronic Information - Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China
[9] Changping Laboratory, Yard 28, Science Park Rd., Changping District, Beijing, China
[10] Institute for Cell Analysis, Shenzhen Bay Laboratory, Guangdong 528107, China
[11] College of Chemistry and Molecular Engineering, Beijing National Laboratory for Molecular Sciences, Peking University, Beijing 100871, China

* corresponding authors. E-mail: yanyi@pku.edu.cn (Y.H.) and jianbinwang@tsinghua.edu.cn (J.W.)
[#] equal contribution.

Abstract

Technology development in spatial transcriptomics has propelled further understanding of cell types and their organization in tissues, opening new possibilities for researchers to explore transcript distributions at subcellular levels. However, current methods have limitations in either resolution, sensitivity, or speed. Here we demonstrate SPRINTseq (Spatially Resolved and signal-diluted Next-generation Targeted sequencing), a new in situ sequencing strategy that combines hybrid block coding and molecular dilution to resolve molecular crowdedness, overcoming a fundamental challenge that precluded simultaneous fast and sensitive high-resolution data acquisition. Our method recovered over 142 million transcripts in 453,843 cells from four whole mouse brain slices in less than two days. We then show how this high-resolution transcript map can be used to decipher cellular and subcellular molecular architecture of Alzheimer's disease in an unprecedented way. This improved spatial transcriptomics technology allowed us to link abnormal cellular behaviors and their organization in diseased tissue to changes in their subcellular mRNA distribution, generating new biological insights in health and disease.

Introduction

The well-ordered structural organization of cells and the intrinsic heterogeneity among cells are essential characteristics of multi-cellular organisms. Various spatial transcriptomic technologies have been developed to help understand the nature of such spatial information and its functional properties. In-situ mRNA capture-based approaches offer a spatially-resolved expression landscape [1-5]. Nevertheless, these methods have trouble providing finer information in space, such as subcellular RNA localization, whereas imaging-based spatial transcriptomic approaches offer near-optical limit resolution [6]. An ideal spatial transcriptomic technology needs to be sensitive, accurate, scalable, and robust. Such currently available approaches are divided into two major categories, fluorescent in-situ hybridization (FISH) [7-11] and in-situ sequencing (ISS) [12-16]. Those prevalent approaches face major challenges due to time-consuming workflows, complex reagents, and complicated imaging set-ups. FISH-based approaches are typically built upon single molecule detection schemes that rely on high-power, large NA objective lens and on specialized imaging techniques to overcome a low signal-to-background ratio (SBR) problem, and that usually leads to a sacrifice in the size of imaging field of view (FoV). The length of the mRNA of interest is also limited since many probe binding sites are commonly needed. While ISS-based methods usually apply in-situ amplification to increase the SBR ratio, molecular crowdedness during amplification or optical imaging are major challenges. Another common issue of currently available methods, the long experimental time, greatly hinders the scalability needed to handle a high-quantity of or large-size samples.

Here, we introduce a new ISS-based technology that uses sequencing-by-synthesis (SBS) chemistry to speed up the cyclic reactions needed for highly efficient information acquisition. We created a hybrid block code that is both signal-crowdedness and error robust, and that has a high decoding efficiency. Physical dilution of signal is adopted to further counteract ISS's intrinsic crowdedness issue. This method, SPatially Resolved and signal-diluted Next-generation Targeted sequencing (SPRINTseq), has greatly shortened the sequencing time it took (to within 5 hours) to profile a targeted transcriptome of a whole mouse brain slice, and produced near optical diffraction-limit resolution. SPRINTseq produced a sub-micron precise, whole-slice-scale cellular atlas that contains subcellular location information for each transcript. Such an information-rich sub-cellular distribution of genes can be greatly affected by physiological conditions. Using a 108-gene panel, 4 slices of whole mouse brain (from 2 normal mouse and 2 mouse with Alzheimer's disease) from sample to data could be profiled in 2 days, covering 453,843 cells and 142,957,485 transcripts. We found that the degree of subcellular mRNA dispersion increased as glia cells activated in Alzheimer's disease. We also found that the mRNA distribution produced changes in orientation within the amyloid microenvironment. Additionally, the high heterogeneity we found within inhibitory neurons may correlate with cell-type-specific responses during disease.

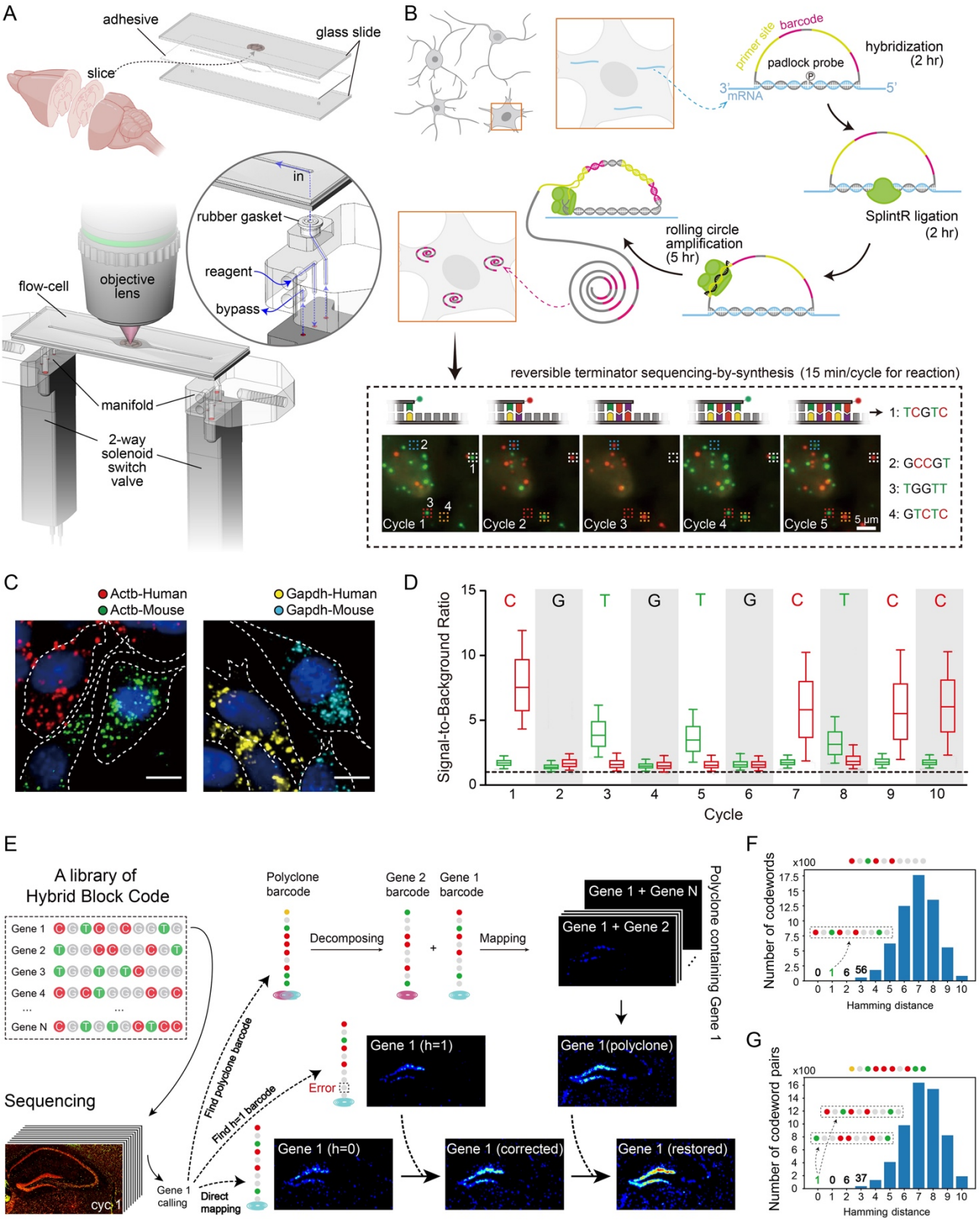

Fig. 1. Workflow and performance of SPRINTseq: in-situ sequencing from sample to data within 15 hours. (A) In-situ sequencing automation. Each mouse brain tissue slice was mounted on a slide and each slide was assembled into a flow cell and mounted between 2 manifolds. The manifolds were controlled by 2-way solenoid switch valves that controlled the fluidic routes. The reagent flowed directly into the flow cell while a bypass prevented reagent cross-contamination. **(B)** Schematic workflow: in situ barcoded clonal amplification and in situ sequencing. Probe hybridization, ligation, and rolling circle amplification were performed sequentially and took about 10 hours total. Each probe contained two identical "sequencing primer + barcode" blocks to increase the signal-to-background ratio. Barcodes were sequenced in situ using 2-color, reversible terminator sequencing-by-synthesis chemistry. Scale bar: 5 µm. **(C)** SPRINTseq specificity demonstrated by inter-species gene detection on a human (HEK293T) and mouse cell (3T3) co-culture system. Scale bar: 10 µm. **(D)** Characterization of 10-bp barcode sequencing. No obvious dephasing or signal decay was observed throughout 10 cycles. The boxes show the interquartile range, the lines in the boxes are the medians, and the bars show min/max values. The horizontal dashed line is SBR = 1. **(E)** Hybrid block codes achieve error-correctable and crowdedness-robust encoding. The barcodes are highly orthogonal between each other. A specific gene is mapped by looking for its exact barcode sequence and the one with correctable error (e.g., Hamming distance [h] = 1). Also, all composed barcodes (composite codewords)

containing that gene are corrected for mapping. By counteracting signal overlapping and sequencing error, a significant number of signals are rescued by this encoding strategy. **(F)** Error correction. When a sample readout sequence "CGTCGCGGGG" (one base error) is aligned with all barcodes (including an original barcode and a composed barcode) in the library, only one barcode (CGTCGCGGTG, Gene 1) with a Hamming distance equal to 1 is found, thus this sequence was mapped to Gene 1 and the error was corrected. **(G)** Polyclone barcode decomposition. When a sample readout sequence "AGTCCCGCTT" (a composed barcode containing two barcodes) is aligned with all the composed barcodes in the library, only one barcode pair (Gene 1: CGTCGCGGTG and Gene 2: TGGCCGGCGT) with a Hamming distance less than or equal to 1 exists. Thus, this polyclone (overlapping amplicons) can be decomposed and mapped to Gene 1 and Gene 2.

Result

Principle, operation, and performance of SPRINTseq.

A SPRINTseq experiment consists of two major parts, in situ barcoded clonal amplification and ISS. We constructed an integrated setup to automatically control fluidics, temperature, mechanical motion, and fluorescence imaging (Fig. 1A, Fig. S1 & S2). A slice of tissue was placed in a microfluidic flow-cell and various reagents were programmed to flow through it. A set of padlock probes we designed directly hybridized mRNAs with high specificity and then ligated to form a circular DNA template for rolling circle amplification (Fig. 1B). In about 10 hours, targeted genes were converted in situ into individual nanoball clones of corresponding barcodes. Each padlock probe contained two identical barcode blocks to double the signal of the sequencing reaction (Fig. S3), which came from fluorescent-labeled, reversible terminator nucleotide substrates. Each barcode block consisted of a gene-specific barcode and a universal sequencing primer binding sequence.

Because of its fast cyclic reaction time (Fig. S2; 15 min/cycle includes tissue blocking, substrate incorporation, and fluorophore cleavage steps), we used mature 2-color reversible terminator SBS chemistry for barcode readout (Fig. 1B). With an approximate 2.2 x 10^7 pixel/s imaging speed, the sequencing process took only about 5 hours for a 10-base barcode reading on a slice of whole mouse brain. The images were further processed for base calling and cell segmentation (Fig. S4 & S5). Common imaging background noise caused by non-specific binding of the fluorescent nucleotide substrate was clearly reduced when free thiol groups in the tissue were blocked (Fig. S3A,B). Also, with multiple-barcode-block design (Fig. S3C-F) the SBR of SPRINTseq was significantly greater than those FISH-based methods (Fig. S6). Amplified barcode clones were stably attached on the sample and drifted negligibly during the whole experimental process (Methods). Using only one probe per gene, targeting sensitivity was as high as 38% (Fig. S6D-F), mainly due to a low loss of amplified clones and no reverse transcription step in the protocol. In addition, during cyclic sequencing reactions, SPRINTseq exhibited high specificity (Fig. 1C) with very low signal decay or dephasing (Fig. 1D, Fig. S7).

We designed a highly efficient and orthogonal coding scheme, 'hybrid block code', that encodes barcodes in a way that tolerates signal crowdedness and corrects sequencing errors (Fig. 1E). We encoded 2^k genes using n-bit (n > k) codewords created from 0/1 string-based block code, and used the redundant bits to correct errors (Fig. S8). Moreover, we folded n-bit binary codewords to (n/2)-bit double-binary codewords by combining every two bits into a dual channel bit, thus shortening signal acquisition time. The dual-channel "1" signal at the same bit was avoided to reduce the overall

signal density and to improve orthogonality between codewords. We then selected a subset of all codewords, among which almost any two could be uniquely split to enable the decomposition of overlapping signals. Hybrid block code naturally fits 2-color reversible terminator SBS. To convert the codeword into a DNA barcode sequence, we used the bases ‘C’ and ‘T’ to represent “1” in the two channels and base ‘G’ to represent “0”. Base ‘A’ was not used in the barcodes. After sequencing, all barcode reads were aligned against a library consisting of the original barcodes and all composed barcodes (e.g., CGG + TCG = ACG). When aligning a barcode that had one sequencing error against the library, only one barcode with a Hamming distance equal to 1 was obtained, thus achieving error correction (Fig.1 F). Polyclone (overlapping amplicons) barcodes were aligned to all possible polyclone barcodes (all barcode pairs) in the library. Thus, a unique polyclone barcode with a Hamming distance equal to or less than 1 was found, and that allowed us to decompose the signal from overlapping amplicons (Fig.1 G). Using barcodes called from an image sequence built with that design resulted in a significantly higher average Shannon entropy per image than those of existing methods (Table S3). Our refined experimental process and efficient coding scheme enables SPRINTseq to profile a slice of whole mouse brain within 15 hours and with high sensitivity and accuracy.

Relieving crowdedness issues yields high quality ISS

Signal crowdedness, a major challenge for ISS, occurs when numerous amplified signal spots fill the limited space within a cell and thus limit the total read count. In a swift scanning scheme, the locations of amplified clones after 2D-stacking projection are prone to signal overlap in crowded regions. In hybrid block code, though, signals are diluted by using the ‘bit-of-silence’ and ‘optical dilution’ is realized. Only a fraction of the barcodes was fluorescently labeled in each sequencing cycle (Fig. S8C). Such signal sparseness can be finetuned by the number and position of the ‘0’ signal in the barcodes, thus staggering the signals from highly expressed genes in different cycles.

However, the optical dilution ratio has an upper limit within a certain number of sequencing cycles as the proportion of silent channels is limited by the Hamming weights of codewords. Inevitably, signal co-occurrence between highly-expressed genes will happen as more genes that need to be identified are included in the panel (Fig. 2A). In some cases, a signal from one super-highly expressed gene is so dense that the spots form ‘plaques’ and the signals from other genes are covered and their information is unextractable. This crowdedness is caused by too many amplicons from highly abundant genes. Also, some amplification events may be precluded in such a physically crowded environment, thus causing an even greater bias in gene profiling. Optical dilution does not alleviate this issue.

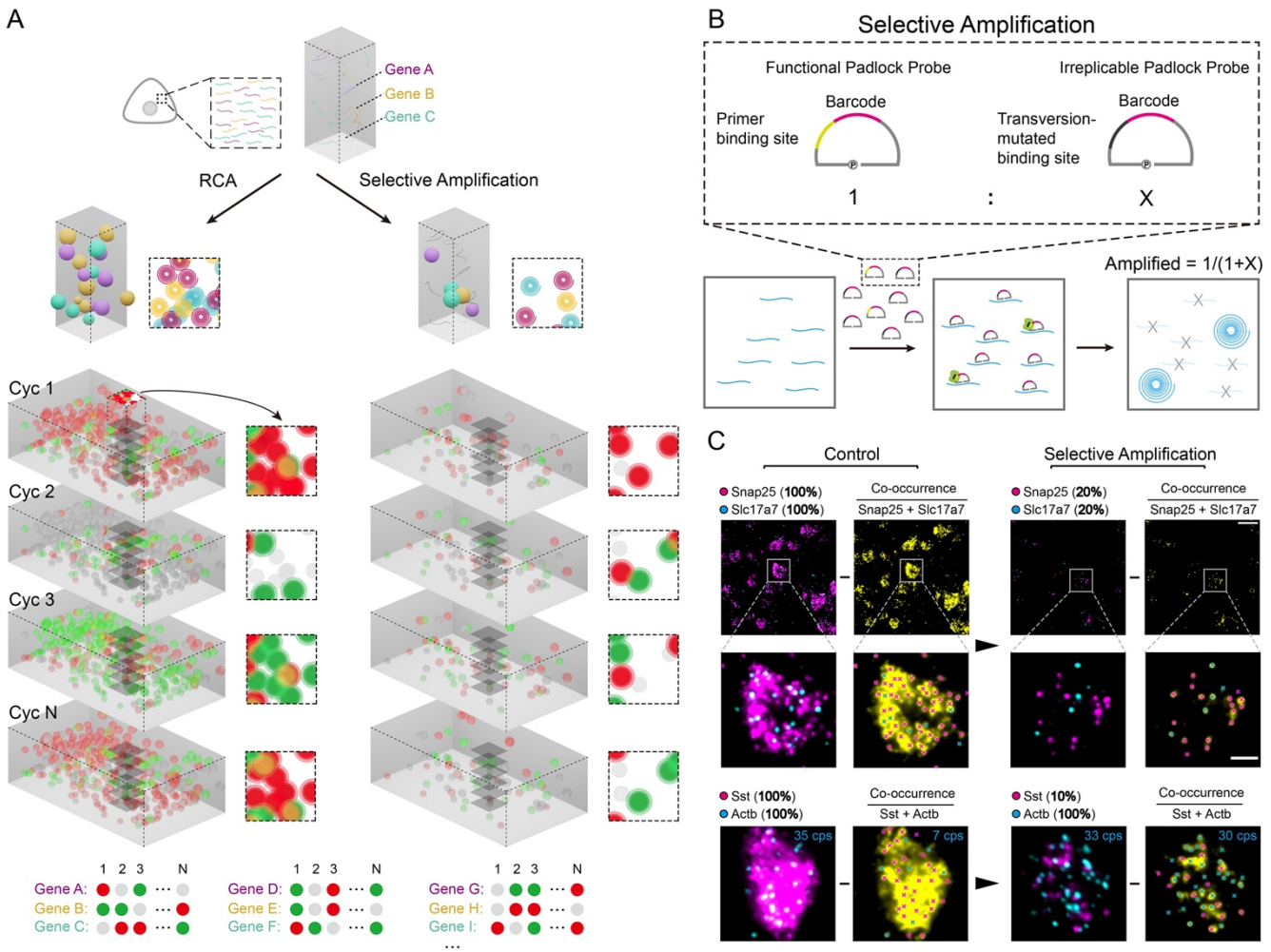

Fig. 2. SPRINTseq uses signal dilution to relieve signal crowdedness. (A) Spatial crowdedness and physical dilution of signals. As signals from rolling circle amplified (RCA) clones were projected onto a 2-D plane during focal-stacking, the signals inevitably overlapped in crowded regions and became unrecognizable, thus causing information dropout. Through selective amplification, signals from partial transcripts were physically eliminated, thus leaving a sparser environment. (**B)** Selective amplification. The proportion of amplification of specific highly expressed genes can be controlled by mixing normal padlock probes and sequencing-primer binding-site transversion probes (irreplicable), which achieve selective gene masking during amplification and yields a desired ratio. **(C)** Signal rescue through selective amplification. Respective signals from Snap25/Sst (magenta) and Slc17a7/Actb (cyan) can be obtained through another sequencing cycle. Because at least one base of the respective gene barcodes is designed to be the same (cycle and channel),a crowdedness situation (co-occurrence, yellow) is simulated. Percentages represent the remaining fractions of gene signals after selective amplification. The crosses represent the recognizable signals based on local maxima identification. In the example image, 20% (7/35) of Actb reads can be extracted during co-occurence with all Sst amplicons, while 91% (30/33) of the reads can be extracted during co-occurrence after masking 90% of the Sst amplicons. Scale bars: 20 μm for the base images and 5 μm for the insets.

We used a selective amplification strategy to relieve physical crowdedness. For those highly expressed transcripts, we doped the padlock probes with irreplicable ones to uniformly dilute their clonal populations. Thus, partial transcripts of such genes are physically eliminated and other genes are able to be better amplified and read (Fig. 2B). To demonstrate selective amplification, we simulated an inevitable situation in which multiple genes are tested in a panel: two genes with dense signals appear in the same reaction cycle and channel. Snap25 and Slc17a7 are highly expressed genes in excitatory neurons and when their signals appeared together, their clonal puncta joined together to form plaques in the images and the puncta were obscured (Fig. 2C). After masking 80% of both genes, each of their transcript clones were clearly differentiated. Though the masked signals' locations could not be recovered, the quantities of transcripts were effectively restored by multiplying the dilution ratios for further analysis (Fig. S9). In another example, Sst, a GABAergic neuron

subtype marker gene, dominated the expression in that cell type. The puncta of clonal signals from other genes such as Actb were difficult to discern. After diluting the Sst signal 10-fold, it was possible to digitally count mRNA clones, and many signals, including those of Actb and Sst, were rescued. Only 20% (7/35) of the Actb reads could be extracted when the Actb signal co-appeared with Sst's. After masking 90% of the Sst amplicon, more than 90% (30/33) of the Actb reads were extracted. Given prior knowledge, we can adjust the physical dilution ratios for genes with different abundances. This ratio needs to be high enough to identify clonal signals in highly expressed cells, but not so high as to cause false-zeroes in lowly expressed cells. Optical dilution and selective amplification are functionally complementary. The final dilution fold is the product of the physical dilution fold from selective amplification and the optical dilution fold.

For regions that are still crowded after two-level dilution, the polyclonal signals that result mainly from the co-localized 2-D projections of two amplicons, can be decomposed into separate barcodes in hybrid block coding and their signals can be rescued (Fig. 1E,G). By combining optical dilution, physical dilution, and polyclone decomposition, SPRINTseq effectively relieves the signal crowdedness issue in ISS.

Single cell profiling with subcellular resolution of whole mouse brain slices

We profiled whole mouse brain slices (approximately 10.2 x 7.6 mm^2) at subcellular resolution using a 108-gene panel (Fig. 3A). The genes included both marker and disease-related genes selected for cell classification and status characterization (Extra Table 1) [17]. To encode these genes, we designed a 10-base barcode set based on the hybrid block code (Fig. S4A-C). The minimum Hamming distance between any barcode pair in the 108-gene panel was 3, and the minimal Hamming distance between almost any (> 99.8%) barcode pairs that containing polyclone barcode was 3, which ensured error correction and polyclone decomposition. The silent bit (G) content of 52.4%, achieved a 23.8% optical dilution ratio while keeping a sufficient Hamming weight in codewords. Furthermore, selective amplification applied to seven highly-expressed genes (Actb, Mbp, Cst3, Penk, Snap25, Slc17a7, and Sst) further reduced the clone number by 54.4%. As a result, about 10% of all signals were acquired per sequencing cycle, significantly diluting crowdedness. While 49.8% of all monoclonal reads were rescued from error correction, 31.6% of the total reads were rescued from polyclone decomposition.

We identified 16,606,784 raw reads from the whole brain slice, and that number increased to 36,414,561 after selective amplification was restored. Probe binding false positives and gene mapping were 0.05 and 0.02 events per cell, respectively. The border of each general brain region was drawn using corresponding marker genes (Fig. 3A). Highly accurate and specific expression patterns were consistent with the in-situ hybridization results in the Allen Brain Atlas (Extra Supplement PDF). Brain replicates also showed high concordance at both whole and regional brain levels (Fig. 3B, Fig. S10).

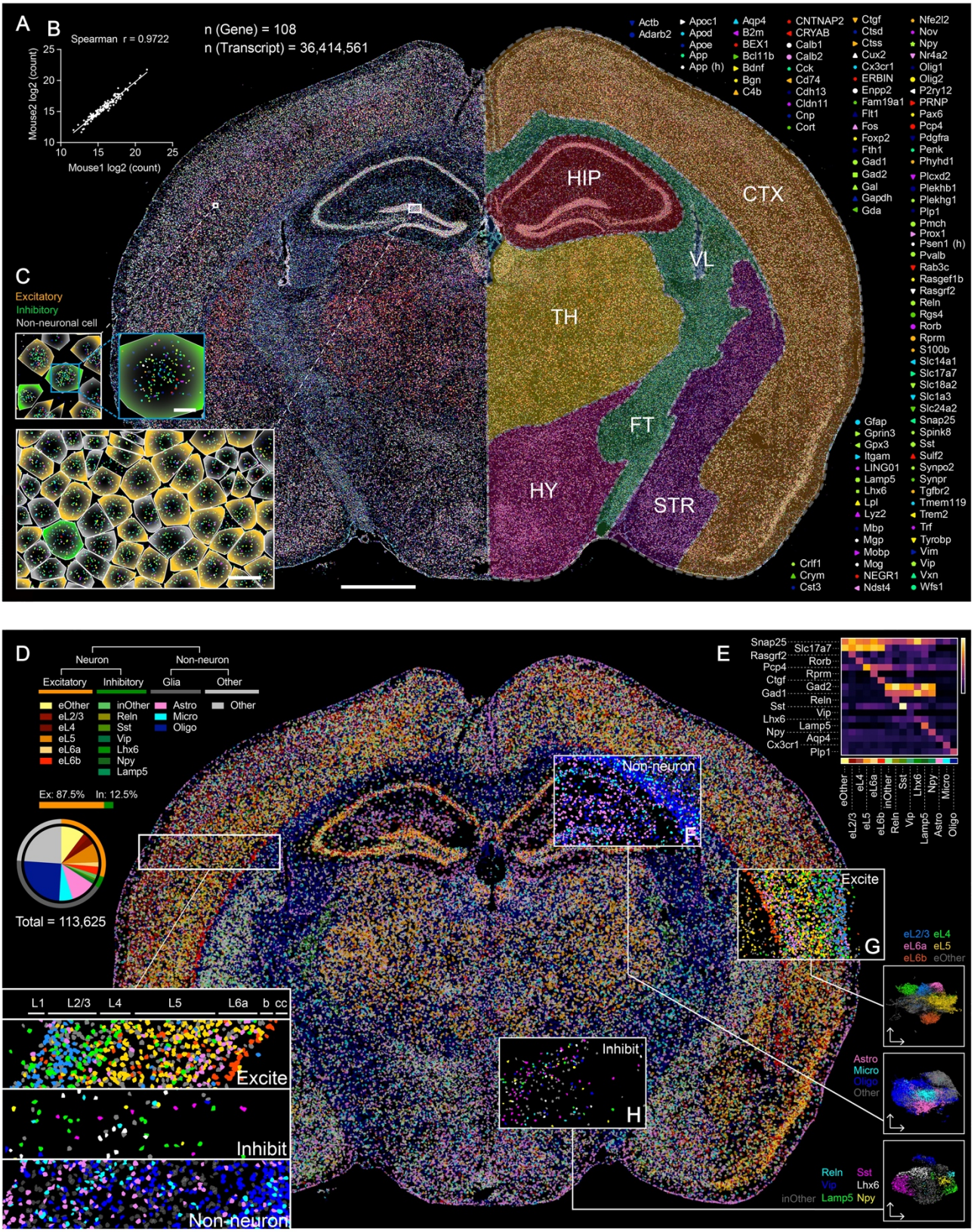

Fig. 3. Single cell profiling with subcellular resolution of whole mouse brain slices. (A) The spatial landscape of 108 genes called by 10-cycle sequencing in mouse brain. Brain regions shown in the right half, the brain was defined by several regional marker genes (Slc17a7, Gad2, Fth1, Enpp2, Pcp4, Pmch), CTX: cortex, HIP: hippocampus, VL: lateral ventricle, TH: thalamus, FT: fiber tract, HY: hypothalamus, STR: striatum. Scale bar: 1 mm. **(B)** Spearman correlation of two mouse sequencing replicates, the brain slices were selected at the same position and the gene panel was the same. **(C)** Enlarged images showing gene locations (dots color/shape-coded to the genes) and cells from the insets in **(A)**. Cell border was showed by lines with different colors: orange for excitatory neuron, green for inhibitory neuron and gray for non-neuronal cell. Scale bar: 2 μm for the single cell image and 10 μm for the multiple cell images. **(D)** Cell composition and spatial projection of a whole brain slice. Marker gene expression levels and Louvain shared nearest neighbor clustering identified 17 cell types (16 defined cell types and 1 other cell type shown in upper left corner), and their proportions are shown in the pie chart. Each color represents one unique cell type. The enlarged images at the bottom left show projections of all 17 cell types at the same locations in the cortex. The anatomical structure is labeled on top beginning with L1 (L: layer in the cortex, cc: corpus callosum) **(E)** Marker gene expression heatmap by each cell type. The cell type color code is the same

as in **(D)**. **(F-H)** Spatial projections of non-neuronal cells (astrocytes, microglia, and oligodendrocytes), excitatory neurons (eL2/3, eL4, eL5, eL6a, eL6b and others), and inhibitory neurons (Reln, Sst, Vip, Lhx6, Lamp5, Npy, and others), as well as their cell clustering visualizations through uniform manifold approximation plots.

We determined cell nuclei centroids using DAPI staining, calculated cellular segmentation with nucleic images, and assigned sequenced RNA reads to their nearest nucleus centroid (Fig. 3C, Fig. S5). We obtained 146,137 cells in total and 113,625 (77.8%) passed quality-control. According to the major marker genes, cells were categorized into three major populations: excitatory neurons (Slc17a7+), inhibitory neurons (Gad1/Gad2+), and non-neuronal cells. Each population was then subdivided into detailed clusters through Louvain shared nearest neighbor clustering and the resulting 17 types were spatially projected to their original positions (Fig. 3 D & E, Fig. S11). Non-neuronal cells were divided into astrocyte, microglia, oligodendrocyte, and other cell types (Fig. 3F). Astrocytes and microglia are scattered across the whole brain, but astrocytes are more frequently distributed at tissue edges and the hippocampal region. Oligodendrocytes have a high density in the fiber tract region. Neuronal cells were classified according to their positions in the cortex (for excitatory neurons, Fig. 3G) and their subtypes (for inhibitory neurons, Fig. 3H). Excitatory neurons in the cortex can be clearly divided into six layers. They are also widely distributed in the hippocampus and thalamus. Inhibitory neurons are sparsely distributed in the cortex, striatum, hippocampus, hypothalamus, and reticular nucleus region of the thalamus.

SPRINTseq offers an informative subcellular distribution and location of genes, as well as their correlations to other spatially distributed components. We assessed the degree of the mRNAs' subcellular dispersity using the average distances to their centroids, and also calculated the mRNA average distance to the nucleus centroid (Fig. S12). We then used those two parameters as coordinates for each gene in the panel and classified them all into one of three quadrants.

The mRNAs of the genes in the first quadrant were widely dispersed within the cell and far from the nucleus. From that we inferred that they were diffused throughout the cytoplasm and expressed the protein needed inside the cell. Actb is one typical example of this type. The mRNAs of the genes in the third quadrant, such as Ctgf, were likely to be distributed on the endoplasmic reticulum near the nucleus and to express membrane proteins or secreted protein. Genes appearing in the fourth quadrant showed a certain polarity in their mRNA distributions; genes including Snap25 and Slc17a7 were in this category. The mRNA distribution polarity of signal-transduction-associated genes may be related to the polarity of neurons and other cells in the brain. No genes appeared in the second quadrant.

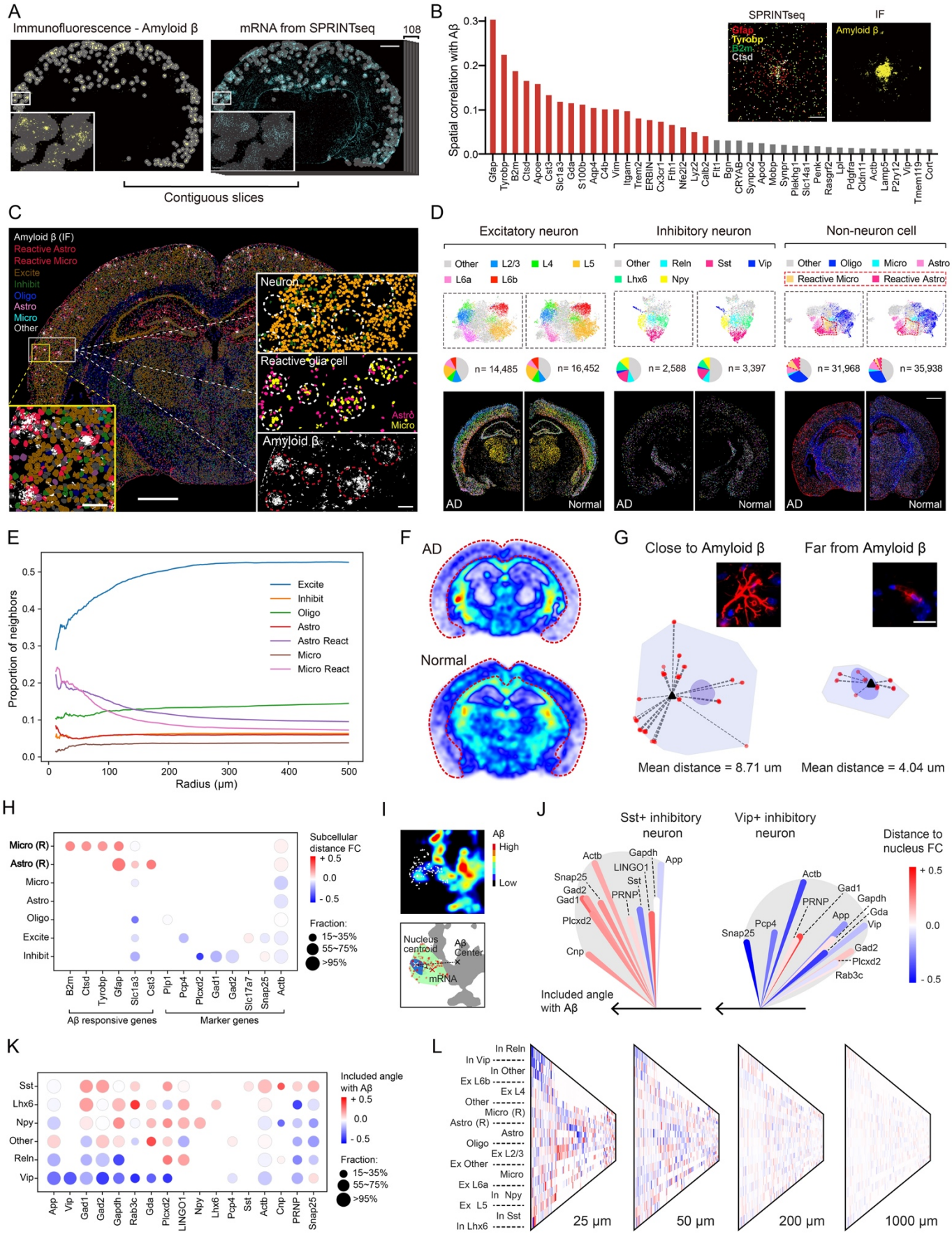

Fig. 4. Spatially-resolved cellular and subcellular changes of mouse brain with Alzheimer's disease (AD). (A) Immunostained amyloid plaque (amyloid β) and SPRINTseq analysis on contiguous slices. After alignment, the amyloid plaque positions were localized on the sequencing image. The 'Pan-plaque region' (regions circled in white) was defined around amyloid, and spatial correlation was calculated between all 108 genes and the amyloid plaque, respectively, in the pan-plaque regions. Bin size: 50 x 50 μm. Scale bars: 80 μm for enlarged images and 1 mm for brain slice images. **(B)** The most enriched genes, as shown by spatial

correlation analysis. The top 20 genes in the graph are shown by red bars. The upper right images show the top 4 genes' (Gfap, Tyrobp, B2m, and Ctsd) actual aggregations (left, sequencing result) and at the same positions with amyloid plaque (right, immunofluorescence [IF], contiguous slice after alignment). Scale bar: 40 μm. **(C)** Whole brain spatial projection for general cell types. Amyloid plaque IF in a contiguous slice is shown along with excitatory and inhibitory neurons and non-neuronal cells: oligodendrocyte (Oligo), astrocyte (Astro), microglia (Micro), and other. Reactive glia cells (responding to amyloid plaque) are labeled in red (lower left image). The 3 enlarged images on the right show neuron and reactive glia cells and the amyloid plaque distributions at the same position. Scale bar: 1 mm for the whole brain image, 100 μm for lower left image and the multiple cell images on the right. **(D)** Normal and AD mouse brain direct cell composition comparison using combined classifications of the 16 cell types (see Fig. 3F-H). Scale bar: 1 mm **(E)** Relative cell density change as a function of distance to amyloid plaque. The measurements start at 11 μm. **(F)** Oligodendrocyte density comparison of AD and normal mouse brain. The cortex region is encircled by a red dotted line. Oligodendrocyte density decreased globally in the AD mouse brain. **(G)** Diagram of subcellular dispersion. The photos show reactive astrocytes with Gfap protein in red and DAPI in blue. IF results confirm that subcellular dispersion of reactive astrocytes differed based on their distances from plaque. In the corresponding diagrams, the red dots represent Gfap mRNA, the black triangles are Gfap mRNA centroids, the pale purple shows cell body, and the dark purple shows cell nucleus. Scale bar: 10 μm. **(H)** Heatmap of the degree of subcellular dispersion change. Y axis shows different general cell types as listed in **(C)**. The values are the subcellular distance of cells close to amyloid plaque divided by that of cells far from amyloid plaque. (R): reactive, FC: fold change. **(I)** Diagram of mRNA orientation characterization. In a given cell, orientation is calculated as the included angle formed by amyloid plaque, the nucleus centroid, and the mRNA. This value is calculated only within cells that are close to amyloid plaque. A smaller included angle might suggest a trend for mRNA to approach plaque. mRNA locations are white in the top image and red in the bottom diagram. **(J)** A combination of mRNA orientation and the change of distance to the nucleus centroid. The change of distance to a nucleus centroid was calculated as a subcellular dispersion change. This change is also calculated by the value of cells close to amyloid plaque divided by cells far from amyloid plaque. Major expressed genes in two cell types were chosen as examples. The length of the scalloped area represents the expression level of each gene (log2). **(K)** Heatmap of mRNA orientation in inhibitory neurons. Y axis shows different subtypes within inhibitory neuron. The genes listed are major expressed genes in inhibitory neurons. **(L)** The orientation pattern as a function of distance. Y axis shows different cell types. The orientation of mRNAs was profiled in the cells whose distance to the nearest Amyloid β was 25 μm, 50 μm, 200 μm and 1000 μm, respectively.

Spatial profiling of the mouse brain with Alzheimer's disease

The 108-gene panel included many genes associated with Alzheimer's disease (AD), which is also pathologically associated with the distinct spatial distribution of amyloid plaques (amyloid β) that accumulate in the cortex and hippocampus, especially during aging [18]. We used SPRINTseq to examine whole brain slices from 10-month-old APP/PS1 and normal mice. First, the genes' overall expression levels were similar between AD and normal mice at the whole brain slice and regional levels (Fig. S13A). Next, we investigated gene-expression changes around amyloid plaques, whose typical size was over 100 μm in diameter. We aligned the Amyloid β immunofluorescence images to our SPRINTseq images and defined the plaques and their surrounding areas as 'pan-plaque regions' (Fig. 4A). Spatial correlation analysis between all genes and Amyloid β in those regions showed that many genes in our panel were potentially responsive to Amyloid β (Fig. 4B). The most relative genes included Gfap, Tyrobp, Ctsd, B2m, and Apoe, which express mainly in reactive astrocytes and microglia cells [19,20]. The aggregation of those genes was confirmed by raw sequencing images (Fig. 4B), and the ranking of spatially-correlated genes were independent of the bin size selected for analysis (Fig. S13B).

At the cellular level, our SPRINTseq results showed that neurons were depleted from the Amyloid β while the reactive glia cells, including microglia and astrocytes, were aggregated around it (Fig. 4C). To directly compare the cell compositions of AD and normal brains, we combined both data sets and classified them together before projection (Fig. 4D, Fig. S13C). The proportion of reactive glia cells in the AD mouse brain was more than twice that in the normal mouse brain. For other cell types, although the difference in quantity between AD and normal brains was marginal, the local densities showed different spatial dependencies to Amyloid β and those changes were not random (Fig. 4E, Fig. S13D). Clearly, the reactive glia cells were clustered favorably around Amyloid β, within 200 μm, which reflected their roles of responding to microenvironmental changes and plaque clearance. The reduction of neuronal density near Amyloid β reflected neuron apoptosis in a plaque-rich

microenvironment. Notably, the pathological effect of AD is not only associated with the local microenvironment but can be scaled to larger regions. For example, oligodendrocytes did not show significant density changes around Amyloid β, but their presence in the cortex region was generally less in the AD than in the normal mouse brain (Fig. 4F, Fig. S13E).

SPRINTseq extended the analysis' spatial resolution to the subcellular level, enabling us to find many cells that showed distinct differences that were related to the cellular distance to Amyloid β. In reactive glia cells, the subcellular dispersion of mRNAs was increased in Amyloid β adjacent cells (distance threshold: 25 μm) than in more distant cells (Fig. 4G & H). This was largely because the cytoplasmic size of reactive glia cells near Amyloid β increased because of increased and elongated surface bumps. Disease-associated microglia (DAM) have been reported to play a vital role in AD and the function is conserved in mice and human [21,22]. Interestingly, the Amyloid β adjacent reactive microglia we found may largely correspond to reported plaque-phagocytic microglia (XO4+) in gene expression pattern (Tyrobp+, Apoe+, B2m+ and low Cx3cr1) and behavior (spatially enriched around plaques) [23]. The enlarged cell body might suggest their function in phagocytosis and other regulation process within the micro-environment, whereas XO4- microglia can't migrate towards plaques and thus are far away from plaques. The spatial information can be used as a new dimension for confirming plaque-phagocytic microglia.

Additionally, in a cell that is directly adjacent to a plaque, the location of one of its gene's mRNA orientations with respect to the nearest Amyloid β can characterize the relative proximity of that gene. So, we calculated the included angle between the mRNA's orientation, the nucleus centroid, and Amyloid β's direction. A smaller angle indicated that the mRNA tended to be closer to its nearest plaque (Fig. 4I). Changes in the distance to the nucleus can also be calculated and combined with the orientation calculation to better describe the mRNA's tendency to approach or retreat (Fig. 4J). Altogether, we found much heterogeneity within inhibitory neurons (Fig. 4K). In the Sst+ inhibitory neuron subtype, which is reported to be one of the most affected inhibitory neurons in AD [24,25], there was a general tendency for mRNAs to get close to plaques, while in other subtypes, such as Vip+/Reln+ inhibitory neurons, mRNAs were generally farther away from amyloid plaques. Orientation was dependent on the distance between a cell and amyloid (Fig. 4L, Fig. S13F). This orientation feature pattern disappeared at longer distances, as the effect of Amyloid β on distant cells became weaker. Altogether, Amyloid β might generally affect its nearby cell morphology and potentially alter sub-compartment architecture in various cell types.

While in-depth investigation is left to future studies, we validated the differential gene expression within Amyloid β microenvironments at a larger transcriptomic landscape. We analyzed an existing capture-based spatial transcriptomic dataset of 18-month AD mice [19,20]. Although the spatial resolution is limited in this reported dataset, we still observed that in cells close to Amyloid β, gene ontology (GO) terms including extracellular region (GO:0005576) and integral component of plasma membrane (GO:0005887) were enriched, which is consistent with our findings, and are possibly

linked to Amyloid β-induced functional and morphological alteration, such as cellular response to organic substance (GO:0071310) and biotic stimulus (GO: 0071216) (Fig. S13G-I) in glia cells.

Discussion

SPRINTseq is an intrinsically high-speed spatial sequencing method that can finish a whole mouse brain slice profile at subcellular resolution within 15 hours, including both sample preparation (10 hours, and several slices can be prepared in parallel) and cyclic sequencing (5 hours). Due to a combination of highly effective barcode coding and robust SBS chemistry, SPRINTseq is significantly faster than most other approaches. The scalability holds a great potential to be further improved by engineering optimization. Besides, we developed three orthogonal approaches to relieve the conventional ISS's signal crowdedness issue. Among those approaches, optical dilution holds great potential and deserves further examination. For example, with constant codeword Hamming weights, more silent bits can be gained by increasing the barcode length. Multiple padlocks with different barcodes can also be applied to detect more transcripts of highly expressed genes, if necessary, which resembles selective amplification but is lossless in spatial information.

One major advantage of this method is the single-molecule high spatial resolution that gives sub-cellular localization information. This is important for addressing questions of cellular interactions in the context of gradients or proximity to specific tissue features such as disease lesions. In mouse models of Alzheimer's disease, subpopulations of microglia with distinct transcriptomic phenotypes have been identified via single-cell RNA-sequencing [26]. These seemingly AD-associated microglia bearing stronger inflammatory signatures and were hypothesized to engulf Amyloid β plaques [26]. We described the distribution of different types of cells near Amyloid β in Alzheimer's disease mouse model, and we identified reactive glia cells with a distinct expression pattern that are responsive to Amyloid β, similar to what was previously reported [21,27]. The spatial information, including their distance to Amyloid β, can further confirm their functional and phenotypic diversity. In addition, subcellular mRNA distribution was also informative, showing specific transcript distribution patterns within the cell including polar, random, and centripetal distributions. Extensive changes in these subcellular distributions in glia cells and neurons are found within the Amyloid β microenvironment, most likely caused by changes in the cytoplasm morphology and membrane position, consistent with previous reports [26]. Cell functions including phagocytosis, stimulus sensing and response, and cell-to-cell interactions are closely linked with these structural alterations during disease development [23,26,28].

Subcellular mRNA distribution has high heterogeneity across cell types and subtypes. All reactive glia cells were classified as such based on their gene expression, but importantly the mRNA dispersion degree is higher in cells that are closer to plaques within the same type (the larger cytoplasm size is a typical phenotypic feature of glia cell activation in addition to specific gene expression). General mRNA orientation within inhibitory neurons around plaques also differs in various subtypes. This new facet of in-situ sequencing data has added one more dimension to the

conventional cellular-level gene expression matrix. We assert that a combination of gene expression and subcellular mRNA distribution will improve the understanding of complex transcriptional molecular functions in tissues and lead to more accurate and more quantitative characterizations of cell types and states.

The current encoding scheme has about 140 original codewords when fewer than 5% of all codeword pairs (including the original codeword and all composite codewords) have Hamming distances less than 3, a preferred condition for error correction. Information acquisition efficiency decreases when more codewords are used to encode a greater number of genes. To encode all 20,000 genes in the human genome, a simple solution is to splice two current 10-bp barcodes into a 20-bp barcode while maintaining the current dilution fold (140^2 ~ 20,000). However, when more genes are read, a large dilution fold is needed to maintain crowdedness-robust signal quality, which likely requires a longer barcode.

Acknowledgments

The authors thank BIOPIC sequencing center for experimental assistant. Funding was provided by National Natural Science Foundation of China Grant 21927802 (Y.H., J.W., P.F.), 22050002 (Y.H.) and T2188102 (Y.H.), Beijing Municipal Science and Technology Commission Grant Z211100003321006 (Y.H.), and Beijing Advanced Innovation Center for Genomics.

Author contributions

Conceptualization: Y.H. and J.W. Experiment: T.C., W.H., M.J., J.L., J.S., Z.C., Y.P., and G.W. Data analysis: T.C., W.H., M.J., P.F., X.R., and Y.H. Writing: T.C., W.H., M.J., J.W., and Y.H.

Competing interests

All authors declare no competing interests.

Supplementary Materials as a separate PDF file.

References

1. P. L. Ståhl *et al.*, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82 (2016).
2. s. g. Rodriques *et al.*, Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467 (2019).
3. S. Vickovic *et al.*, High-definition spatial transcriptomics for in situ tissue profiling. *Nature Methods* **16**, 987-990 (2019).
4. Y. Liu *et al.*, High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **183**, 1665-1681.e1618 (2020).
5. C. R. Merritt *et al.*, Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nature Biotechnology* **38**, 586-599 (2020).
6. M. Asp, J. Bergenstråhle, J. Lundeberg, Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. *Bioessays* **42**, e1900221 (2020).
7. E. Lubeck, A. F. Coskun, T. Zhiyentayev, M. Ahmad, L. Cai, Single-cell in situ RNA profiling by sequential

hybridization. *Nature Methods* **11**, 360-361 (2014).
8. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
9. S. Codeluppi *et al.*, Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature Methods* **15**, 932-935 (2018).
10. C. H. L. Eng *et al.*, Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235-239 (2019).
11. C. Xia, J. Fan, G. Emanuel, J. Hao, X. Zhuang, Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proceedings of the National Academy of Sciences* **116**, 19490-19499 (2019).
12. J. H. Lee *et al.*, Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* **343**, 1360-1363 (2014).
13. X. Chen, Y. C. Sun, G. M. Church, J. H. Lee, A. M. Zador, Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Research* **46**, e22 (2018).
14. X. Wang *et al.*, Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, doi: 10.1126/science.aat5691 (2018).
15. X. Y. Chen *et al.*, High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. *Cell* **179**, 772-786 (2019).
16. Y.-C. Sun *et al.*, Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. *Nature Neuroscience* **24**, 873-885 (2021).
17. B. Tasic *et al.*, Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience* **19**, 335-346 (2016).
18. D. S. Knopman *et al.*, Alzheimer disease. *Nature Reviews Disease Primers* **7**, 33 (2021).
19. H. Mathys *et al.*, Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337 (2019).
20. W.-T. Chen *et al.*, Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. *Cell* **182**, 976-991.e919 (2020).
21. H. Keren-Shaul *et al.*, A Unique Microglia Type Associated with Restricting Development of Alzheimer' s Disease. *Cell* **169**, 1276-1290.e1217 (2017).
22. S. Krasemann *et al.*, The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. *Immunity* **47**, 566-581.e569 (2017).
23. A. Grubman *et al.*, Transcriptional signature in microglia associated with Aβ plaque phagocytosis. *Nature Communications* **12**, 3015 (2021).
24. Y. L. Xu, M. N. Zhao, Y. Y. Han, H. Zhang, GABAergic Inhibitory Interneuron Deficits in Alzheimer's Disease: Implications for Treatment. *Frontiers in Neuroscience* **14**, 14:660 (2020).
25. Y.-H. Song, J. Yoon, S.-H. Lee, The role of neuropeptide somatostatin in the brain and its application in treating neurological disorders. *Experimental &amp$\mathsemicolon$ Molecular Medicine* **53**, 328-338 (2021).
26. H. Chun, C. J. Lee, Reactive astrocytes in Alzheimer' s disease: A double-edged sword. *Neuroscience Research* **126**, 44-52 (2018).
27. D. S. Bouvier *et al.*, High Resolution Dissection of Reactive Glial Nets in Alzheimer' s Disease. *Scientific Reports* **6**, 24544 (2016).
28. Y. Verdier, M. Zarándi, B. Penke, Amyloid beta-peptide interactions with neuronal and glial cell plasma membrane: binding sites and implications for Alzheimer's disease. *J Pept Sci* **10**, 229-248 (2004).
29. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359 (2012).
30. B. J. Beliveau *et al.*, OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proceedings of the National Academy of Sciences* **115**, (10) E2183-E2192 (2018).
31. A. Sountoulidis *et al.*, SCRINSHOT enables spatial mapping of cell states in tissue sections with single-cell resolution. *PLOS Biology* **18**, e3000675 (2020).

32. S. Liu *et al.*, Barcoded oligonucleotides ligated on RNA amplified for multiplexed and parallel in situ analyses. *Nucleic Acids Res* **49**, e58 (2021).
33. K. Smith *et al.*, CIDRE: an illumination-correction method for optical microscopy. *Nature Methods* **12**, 404-406 (2015).
34. S. van der Walt *et al.*, scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).
35. J. Chalfoun *et al.*, MIST: Accurate and Scalable Microscopy Image Stitching Tool with Stage Modeling and Error Minimization. *Scientific Reports* **7**, 4988 (2017).
36. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018).
37. E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).