Color-barcoded Super-multiplex RNA FISH in Three-dimensional Tissues through Single-round Imaging

Tianyi Chang,[1,#] Shihui Zhao,[1,#] Kunyue Deng,[1,2,#] Zhizhao Liao,[1,3,#] Mingchuan Tang,[1,4,#] Yanxi Zhu,[1] Wuji Han,[1,4] Chenxi Yu,[1] Wenyi Fan,[1] Mengcheng Jiang,[1] Guanbo Wang,[1,3,5] Dongfang Liu,[6] Jirun Peng,[6,7] Yuhong Pang,[1] Peng Fei,[8] Jianbin Wang,[3,9] Chunhong Zheng,[1,10,*] and Yanyi Huang[1,2,3,5,*]

[1] Biomedical Pioneering Innovation Center (BIOPIC), Peking University International Cancer Institute, School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China
[2] College of Chemistry and Molecular Engineering, Beijing National Laboratory for Molecular Sciences, Peking University, Beijing 100871, China
[3] Changping Laboratory, Yard 28, Science Park Rd., Changping District, Beijing, China
[4] Yuanpei College, Peking University, Beijing 100871, China
[5] Institute for Cell Analysis, Shenzhen Bay Laboratory, Shenzhen 518132, China
[6] Department of Surgery, Beijing Shijitan Hospital, Capital Medical University, Beijing 100038, China
[7] Ninth School of Clinical Medicine, Peking University, Beijing 100038, China
[8] School of Optical and Electronic Information – Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China
[9] School of Life Sciences, Tsinghua University, Beijing 100084, China
[10] State Key Laboratory of Molecular Oncology, Beijing Key Laboratory of Carcinogenesis and Translational Research, Peking University Cancer Hospital & Institute, Beijing 100142, China.

[#] These authors contributed equally to this work.
* Correspondance should be addressed to chunhong_zheng@bjmu.edu.cn (C.Z.) and yanyi@pku.edu.cn (Y.H.)

Abstract

To resolve diverse RNA species in situ, cyclic reactions are often required to increase multiplexity, as conventional fluorescence microscopy is limited to only a few channels. This typically necessitates sophisticated instrumentation for techniques such as in-situ sequencing or sequential fluorescence in-situ hybridization (FISH) imaging, which has hindered the widespread adoption of spatial RNA imaging in biological research. Here, we introduce 'Profiling of RNA In-situ through Single-round iMaging' (PRISM), a method that expands coding capacity by leveraging color intensity grading. Using a radius vector filtering strategy to ensure the distinguishability of codewords in color space, PRISM achieves up to 64-plex color-barcoded RNA imaging in a single imaging round with conventional microscopes. We validated PRISM's versatility across various tissues, including mouse brains, mouse embryos, and human hepatocellular carcinoma (HCC) samples, collectively annotating over 5.6 million cells. Through quasi-3D spatial landscapes, we tracked major cell types in multiple organs during embryonic development from E12.5 to E14.5. Additionally, we uncovered the critical role of cancer-associated fibroblasts (CAFs) in mediating immune infiltration and immune response heterogeneity within and between tumor microenvironments. PRISM was further applied to 100-μm-thick mouse brain slices, generating an accurate 3D cell atlas and subcellular RNA localization landscapes. Robust, user-friendly, and capable of fast turnaround times with sub-cellular resolution, PRISM provides biologists with a powerful super-multiplex FISH toolkit.

Introduction

The spatial arrangement of RNA transcripts is essential for revealing biological interactions at cellular and sub-cellular levels, exposing the diverse heterogeneity and structural compositions across various tissues. As demand for spatially resolved methods within the scientific community continues to grow[1–6], existing methods face significant challenges regarding accessibility due to complex instrumentation, high costs, and low throughput[7,8], particularly with imaging-based approaches[9]. Standard fluorescent microscopes, for instance, are generally limited to resolving no more than five channels due to spectral overlap[10,11]. Consequently, increasing multiplexity typically requires multiple rounds of labeling-stripping reactions or the addition of fluorescence channels in existing FISH-based methods[12–18] and in-situ sequencing approaches[19–27].

Although multiplexing potential can theoretically be increased exponentially, fluidics-dependent methods have limited widespread adoption because they rely on sophisticated, delicate equipment that combines fluidics, optics, and temperature regulation, which are typically costly and time-intensive[9]. Ensuring precise registration of signal spots across reaction cycles is further complicated by sample deformation or displacement over extended experiment durations. Additionally, three-dimensional RNA profiling presents unique challenges, such as slow reagent diffusion in thick tissues and the alignment of three-axis imaging in fluidics-dependent techniques.

Researchers are increasingly turning to fluidics-free methodologies. To exceed traditional channel limitations and encode more molecular species, some additional microscopic imaging modalities, such as super-resolution[28], fluorescence lifetime[29], and coherent Raman scattering microscopy[30], have been employed. Recently, pseudo-thermal-plex channels[31] and binding-affinity labeling[32] were utilized to expand multiplexity on traditional microscopes. While these efforts achieve higher coding capacities without fluidics, the improvement in multiplexing remains modest. Furthermore, the requirement for specialized microscopy setups and temperature control equipment restricts broader application in biological laboratories.

Here we present PRISM (**P**rofiling of **R**NA **I**n-situ through **S**ingle-round i**M**aging), a super-multiplex FISH method that employs a multi-channel color barcoding on rolling circle products to distinguish dozens of RNA transcript species in a single imaging round. As a fluidics-free approach, PRISM addresses accessibility challenges by requiring only a standard fluorescence microscope, with the entire workflow from sample preparation to data acquisition completed within a day. We tested PRISM across various tissue samples, including mouse brains, mouse embryos and human hepatocellular carcinoma (HCC) samples. Using a 30-gene panel, we developed a 3D cell atlas of mouse embryos at developmental stages E12.5, 13.5, and 14.5, encompassing over 4.2 million cells categorized into 12 types, showcasing cellular compositions and organizations in different organs at various developmental phases. Additionally, using a 31-gene panel on 20 consecutive sections from an HBV-positive patient, we created a quasi-3D HCC tumor-normal transition landscape with 1.2 million annotated cells across 32 cell types, uncovering heterogeneities in the tumor microenvironment and demonstrating how cancer-associated fibroblasts (CAFs) construct a physical barrier against immune infiltration. PRISM's compatibility with intact 3D imaging was further validated using 100-μm-thick mouse brain slices, where a 30-gene panel effectively identified over 20 cell types within the cortex, hippocampus, thalamus, and hypothalamus regions in a single staining round using conventional confocal microscopy. With a coding capacity reaching up to 64-plex, PRISM has been validated across different tissue types and thickness. It stands as a robust

and accessible super-multiplex FISH method for in-situ imaging of RNA transcripts, relying solely on standard laboratory equipment.

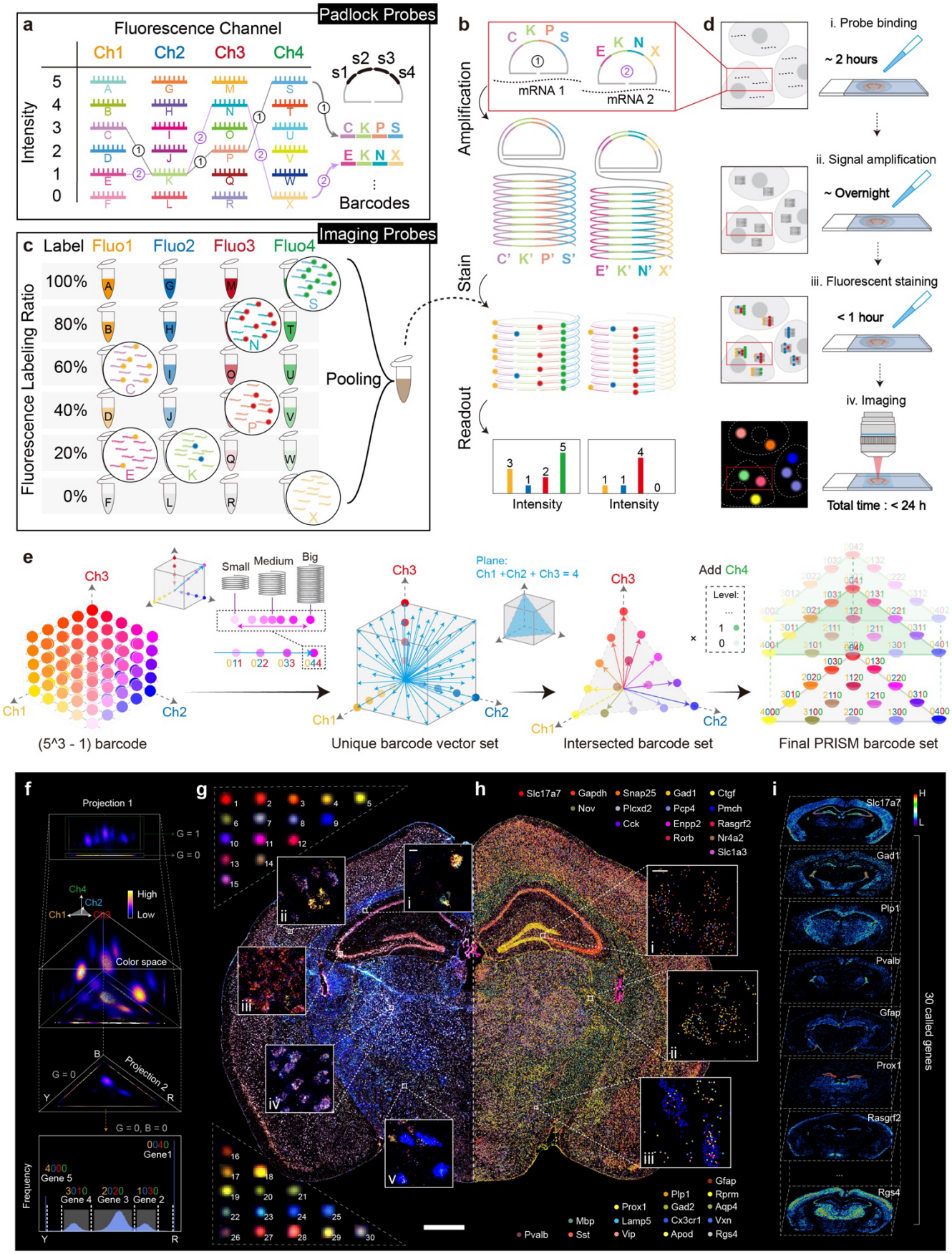

Fig. 1 | Workflow and principle of PRISM encoding. a, PRISM barcoding on padlock probes. Barcode on padlock probe consists of four segments, with each segment corresponding to one spectral channel. Each segment can accommodate various sequences

(denoted as A, B, C, … in Ch1), with each sequence represents a distinct intensity level. The barcode on padlock probe is a combination of different segment sequence (intensities) under each channel, e.g. C-K-P-S, signifies level 3 in Ch1, level 1 in Ch2, level 2 in Ch3 and level 5 in Ch4. Therefore, each barcode is spectroscopically unique, visualized as a color. **b**, Molecular events in the PRISM workflow. Gene transcripts are targeted using a set of padlock probes with unique barcodes. Following ligation, each padlock probe undergoes rolling circle amplification (RCA) to produce hundreds of identical barcode sequences. These barcodes are then read out in-situ through fluorescent staining and imaging. **c**, Decoding through staining with pooled imaging probes. Each kind of imaging probe is designed with sequences identical to each segment sequence, with two types: fluorescent labeled and unlabeled. Two types are pre-mixed in ratios according to the designed intensity level for that segment sequence. During staining, competitive binding ensures that the desired fraction of fluorophore binds to the rolling circle product, thereby exhibiting the intended intensity. All imaging probes are pooled together for one-pot staining. **d**, Experimental workflow of PRISM. Major steps include padlock probe binding, signal amplification (probe ligation and rolling circle amplification), fluorescent staining, and imaging. The entire experiment can be completed within 24 hours using standard equipment. **e**, Radius vector filtering to ensure ample barcode discriminability. Barcode candidates result from combinations of intensity levels and spectral channels. Due to the amplicon size heterogeneity during isothermal RCA, the absolute signal intensity varies, causing the spot from a certain barcode to shift along its radial ray in the color space, leading to indistinguishability. To address this, all barcodes were represented as their radius vector, forming a unique set of barcode vectors. Barcode vectors with small neighbor angles are difficult to distinguish. To maximize the dispersion of vectors, the barcode set intersected by plane (Ch1 + Ch2 + Ch3 = K) was selected. By quantizing three channels into quarters and setting K to 4, 15 highly distinguishable barcodes can be generated. The addition of a fourth channel, Ch4, multiplies the coding capacity by Ch4 quantities. For example, with a 2-intensity-level, which is 0 (No) or 1 (Yes), the total available barcodes are 30 or 31 (including only Ch4 itself). **f**, Gene decoding in color space. The signals from four colors were converted into three-dimensional coordinates, and each spot was plotted in a "color space." 30 clusters located in different positions, each representing a unique barcode. Each barcode was identified by its specific location within the color space. **g**, Spots in the raw image corresponded to 30 different barcodes. Each spot consists of four channels, shown within the same dynamic range. Scale bar: 10 μm, zoom-in image; 1 mm, below. **h**, Decoded genes. Each gene was assigned a pseudo-color corresponding to its barcode. Scale bar: 10 μm. **i**, Spatial distribution of 30 decoded genes.

Coding principles of PRISM

The PRISM workflow begins by coding the targeted gene transcripts using padlock probes, each containing multiple segments of specific sequences. The barcode of each probe is a combination of segments, with each segment corresponding to a fluorescence channel (color). Although the number of available channels is limited due to spectral indistinguishability among fluorophores, each channel can accommodate multiple unique segment sequences, each encoding specific intensity information (Fig. 1a). By using this spectral barcoding scheme, the coding capacity can greatly exceed the channel limit through the combination of segmental sequences.

Each transcript-bound padlock probe is then clonally amplified in situ via rolling circle amplification (RCA), creating a large number of identical barcodes for each clone (Fig. 1b). An array of imaging probes is also prepared, each containing a pair of identical single-stranded oligos, one with and one without fluorescent labels (Fig. 1c). The ratio of labeled to unlabeled twin-probes is carefully determined, and all probes are pooled together as a single staining mix. When this mix is applied to hybridize with the RCA amplicons in situ, each segment binds both labeled and unlabeled probes in the preset ratio, forming a spectral barcode for each clone through competitive hybridization. Quantitative assessment of fluorescence intensity across all channels in each clone then reveals its barcode (Fig. 1c,d).

Theoretically, an array with m channels and n intensity-level segments can provide a coding capacity of $n^m - 1$. For example, with three spectral channels, each quantized into quarters (0, 1/4, 2/4, 3/4, 4/4), it is possible to write codes with 3-digit quinary numbers, enabling encoding of up to $5^3 - 1 = 124$ genes. However, in practice, the number of usable barcodes is reduced due to variations in the sizes of amplified nanoball clones, which can cause certain codes indistinguishable in spectral color space (Fig. 1e). For instance, a small amplicon barcoded as (020) would be indistinguishable from a twice-sized amplicon barcoded as (010), as would (033) and (022).

To better describe barcode discriminability under these conditions, each barcode is conceptualized as a 'radius vector' in color space. Indistinguishable clones share the same vector direction but differ

in magnitude; therefore, only angularly distinct barcode vectors are considered legitimate (Fig. 1e, Extended Data Fig. 1). To ensure ample dispersion of barcode vectors, we constrained the coding space to the plane where the sum of intensities across three channels (Ch1+Ch2+Ch3) equals 4. This approach optimizes the angles between barcode vectors, minimizing crosstalk and maximizing the distinguishability of barcodes. As a result, with three quarterly quantized channels, we achieve 15 highly distinguishable barcodes.

With a 3-channel setup that minimizes amplification bias, the coding capacity can be further expanded by adding a fourth channel. In our demonstration, we selected common fluorophores, Texas Red (TR, Ch1), Alexa Fluor 488 (AF488, Ch2), Cy5 (Ch3), and Cy3 (Ch4), due to their wide availability. To address interference from tissue autofluorescence, the intensity level of Ch4 (Cy3) is limited to 2 or 3, effectively doubling or tripling the coding capacity to more than 30.

Using conventional equipment and a single round of staining, we conducted cell typing on a 10-μm-thick mouse brain coronal section with a 30-gene panel (Fig. 1f-i). Fluorescent puncta were processed to determine the spectral code of each amplicon, revealing the barcode of each transcript in situ (Extended Data Fig. 2). The barcodes formed into 30 distinct clusters in color space, accurately reflecting the original design. Clusters were identified using manually adjustable Gaussian fitting for probability prediction, achieving over 80% spot-preserving efficiency (Fig. 1f, Extended Data Fig. 3).

Mapping each barcode back onto the sample allowed for a submicron-resolution spatial distribution plot of 30 genes, with spatial patterns closely matching RNA FISH results from the Allen Brain Atlas (Fig. 1i, Extended Data Figs. 4, 5). The decoding accuracy was 94.6%, as verified by overlap rate in strip-rehybridization experiments (Extended Data Fig. 6). Cell segmentation and typing were completed by integrating 30 RNA spatial images with DAPI-stained nuclei territories (Extended Data Fig. 7). In this brain section, we identified 71,435 cells, categorizing them into five major types: excitatory neurons (39,757, 55.6%), inhibitory neurons (13,134, 18.4%), and non-neuronal cells, including oligodendrocytes (8,244, 11.5%), astrocytes (7,062, 9.9%) and microglia (3,238, 4.5%). Further differentiating into subtypes revealed the architectural organizations of the mouse brain at single-cell resolution (Extended Data Fig. 8).

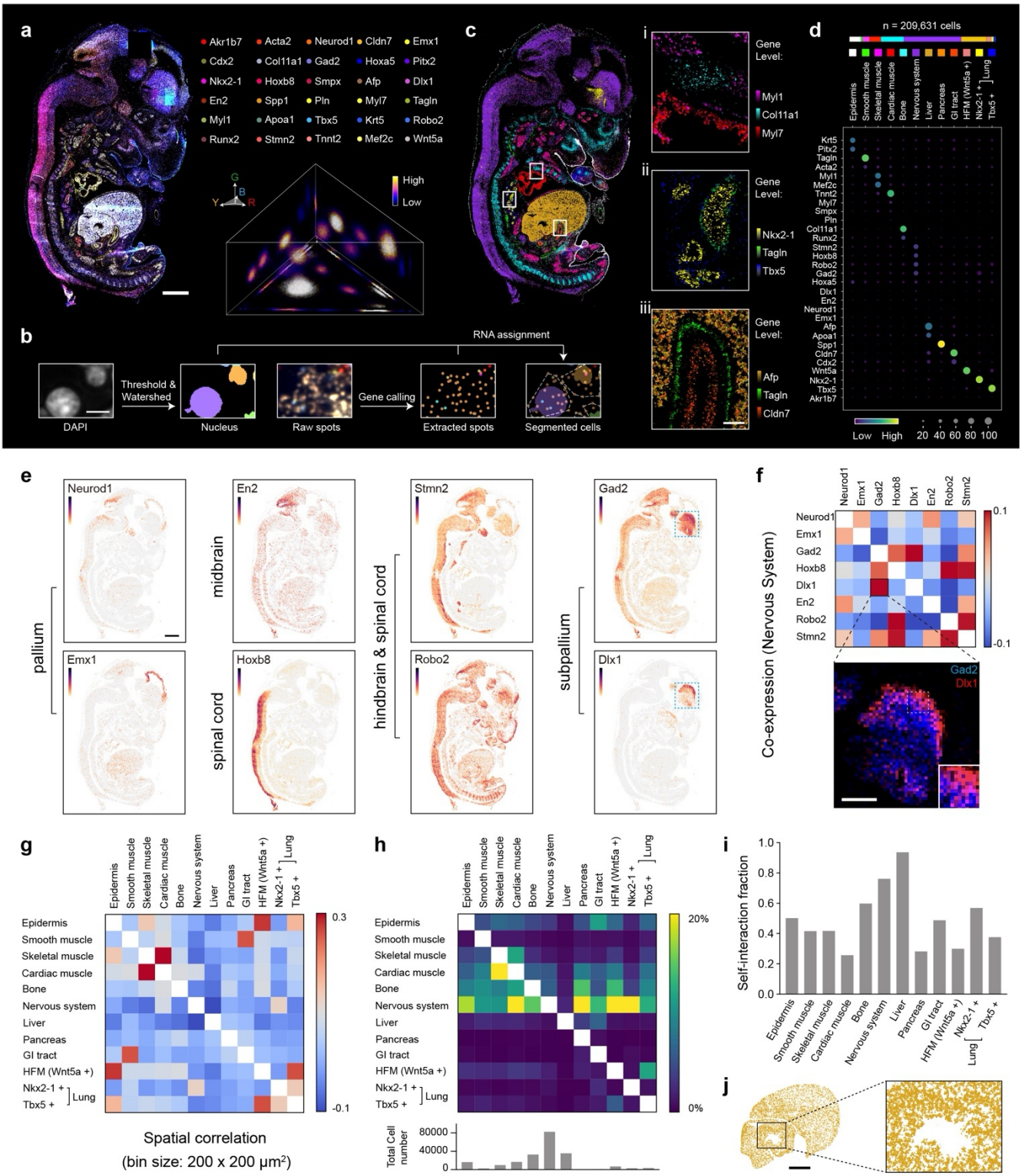

Fig. 2 | Spatial profiling of mouse E13.5 embryo by PRISM. a, A composite raw image of four channels of the mouse embryo E13.5 (sagittal section) by PRISM and the distribution of 30 clusters of spots in color space. Scale bar: 1mm. **b**, Cell segmentation. Cells were segmented through a DAPI image, and each identified gene spot was assigned to the nearest nucleus through Euclidean distance. Scale bar: 5 μm. **c**, Annotated cell type distribution of the mouse embryo. Cell type was determined by the most abundant gene within each cell. Zoom-in figs (i - iii) were colored based on gene abundance. Scale bar: 100 μm. **d**, Dot plot of each cell type. **e**, Heterogeneity in the nervous system can be revealed through gene expression patterns. The cells were colored based on the abundance of their respective gene expressions and grouped according to the organ's location. Scale bar: 1 mm. **f**, Co-expression analysis revealed the developmental role of transcription factor DLX1 in forebrain GABAergic-neuron. Scale bar: 500 μm. **g**, Coarse-grain spatial correlation analysis between general cell types. The spatial binning size is 200 x 200 μm^2. **h**, Direct intercellular interaction between general cell types. The interaction degree was defined as the proportion of cells directly adjacent to another cell. The figure below shows the total cell number of each general cell type. **i**, Direct intercellular interaction between each cell type itself. **j**, Spatial distribution of liver cells. The zoom-in image shows numerous structures of hepatic lobes, each composed of several liver cells. Scale bar: 500 μm.

Spatial profiling of mouse embryos reveals developmental intracellular interaction

We utilized PRISM with a specific panel of 30 marker genes to explore the intricate structural dynamics and cellular interactions within various tissues and organs during embryonic development (Fig. 2a-d, Extended Data Figs. 9-11). A 10-μm whole-mount section of an E13.5 mouse embryo revealed 12 primary cell types, including the epidermis, bone, and gastrointestinal tract, distinguished by the most abundantly expressed genes (Fig. 2c,d). Notably, the nervous system, constituting 39.3% of the total cell count, can be further classified based on marker gene expression, revealing spatial organization within the embryonic brain at E13.5, spanning the forebrain (pallium, subpallium), midbrain, hindbrain, and spinal cord (Fig. 2e). Gene co-expression analyses were conducted to determine the co-localization of genes within the nervous system (Fig. 2f), with significant co-localization observed for *Hoxb8*, *Robo2*, and *Stmn2* in the spinal cord region. Additionally, the co-presence of *Gad2* and *Dlx1* in the subpallium supported the hypothesis that DLX1 transcription factors directly influence *Gad2* expression, a crucial interaction for the development of GABAergic neurons in the forebrain and illustrated a key aspect of neuronal differentiation and specialization[33].

A coarse-grained spatial correlation analysis was performed to understand the relationship between various cell types across tissue locations (Fig. 2g, Extended Data Fig. 12). The co-occurrence of skeletal and cardiac muscle cells, as well as GI tract and smooth muscle cells, indicates their collaborative roles in specific developmental functions. Both *Wnt5a*$^+$ and *Tbx5*$^+$ cells play critical roles in coordinating limb development. In our analysis, *Wnt5a*$^+$ Limb mesenchymal cells are primarily distributed in the distal regions of the limb and the terminal mouth area but spatially proximal to epidermis cells. Additionally, *Tbx5*$^+$ limb fibroblasts demonstrate a high spatial correlation with *Wnt5a*$^+$ cells. They exhibit a greater abundance in the limb and are primarily distributed in the proximal regions, exemplifying their joint contribution to the formation and functioning of the limbs.

The further refined analysis assessed direct intercellular interactions (Fig. 2h), revealing a complex network of developmental interactions between the nervous system and other organs at E13.5. Notably, pancreatic cells marked by *Spp1* exhibited significant interactions with chondrocytes and osteoblasts. The presence of islands of *Spp1*$^+$ cells within the skeletal structure suggests their possible involvement in the early formation of the intervertebral disc[34] (Extended Data Fig. 13). The previously noted spatial correlation between *Wnt5a*$^+$ and *Tbx5*$^+$ cells was confirmed to be a direct interaction. Additionally, *Nkx2-1*$^+$ cells, highly expressed in certain neural precursors[35], were found predominantly in the subpallium and demonstrated extensive interactions with the nervous system, including the spinal cord. Remarkably, numerous *Nkx2-1*$^+$ cell clusters, with their distinctive disk-like structure, are scattered across the alar plates of the spinal cord, suggesting a role in cell differentiation and migration within the spinal cord (Extended Data Fig. 13). Not surprisingly, interactions within the same cell type are considerable (Fig. 2i). Particularly, 94% of hepatocytes interact with each other in close proximity (Fig. 2j), indicating early multicellular structure formation within the hepatic lobe and the establishment of basic hepatic architecture by E13.5.

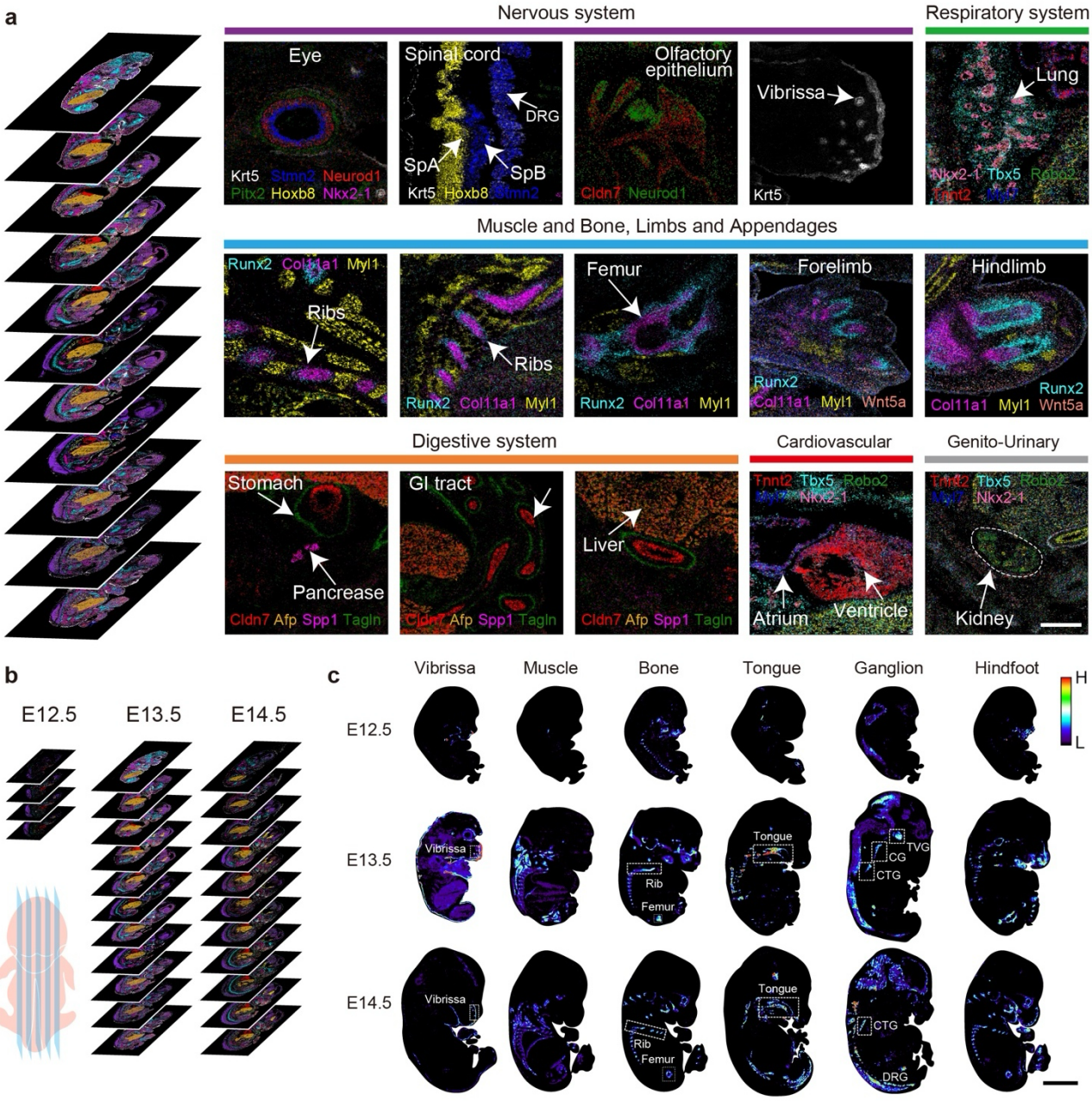

Fig. 3 | High-throughput spatial profiling of mouse embryos E12.5, E13.5, and E14.5. a, Developmental structures across different systems from a sampled-3D spatial atlas of the E13.5 embryo, constructed from 11 sagittal sections at an average interval of 280 μm. Fine structures from the nervous, respiratory, motion, digestive, cardiovascular, and genitourinary systems are collected from different sections. Cells within these structures are colored based on the abundance of their respective marker genes. Scale bar: 500 μm. **b**, Sampled-3D spatial atlas of E12.5 (4 sections), E13.5 (11 sections) and E14.5 (11 sections) embryo. **c**, Developmental progression at different stages is revealed by annotated cell distribution. DRG (Dorsal Root Ganglion); CG (Cervical ganglion); TVG (Trigeminal V ganglion); CTG (Cervico thoracic ganglion). Scale bar: 2.5 mm.

Spatial-temporal atlases illustrate embryonic organ development

Single-round 30-plex imaging has enabled rapid spatial profiling on a large scale. We created a sampled 3D cell atlas of an E13.5 mouse, utilizing 11 whole-mount sections spaced 280 μm apart, capturing a total of 2,331,513 annotated cells (Fig. 3a). This comprehensive experiment, from sample preparation to imaging, was completed within three days to generate an annotated cell atlas, allowing for the visualization of various early developmental structures.

At the E13.5 stage, the nervous system showcases distinct structures such as the multi-layered composition of the eye (marked by *Neurod1*$^+$, *Pitx2*$^+$, *Stmn2*$^+$), the spinal cord's alar and basal plates

(*Hoxb8*$^+$, *Stmn2*$^+$), and the olfactory epithelium (*Neurod1*$^+$, *Cldn7*$^+$). Notably, *Pitx2* (cornea), *Neurod*1 (outer layer), and *Stmn2* (inner layer) of the optic cup exhibit a gradient of expression from outer to inner layers, contributing to eye development. Conversely, in the pallium, an inverse pattern of expression was observed, likely due to the invagination process during the optic cup formation[36] (Extended Data Fig. 14). The respiratory system at this stage, indicated by *Nkx2-1* and *Tbx5* expression, reveals lung-lobe and mesenchymal cells' distribution. The early motion system, including muscles, bones, and appendages, can be observed with cellular resolution from various angles, laying the groundwork for future movement and coordination. Skeletal development is marked by specific gene expressions that present the differentiation direction of cell types: *Runx2* for osteoblasts and *Col11a1* for cartilage. The digestive system begins to differentiate into the gastric and intestinal primordia, with *Cldn7* marking this early form. The circulatory system, particularly the heart, shows a clear chamber structure and atrium/ventricle differentiation (*Myl7*$^+$, *Tnnt2*$^+$), while the renal system's fine structure is delineated by *Robo2*$^+$ and *Nkx2-1*$^+$ expressions. PRISM's high resolution and accuracy effectively reveal mouse embryo's intricate structures and cell types at E13.5.

Expanding the atlas to include E12.5 and E14.5 mouse embryos offers valuable insights into their developmental progression, creating a temporal-spatial atlas with 26 whole-mount sections, 4,257,418 cells, and 107,655,795 transcripts from 30 RNAs at sub-micron resolution (Fig. 3b, Extended Data Fig. 15). The period from E12.5 to E13.5 witnessed significant transformations in various tissues (muscle, skeleton, and ganglia) and organs (such as vibrissa and tongue) (Fig. 3c). By E12.5, the pancreas was identifiable as a cluster of *Spp1*-positive cells, which subsequently enlarged and underwent bifurcation by E14.5 (Extended Data Fig. 16). Also, a group of *Wnt5a*$^+$ cells in the hindbrain has a significant migration behavior, suggesting morphogenic behavior of the dorsal hindbrain during development[37]. Lung development is characterized by volume reduction and increased branching and bifurcation within the lung structure (*Nkx2-1*$^+$, *Tbx5*$^+$), and a more developed trachea is observed (*Nkx2-1*$^+$). At the same time, the GI tract shows rapid development (*Cldn7*$^+$, *Tagln*$^+$), highlighting the dynamic nature of embryonic growth and organ formation.

PRISM depicts complex histological profiling of liver tumor microenvironment

We conducted a detailed analysis of a hepatitis B virus (HBV) positive hepatocellular carcinoma (HCC) sample (around 6 mm x 5 mm) using a 31-gene panel, comprising 30 selected markers associated with cell proliferation, liver tumor cells, cancer-associated fibroblasts (CAFs), endothelial cells, and immune cells, in addition to a probe targeting a 28-bp sequence encoding the HBV core protein (Fig. 4a,b, Extended Data Figs. 17-19). This comprehensive panel facilitated the spatial exploration of the tumor microenvironment at single-cell resolution.

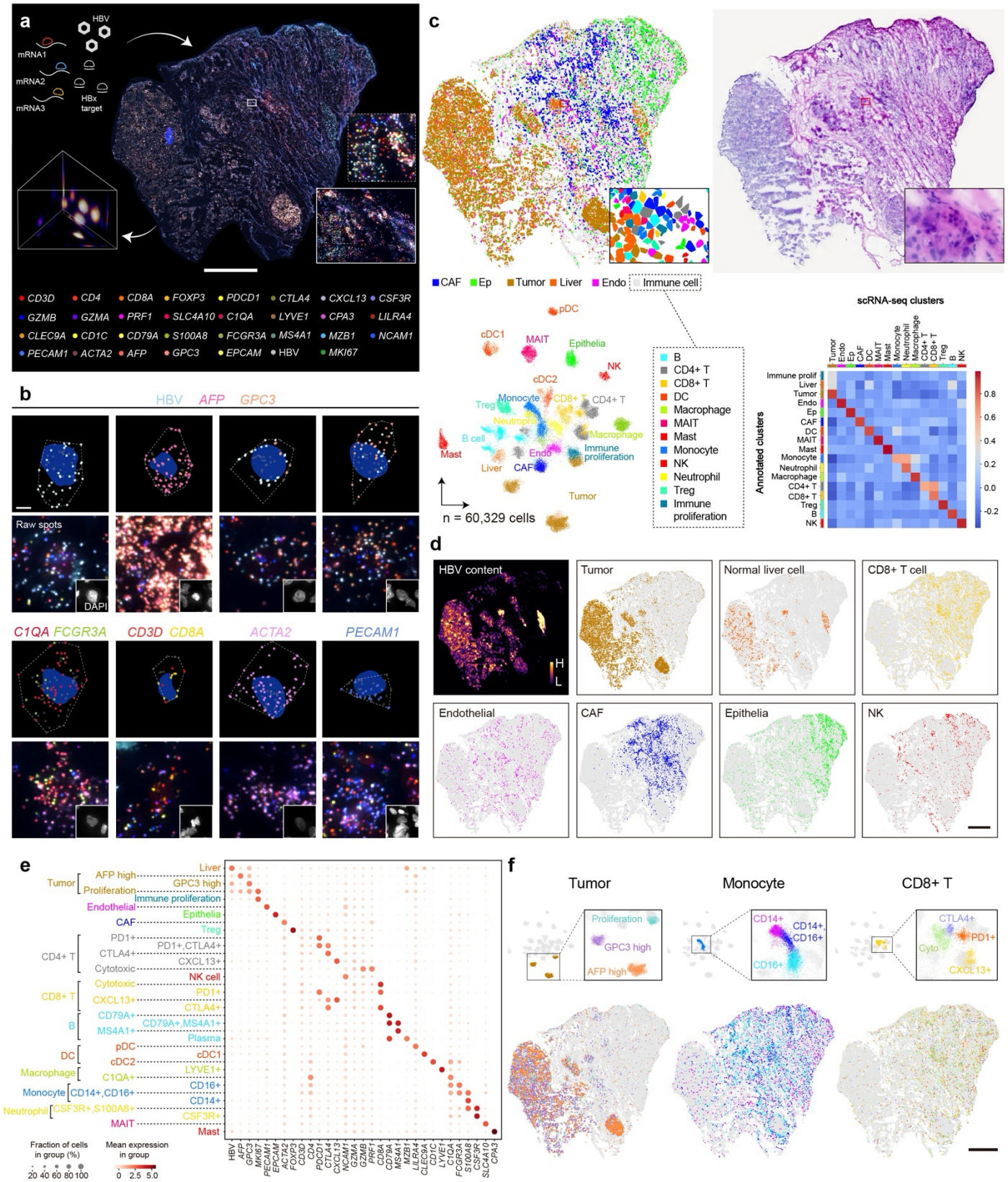

Fig. 4 | Spatial profiling of human hepatocellular carcinoma cancer microenvironment by PRISM. a, Workflow and signal results of HCC sample profiled by PRISM 31-gene panel. The panel includes 30 marker genes and an HBV core-protein-encoding RNA. Scale bar: 1 mm. **b**, Zoom-in image of called signal spot and the corresponding raw image composed of four channels and DAPI. The blue region in the first row shows the segmented nucleus and the white dot line shows the border of the segmented cell. Scale bar: 5 μm. **c**, Cell classification in HCC sample. Left: Cell annotation result is achieved through Harmony embedding and Leiden clustering, along with the UMAP plot. Upper left: projection of all cells. Upper right: H&E staining of adjacent sample section. Lower right: Pearson correlation of PRISM cell classification results with single-cell transcriptome data. **d**, HBV distribution and cell type projection. Scale bar: 1 mm. **e**, Dot plot of each cell subtype. **f**, Projection of cell subtype. Scale bar: 1 mm.

Our analysis identified 1,695,418 transcripts and assigned them to 60,329 cells, predominantly classified into 12 immune cell populations—including B cells, T cells, natural killer cells, mast cells, neutrophils, monocytes, and macrophages—and 5 non-immune cell types, such as CAFs, tumor cells, epithelial cells, endothelial cells, and normal liver cells (Fig. 4c,d, Extended Data Figs. 20, 21). This classification was achieved using Harmony embedding and unsupervised clustering analysis based on distinct PRISM gene marker profiles. The accuracy of cell identification by PRISM closely aligned with findings from prior single-cell RNA-sequencing studies[38–40] (Fig. 4c), validating the effectiveness of this single-round imaging technique.

By mapping single cells to their original spatial contexts, we found that the spatial distributions and positions of cells, as revealed by PRISM, well correlated with those derived from conventional fast-diagnostic hematoxylin and eosin (H&E) and immunohistochemistry (IHC) staining methods (Fig. 4c, Extended Data Fig. 22). Further analysis delineated dominant cell populations into subtypes based on their unique PRISM profiles (Fig. 4e,f, Extended Data Figs. 23). For instance, tumor cells were subdivided into three distinct clusters characterized by the high expression of *AFP*, *GPC3,* and *MKI67*, indicating intrinsic heterogeneity. Immune cells, such as monocytes, were further categorized into classical ($CD14^{+}CD16^{-}$), non-classical ($CD14^{dim}CD16^{+}$), and intermediate ($CD14^{+}CD16^{+}$) subtypes. Additionally, cytotoxic, exhausted and $CXCL13^{+}$ tumor-reactive CD8 T cells were identified as unique subpopulations among CD8 T cells (Fig. 4f). This nuanced classification and the investigation of spatial interactions and communications among individual cell types, especially between different subtypes, enrich our understanding of the complex dynamics within the tumor microenvironment.

Quasi-3D atlas exhibits microscale stromal heterogeneity in liver tumor

To comprehensively understand the tumor microenvironment of this HBV-infected HCC, we advanced our spatial analysis beyond simple 2D projections by constructing a quasi-3D landscape. This was achieved by stacking PRISM data from 20 consecutive tissue sections to form a thick slab-like structure (Fig. 5a). Within this structure, based on gene density, the tissue was categorized into four primary areas: tumor (Territory 1), normal liver (Territory 2), epithelia (Territory 3), and CAFs (Territory 4) (Fig. 5b).

Investigating the tumor's 3D architecture revealed its complex structural and functional characteristics, which could be oversimplified by 2D analysis alone. For instance, tumor nests appearing separate in 2D images were found to be interconnected in 3D, highlighting the intricate nature of tumor growth in complex microenvironment (Fig. 5c). The quasi-3D PRISM analysis, utilizing a 31-gene panel, maintained high consistency in cell identification with 2D and single-cell transcriptomics datasets, revealing 16 main cell clusters and identifying 1,218,279 single cells (Extended Data Fig. 24). This 3D analysis more realistically represented the heterogeneity within the tumor microenvironment, showing significant variations in cell composition across the defined territories, with a predominance of tumor cells in Territory 1 and CAFs in Territory 4 (Fig. 5d).

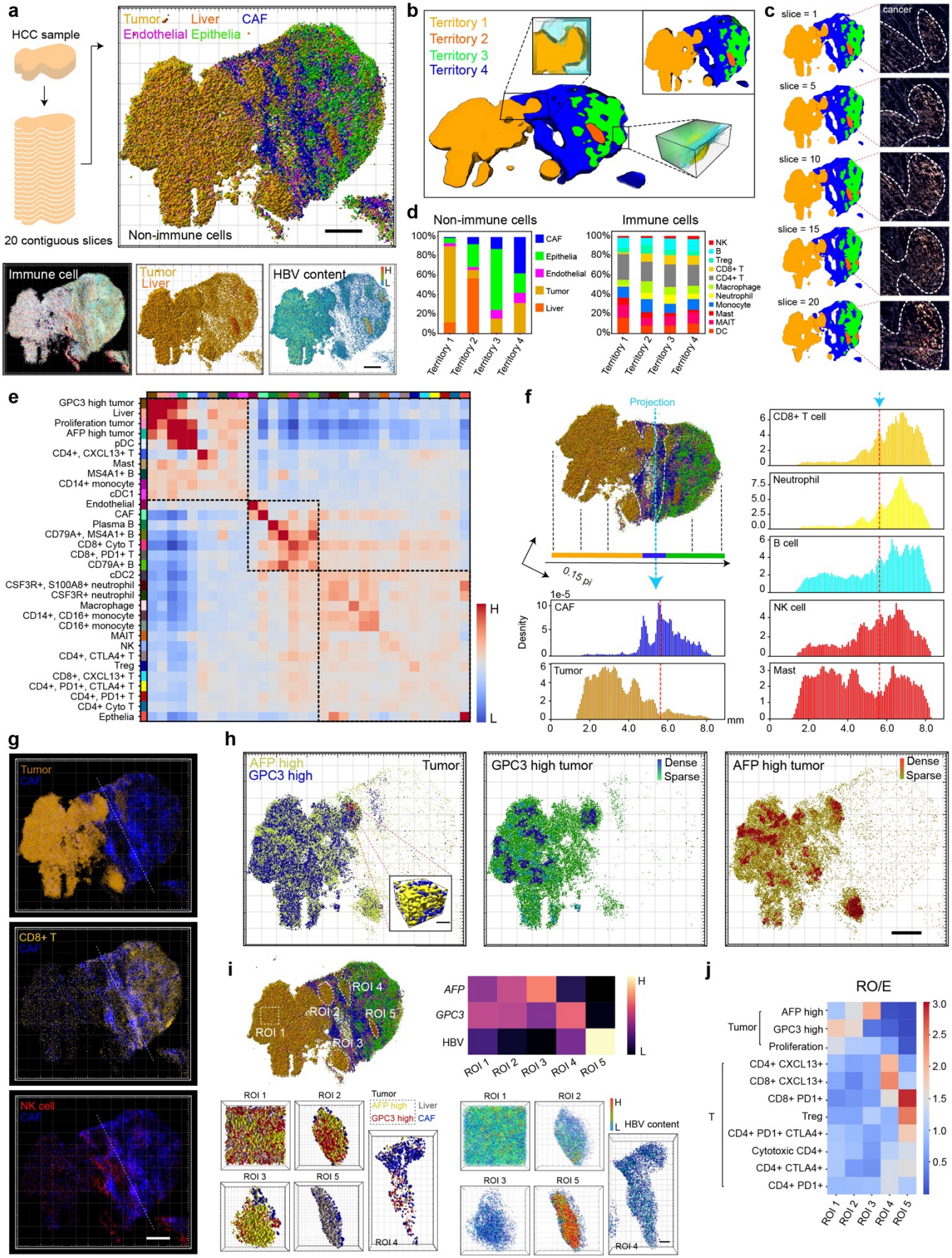

Fig. 5 | Quasi-3D atlas reconstitution of HCC microenvironment through PRISM consecutive profiling. a, reconstituted 3D cell distribution from 20 consecutive sections. Scale bar: 1 mm. **b**, 3D territory segmented based on gene distribution: tumor (Territory 1), normal liver (Territory 2), epithelia (Territory 3), and CAFs (Territory 4). **c**, Sections in 3D territory. The right part shows the raw image in a tumor-protrusion region. **d**, Percentages of cells in four territories. **e**, Neighborhood enrichment analysis for all subtypes. The enrichment value reflects the spatial correlation between cell subtypes. Three circled modules are tumor module, CAF module, and

epithelia module. **f**, CAF barrier effect on immune infiltration through cell density projection. The sample was rotated by 27˚ to align its 'long-axis' horizontally, and all cells were projected onto this long axis. The distribution histograms from several cell types were presented. The red dot line marks the peak distribution locus of CAF. **g**, 3D projection of CAF cells (blue color) and immune or tumor cells, displayed in the respective color. Scale bar: 1 mm. **h**, Intra-tumor heterogeneity. GPC3- and AFP-high tumor cells exhibit different distribution pattern, reflecting the extent of their aggregation. The color is bluer/ redder in a denser region for the respective two subtypes. Scale bar: 100 μm, zoom-in image; 1 mm, right. **i**, ROI analysis for the tumor and HBV-infected region. Upper left: the locations of 5 ROI. Upper right: AFP, GPC3, and HBV core-protein expression levels in 5 ROIs. Lower left: Distribution of AFP-high tumor cells, GPC3-high tumor cells, normal liver cells, and CAFs in 5 ROIs. Lower right: HBV content in 5 ROIs. Scale bar: 200 μm. **j**, The ratio of Observed to Expected (Roe) cell type enrichment analysis in 5 ROIs.

Further analysis segmented these main clusters into 32 subpopulations based on distinct PRISM gene markers (Extended Data Fig. 24). Spatial neighborhood enrichment analysis revealed three distinct cellular modules characterized by specific co-enrichment patterns: tumor, CAF, and epithelia modules, each demonstrating unique cell contexts (Fig. 5e, Extended Data Fig. 25), with most immune cells located in non-tumor modules, while the tumor module displayed a lower level of immune infiltration. Notably, CAFs formed a physical barrier that restricted the infiltration of various immune cells (Fig. 5f, Extended Data Fig. 26), including CD8+ T cells, and neutrophils, into the tumor module. However, NK cells were observed to migrate across the CAF wall, appearing at the forefront of the tumor tissue, indicating their specific immune activity and anti-tumor roles (Fig. 5g).

We also explored intra-tumor heterogeneity using HCC markers AFP and GPC3 (Fig. 5h). Distinct spatial aggregations of GPC3-high and AFP-high cells were noted throughout the tumor territory, suggesting the potential subclone lineage differences. Analysis of four tumor and one normal region of interest (ROIs) revealed higher HBV core-protein expression in the normal liver tissue, suggesting early HBV expression before liver cancer development, and this sub-clone dynamic difference has high-level heterogeneity at the scale about 100 μm (Fig. 5i). The immune landscape varied across these ROIs, with higher immune cell enrichment in ROI4 and ROI5, contrasting with other tumor ROIs, due to the barrier effect of CAFs (Fig. 5j). Furthermore, exhausted T cells expressing PD1 and/or CTLA4 were relatively more abundant in the HBV-positive normal liver ROI5 region than in tumor ROI4. In contrast, CXCL13+ CD8 and CD4 T cells, defined as tumor-reactive T cells[41], were more enriched only in ROI4, potentially due to different immune responses against the HBV virus and tumor, respectively. This detailed investigation underscores the complex cellular interactions within HCC, offering insights that could significantly enhance our understanding of liver cancer's development and progression.

Large-scale 3D in-situ profiling unveils subcellular RNA distribution heterogeneities in intact thick tissues

We expanded the application of PRISM to analyze the intact 3D cell composition of 100-μm thick mouse brain tissue, demonstrating its inherent compatibility with the processing and imaging of thick tissues (Fig. 6a). This approach preserves the three-dimensional structure with high precision and accuracy. By employing single-round imaging, we avoid repetitive macromolecule penetration within thick tissues and effectively circumvent the challenges of fluorescent spot alignment and registration across multiple imaging sessions, which are common issues in existing multiplex RNA imaging techniques.

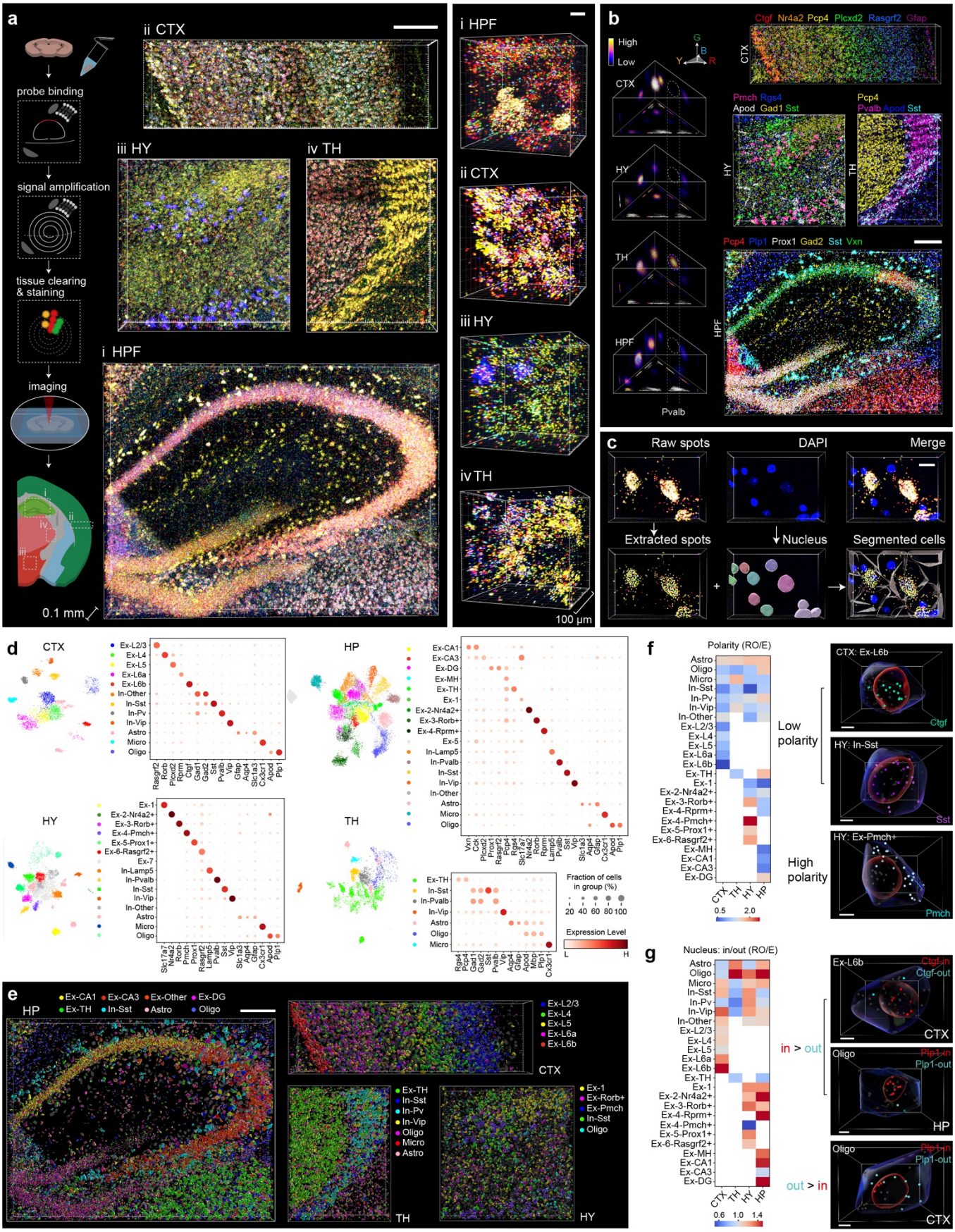

Fig. 6 | Three-dimensional RNA in-situ staining in intact mouse brain tissue. a, Workflow of PRISM 3D staining and the result. A liquid removal step is added after rolling circle amplification. RI matching was performed after imaging probe staining to enhance optical transparency. Resulting images depict tissues from four brain regions (HPF: hippocampus formation, CTX: cortex, HY: hypothalamus, and TH: thalamus). The images are composited using four channels (488nm, 532nm, 594nm and 647nm). Scale bar: 200 µm, upper left; 10µm, right. **b**, Gene calling in color space for four brain tissues. The distribution of selected called genes in the four tissues is shown on the right side. Scale bar: 200 µm. **c**, Three-dimensional cell segmentation. Identified gene spots were assigned to nearest nucleus centroid through Euclidean distance. Scale bar: 15 µm. **d**, Cell classification by Harmony embedding

and Leiden clustering. The color of cell type is unified in UMAP plot and dot plot. **e**, Projection of selected cell type in the four brain regions. Scale bar: 200 μm. **f**, Subcellular RNA polarity analysis across brain regions and cell types. The polarity analysis was based on the scaled distance of RNAs' centroid and nucleus centroid within one cell. The top 20% was designated as high polarity, and the bottom 20% as low polarity. The values in heatmap were calculated through Roe analysis using the ratio of high to low polarity. Polarity values from 30 genes were averaged in each cell type. On the right: examples of cells with high / low RNA polarity. Scale bar: 5 μm. **g**, Subcellular RNA enrichment in nucleus: this was calculated by RNA inside / outside of the nucleus. Similarly, the values were based on Roe analysis. On the right: examples of cells with RNA spots tend to distributed inside/outside of nucleus. Scale bar: 5 μm.

To improve optical transmission through thick tissues, we incorporated a post-amplification lipid-removal step and used iohexol for refractive index matching to ensure a deep light penetration (Extended Data Fig. 27). PRISM decoding process utilized signal ratios rather than absolute intensities, offering greater tolerance to optical attenuation (Extended Data Fig. 28). Utilizing a 30-gene panel, we profiled large-scale ROIs within the brain, including the cortex (1330 μm x 400 μm x 100 μm, 207,115 RNA transcripts), hypothalamus (770 μm x 770 μm x 100 μm, 214,671 RNAs), thalamus (585 μm x 775 μm x 100 μm, 156,368 RNAs), and hippocampus (1520 μm x 1150 μm x 100 μm, 548,220 RNAs) (Fig. 6a). The detailed morphology and distribution of marker genes within these areas align with those observed in 2D experiments, indicating consistency across dimensions (Fig. 6b).

Cell segmentation and classification within each ROI were performed using StarDist[42,43] for 3D nucleus segmentation(Fig. 6c) (resulting in 4,792 cells in the cortex ROI, 9,838 cells in the hypothalamus ROI, 5,370 cells in the thalamus ROI, and 20,721 cells in the hippocampus ROI, respectively) and Harmony embedding coupled with Leiden cell clustering, respectively, allowing for the differentiation of glial cells and neurons into subtypes (Fig. 6d). These classifications show high correlation with single-cell transcriptome data, affirming the robustness of our approach (Extended Data Fig. 29). Furthermore, we projected the cells back to their original locations to construct a 3D neuron-architecture landscape (Fig. 6e).

Beyond cellular analysis, we delved into subcellular RNA distribution, assessing RNA polarity within each cell type by averaging the distances from RNA centroids to the cell centroid for each gene (Fig. 6f). This revealed notable heterogeneity in RNA polarity among neurons. For example, In-*Sst*, Ex-L6b has little polarity with given genes, but *Pmch*$^{+}$ cells in the hypothalamus display pronounced polarity, with most RNA gathering on one side of the cell body, indicative of unique cell orientations. Additionally, we explored gene enrichment within the nucleus across different cell types and brain regions, uncovering a heterogeneous RNA distribution that suggests variations in cell morphology, RNA nascency, and transcriptional activity (Fig. 6g, Extended Data Fig. 30). Interestingly, the nuclear distribution may be associated with brain regions. Within the oligodendrocytes, the majority of RNA tends to distribute inside of the nucleus in the hippocampus and thalamus (mainly reflected in oligodendrocytes in fiber tracts within these two ROIs), while in the cortex, the behavior is the opposite. This subcellular RNA distribution adds another dimension to our understanding of cell types and states, enriching the comprehensive profile provided by PRISM.

Discussion

PRISM leverages multi-channel color barcoding, a strategy that utilizes specific spectral combinations to enhance information density within a single round of fluorescent in situ hybridization staining and imaging, while facilitating size-independent barcoding of clonal RCA products. PRISM's coding space can be expanded by increasing the levels of division without the need to lengthen

barcode sequences. This scalability is achieved by setting more divisions and reallocating each quantity of level. Each level can also be fine-tuned by adjusting the mixing ratio of fluorescent probes to non-fluorescent probes corresponding to each barcode segment in the padlock probes. This adjustability of the fluorophore proportion for each level and channel also allows for optimal differentiation in color space upon decoding, making PRISM adaptable to various optical imaging systems regardless of their specific configurations and performance characteristics.

Furthermore, considering the inherent autofluorescence of many tissues, PRISM's design accommodates quantization of each spectral channel to different levels to ensure distinguishability within a fluorescence-background environment. Demonstrated effectiveness in diverse sample types—including mouse brain, mouse embryo, and human tumor sections—and the ability to detect viral genome fragments as short as 28 bp highlight PRISM's broad applicability.

With PRISM, encoding capacity is enhanced by creating higher intensity levels within respective channels, achieving an experimentally verified encoding capability of up to 64 genes in a single imaging round using four channels (Extended Data Fig. 31). Specifically, three spectral channels Ch1, Ch2, and Ch3, were quantized into fifths (0, 1/5, 2/5, 3/5, 4/5, 5/5), and the channel (Ch4) was divided into three levels: high (2), low (1), and none (0). This capacity allows for detailed spatial mapping, as demonstrated in an E13.5 stage mouse embryo, mouse brain and 100-μm thick tissue (Extended Data Fig. 32-34). The information entropy of imaging is much higher in PRISM than in other multi-cycle methods. While cyclic strip-and-rehybridization methods or adding more optical channels could theoretically further increase multiplexity for PRISM, they would necessitate extensive instrumentation and image processing, compromising the ease of use. Challenges such as signal crowdedness, a common issue in amplification-based RNA in-situ detection technologies, are acknowledged as multiplexity increases. This could be addressed through expansion microscopy or introducing pseudo-channels via light-activable/switchable fluorescent dyes or fluorescence resonance energy transfer (FRET).

PRISM distinguishes itself through cost efficiency, simplicity, and rapid experimental turnaround. Using standardized reagents and conventional benchtop setups, it serves as a valuable tool for diverse applications, from molecular pathology to cell atlas construction, while minimizing reliance on specialized equipment and extensive resources. With an open-source design, PRISM aims to simplify super-multiplex RNA imaging, making it as accessible as standard immunostaining techniques.

Materials and Methods

Mice

Animal handling and tissue harvesting methods were conducted in strict adherence to local animal welfare laws and approved by the institutional ethics committee, ensuring alignment with recognized ethical standards for laboratory research. C57BL/6J mice, aged over 8 weeks, were euthanized to collect brain tissue, while embryos at 12.5-, 13.5-, and 14.5- days post-fertilization were also harvested. The mouse brains and embryos were embedded in Tissue-Tek O.C.T. Compound (Sakura), snap frozen, and stored at -80 °C. The frozen tissues were sectioned into 10-μm-thick sections using a CM1950 Cryostat (Leica) and affixed to a Superfrost Plus glass slide (Epredia). The mounted sections can be stored for one year at -80 °C.

For 100-μm-thick tissue collection, C57BL/6J mice (> 8 weeks) were perfused with pre-chilled 4% (w/v) RNase-free paraformaldehyde (PFA, Shanghai Yuanye). Subsequent to the brain harvest, the samples were immersed in 4% PFA for overnight fixation at 4 °C. The brain was sectioned into 100-μm slices using a VT1200 Vibratome (Leica). These slices were preserved in 70% ethanol at 4 °C and can be stored for more than 3 months.

Human

A human hepatocellular carcinoma sample was placed in RPMI 1640 medium after being obtained from the patient and then processed following the aforementioned protocol (O.C.T embedded, snap frozen and sectioned into 10-μm-thick sections). This study was approved by the Ethics Committee of Beijing Shijitan Hospital, Capital Medical University. The patient in this study provided written informed consent for sample collection and data analyses

Probe and barcode design. For each gene of interest, a 40-nucleotide (nt) sequence window was identified to serve as a binding target, with 2 binding targets designated for mouse brain and embryo samples and 3 for human hepatocellular carcinoma samples. The selection of these 40 nt sequences adhered to stringent criteria: a melting temperature (Tm) in the range of 65-75 °C, the absence of 5 consecutive guanine (G) bases, coverage of all transcript variants (if possible), no blast results matching other genes in the NCBI database and no significant DNA secondary structure ($\Delta G > -10$). Each 40nt sequence was divided into two segments of 20nt each for the construction of a padlock probe.

A “selective amplification” strategy was employed to mitigate signal crowdedness issues[26]. Briefly, padlock probes targeting highly expressed genes were doped with irreplicable ones at proper ratios during probe pooling. This approach can dilute the clonal populations by partially suppressing the amplification of abundant transcripts. During analysis in a gene-by-count expression matrix, the quantities of those transcripts can be recalibrated by multiplying respective dilution ratios. The dilution ratios we used for all samples were listed in **Supplementary Table 1**.

Each gene was assigned by a specific barcode (Fig. 1a,b). The barcode consists of four segments (20nt for each), with segments specifically binding to fluorescent probes labeled with Texas Red (TR), Alexa Fluor 488 (AF488), Cy5, and Cy3, respectively. For 30- or 31-plex coding, each segment variation allowed for 0-4 intensity levels per channel except for Cy3, which was compromised to two levels due to autofluorescence background. Color Intensity grading is realized by mixing fluorophore-labeled probes with unlabeled probes in a specific ratio. A standard mixing ratio for all imaging probes was listed in **Supplementary Table 2**. The ratio for each probe pair can

be finetuned based on actual optical setups. All padlock probes and imaging probes were ordered from Sangon Biotech (Shanghai).

Sample treatment for thin tissue. For a 10-μm-thick frozen tissue sections, they were fixed in 4% PFA with 0.1% Glutaric dialdehyde at room temperature for 15 min after taking out from -80 °C refrigerator, and then rinsed with PBST (0.25% Tween in 1X PBS). The sample was then permeabilized with 0.01% pepsin in 0.1M HCl at 37 °C for 2 min (10 min for mouse embryos), followed by PBST washing. After that, the samples were dehydrated in a series of ethanol washes (10 min in 80% ethanol followed by 2 min in pure ethanol) and rehydrated with three-time PBST washes. The samples were then blocked with oligo-dT (100 nM oligo-dT, 50 mM KCl, 20% formamide, 20 μg/mL BSA, 20 μg/mL Yeast tRNA (AM7119, Invitrogen), and 1U/μL Ribolock RNase inhibitor (Thermo Scientific) in Ampligase buffer (Lucigen)) for 10 min at room temperature. The probes were hybridized to sample by incubating them in a hybridization mix (200 nM padlock probe for each, 50 mM KCl, 20% formamide, 20 μg/mL BSA, 20 μg/mL Yeast tRNA, and 1U/μL Ribolock RNase inhibitor in Ampligase buffer) at 55 °C for 10 min and then 45 °C for 1 h 50 min. The samples were placed in a humidified chamber during incubation. After hybridization, the samples underwent three 10-min washes with washing buffer (10% formamide, 2X SSC buffer) to remove non-specifically bound probes, then rinsed twice with PBST. The samples were then incubated in a ligation mix (2.5 U/μL SplintR ligase (NEB), 20 μg/mL BSA, and 1U/μL Ribolock RNase inhibitor in SplintR buffer) at 37 °C for 2 h in the humidified chamber to ligate the nick on the padlock probe, and then the samples were rinsed twice with PBST. Then rolling circle amplification (RCA) was performed using 0.25 U/μl Phi29 polymerase in Phi29 polymerase buffer (Thermo Scientific) with 250 μM dNTP (Thermo Scientific), 50 μM aminoallyl-dUTP (Thermo Scientific), 10% glycerol, 20 μg/mL BSA, and 600 nM RCA primer at 30 °C overnight. Following RCA, the samples were rinsed twice with PBST, then fixed with 10 μg/μL $BS(PEG)_9$ (Thermo Scientific) in PBST at room temperature for 15 min, and subsequently rinsed three times with PBST. Finally, the samples were washed three times (5 min per wash) with 65% formamide at room temperature, followed by two PBST rinses.

Sample treatment for thick tissue. Tissue slices were rehydrated through three 20-min incubations in PBST after taking out from ethanol, then blocked with oligo-dT (the reagent was the same as mentioned above) for 30 min at room temperature. After that, the probes were hybridized to samples as mentioned above at 45 °C for 24 h with gentle shaking, followed by washing three times for 20 min each in washing buffer, then three times in PBST, then overnight incubation at 37 °C in ligation mix as mentioned above with gentle shaking. After two PBST washes, RCA was performed at 30 °C for 12-24 h, followed by two gentle PBST washes and a 30-min fixation in $BS(PEG)_9$ as mentioned above. After being washed three times with PBST and 65% formamide at room temperature, the samples were treated with Cubic L (TCI, T3740) for lipid removal at room temperature for 1 h, followed by three-time washes with PBST.

Imaging probe staining. The samples were incubated with a mixture of 17 fluorescence-labeled probes (each paired with an unlabeled probe at a predetermined ratio, with a final concentration of 120 nM per pair, 20% formamide, 2X SSC buffer) at 50 °C for 9 min, then 37 °C for 21 min. Following incubation, the samples were washed three times with washing buffer, stained with DAPI (5 μg/μL, Beyotime) solution at room temperature for 5 min, and washed twice with PBST. Finally, the samples were mounted for imaging with a cover slip.

For staining 100-μm-thick samples, the imaging probe incubation was performed at 37 °C for 2

h, followed by three washes with washing buffer, a 30-min DAPI staining at room temperature for 30 min, and two PBST washes. The samples were then immersed in 0.9 M Iohexol (2X SSC buffer) for 1 h at room temperature to adjust the refractive index. Finally, the sample was mounted for imaging with a cover slip and sealed with varnish to avoid evaporation.

Imaging. Imaging can be performed using any proper instrument. Our experiment used both home-built setups and conventional equipment in the core facilities. For our thin tissue imaging, the sample was placed on a home-built epifluorescence microscope, using MIM microscopic imaging framework (Applied Scientific Instrumentation: RAMM-Basic, RAMM-URKIT2, MIM4-OSM25, TN200-MMC, LLG3-E100, MS4, S551-2201B, C60-3WMS-Mx) equipped with an S551-2201B motorized stage (Applied Scientific Instrumentation), an ATF6.5 SYS 785 automated focusing module (Wise Device Inc.), a CFI S Plan Fluor ELWD (40X NA 0.60) objective lens (Nikon), an X-Cite Turbo LED light source (Excelitas Technologies), and an Orca Fusion BT scientific CMOS camera (Hamamatsu Photonics). To image a 10-μm thick tissue, we collected a z-stack of 9 planes at ~1 μm intervals for each tile. 4% overlap was preserved for image registration and stitching. Cy3 / Cy5 (543/647 nm) and AF488 / TxRed / DAPI (488/594/405 nm) channels were serially imaged. The sampling rate was 0.1625 μm x 0.1625 μm per pixel. For a tissue section similar in size to a coronal section of a mouse brain (~1 cm x 0.6 cm), approximately 600 tiles were imaged, and it took around 90 min for imaging. The imaging can also be performed with a commercialized fluorescent microscope or a slide scanner with appropriate channels. Conventional confocal microscopes, such as LSM880 and LSM980 (Zeiss), were used for thick tissue imaging. Fast DAPI scanning across the whole slice was performed to find regions of interest. 647 nm, 594 nm, 543 nm, 488 nm, and 405 nm were used for imaging. The sampling rate is 0.207 x 0.207 x 0.700 μm/voxel for a 100 μm-thick brain slice. The imaging software was Zen or Zen Blue for LSM880 and LSM980, respectively.

Image processing. The thin tissue imaging processing was conducted as described in SPRINTseq[26]. Briefly, images from multiple planes were stacked along the z-axis through an all-in-focus algorithm, followed by shade correction with CIDRE[44] and inter-channel image registration (Fast Fourier Transform and Maximum Cross-Correlation). We resized the image from different channels to correct color aberration. The tiles were stitched using the Microscopy Image Stitching Tool[45]. For the alignment of 20 consecutive sections of HCC, we manually identified and marked 12 anchor points on each slide. The initial slide served as the template, and affine transformations were applied to the subsequent 19 slides, utilizing the established anchor points. This procedure facilitated the creation of a comprehensive, albeit artificial, three-dimensional representation of the HCC sample. For thick tissue, registration from each channel and stitching from each tile were performed in ImarisStitcher (Oxford Instruments). Given that dehydration, refractive index (RI) matching during sample preparation can cause a slight tissue shrinkage in the z-direction. Z-length was rescaled to 100 μm during image processing.

Gene decoding. Signal spots were extracted from all four channels, using local maxima on tophat-filtered images for thin tissue (Extended Data Fig. 2). 3D local maxima and Gaussian fitting (Airlocalize)[46] were used for spot extraction for thick tissue. Spot signal coordinates across all channels were combined to get 4-dimensional (from Ch1 to Ch4) intensity information. Four intensity values were normalized through the mean values of respective channels. Spots with low sum-intensity values were filtered out.

In barcode design, Ch1, Ch2, and Ch3 sum into a fixed value (for a 30/31-barcode set, this sum is 4), corresponding to 15 intersected barcodes (Fig. 1e). Correspondingly, all spots from real data

made up 15 rays in color space (x=Ch1, y=Ch2, z=Ch3, Extended Data Fig. 2). Due to experimental noise, such as amplification variations and the fluorophores' chemical environments, spots belonging to one ray may exhibit behavior indicative of a joint distribution. This joint distribution resembles a Poisson-like distribution along the direction of the ray and a Gaussian-like distribution in the direction perpendicular to the ray. For the ease of following analysis, we projected all spots onto the plane (x+y+z=4); this was accomplished by scaling the sum value of each spot to 1 (L1-normalization). Therefore, the values in each spot were transformed to their respective proportions: Ch1/(Ch1+Ch2+Ch3), Ch2/(Ch1+Ch2+Ch3), Ch3/(Ch1+Ch2+Ch3). Thus, in color space, the Poisson-like distribution along the ray can be omitted, and spots on the plane can be seen as the results of 15 Gaussian distributions with different means and variances or 15 clusters. The channel Ch4 was also scaled to Ch4/(Ch1+Ch2+Ch3), set as a new z-axis for the projected 2-D plane. The new color space axes were defined as x = (Ch1-Ch2)/(Ch1+Ch2+Ch3), y = 2*Ch3/(Ch1+Ch2+Ch3)-1, and z = Ch4/(Ch1+Ch2+Ch3). As a result, 15 clusters (determined by Ch1, Ch2, and Ch3) x Ch4 (Yes or No) make up a new color space containing 30 clusters. Barcode 31 (Ch4 only) was not in this color space, but it could be easily extracted since it has a high-scaled Ch4 value.

The position of each cluster in the color space corresponds to a specific barcode by design. Gaussian fitting was used to evaluate the degree of separation between clusters (Extended Data Fig. 3c). Since the barcode at endpoint “0” and “1” (*e.g.*, ‘4000’, ‘0040’) was not a cluster but a spot in color space, we introduced a small variance to coordinates from all spots to make the endpoint cluster a pseudo-Gaussian distribution. This way, an equal assessment can be achieved between these endpoint clusters and other clusters. The initial gene calling step for all signal spots was based on thresholding their confidence level of belonging to a certain cluster. For a refined gene calling confirmation, three-dimensional boundaries were manually delimited for barcodes in color space. Direct gene calling (all spots within such delimited boundary were assigned to a specific barcode) or iterative Gaussian fitting within delimited boundary before thresholding confidence level could be performed to ensure decoding accuracy (Fig. 1f, Extended Data Fig. 3c).

Cell segmentation. Cell segmentation and RNA assignment for thin tissue were performed as described in SPRINTseq[26] (Extended Data Fig. 7, Fig. 2b). Briefly, nuclei were segmented using adaptive thresholding of DAPI-stained images to generate binary images. The binary image underwent Euclidean transform and was further segmented using the watershed algorithm. Identified RNA spots were assigned to the nearest nucleus centroid using a K-D tree, with a "Cell Index" assigned to each RNA in the result table to create an expression matrix for the cells. For 3D data from thick tissue, nuclei segmentation was achieved using the StarDist algorithm with a pre-trained model.

Annotating and mapping cell types. Different cell-type classification methods were applied. For data derived from mouse embryos, due to the large number of cell types and the lack of comprehensive reference single-cell transcriptome datasets, cell types were directly annotated based on marker gene expression levels. For other datasets, we primarily used Harmony[47] algorithm to integrate our data with published single cell transcriptome datasets. After integration, a graph-based structure was created, enabling the division of cells into multiple clusters using Leiden algorithm at high resolution, followed by manual annotation of cell type based on marker genes. This process was applied to the 2D mouse brain section (Extended Data Fig. 8), 3D mouse brain data (Fig. 6d), and Human Hepatocellular Carcinoma (HCC) samples (Fig. 4c, Extended Data Fig. 24). Each cell's boundary was delineated by constructing a convex hull based on its peripheral RNA spot. Cells were visualized in 2D or 3D using uniform manifold approximation and projection, with colors

assigned to each cell type.

For the 2D coronal section of the mouse brain, single-cell transcriptome datasets from mousebrain.org for four tissues (CTX: cortex, HP: hippocampus, TH: thalamus, HY: hypothalamus) were integrated with our data. After preprocessing, PCA, neighbor detection, Leiden clustering, and annotation, the cells were divided into 14 types, including excitatory neurons, inhibitory neurons, and glial cells (Extended Data Fig. 8).

For 3D mouse brain, the cell classification was performed by 4 tissue types respectively, and the strategy was similar. Different thresholds were used based on the differences between tissues (CTX, HP, TH, HY). The integration of single-cell datasets for these tissues, obtained from mousebrain.org, was conducted using the Harmony algorithm. Following integration, a graph-based structure was created, enabling the division of cells into multiple clusters using the Leiden algorithm at a high resolution and then manually annotated as various cell types and subtypes based on marker genes and tissue type (Fig. 6d). Detailed parameters are shown in the corresponding Jupyter Notebook.

For HCC, the data passed quality control was processed standard pipeline, including normalization, log1p, regress_out, and scaling, before performing Principal Component Analysis (PCA). This foundational work facilitated the integration of spatial transcriptomics data with three distinct single-cell transcriptome datasets (GSE151530, GSE140228, CNP0000650) using the Harmony algorithm within the Scanpy framework. Utilizing high-resolution Leiden shared nearest neighbor clustering. We identified a fine-grained set of subclusters, which were meticulously annotated based on the expression of numerous housekeeping and marker genes associated with HCC, revealing 35 subclusters representing a diverse array of immune and nonimmune cell types. Notably, the challenge posed by the absence of well-defined marker genes for liver cells was addressed by leveraging elevated Hepatitis B Virus (HBV) expression as a proxy, allowing for the precise annotation of liver cell types. This approach helped to reclassify cells initially tagged as 'other' types, ultimately enriching the final dataset with 60,329 cells categorized into 17 major types and 35 subtypes (Fig. 4e, Extended Data Fig. 20). Each cell type was visually distinguished using a unique color scheme in the convex hull representation, with the dataset displayed in two dimensions using Uniform Manifold Approximation and Projection (UMAP)[48].

A similar methodology was applied to analyze the quasi-3D HCC dataset, yielding annotations that identified the same major cell types while revealing slight variations in cell subtypes.

Cell classification evaluation using single-cell transcriptome data. To verify the accuracy of cell classification, we subjected the single-cell data from the mouse brain cortex, comprising 50,478 cells with 27,998 genes from mousebrain.org, to Leiden clustering. We manually annotated these cells based on the expression of the 30 marker genes we were interested in. Subsequently, we took the expression matrices processed by normalization, log1p, regress_out, and scale of corresponding cell types from both spatial transcriptomics and single-cell transcriptome data, and then averaged the data across the cells within each cell type. This process transformed the gene expression of each cell type into a 1x30 vector. We then conducted a Pearson correlation analysis on the representative vectors of these nine cell types, obtaining a correlation between PRISM data and single-cell transcriptome data (Fig. 4c, Extended Data Fig. 24b, Extended Data Fig. 29b).

Spatial analysis. Spatial-related quantitative analysis in mouse embryos involved coarse-grained spatial correlation and intracellular interaction analyses. For coarse-grained spatial correlation analysis, annotated cells were spatially sub-sampled to proper bin size (~ 200 μm x 200 μm). Pearson correlation was calculated across each bin and cell type to analyze the co-occurrence

between cell types (Fig. 2g). For intracellular interaction analysis, the nearest neighbor cell for each cell was first identified through the KD-Tree search, and such neighbor cell pair counts between any two cell types were calculated. Each pair was normalized by dividing the total cell numbers in each cell type, eliminating the bias caused by the total cell count difference (*e.g.*, $muscle_{neuron}$ / muscle, $muscle_{neuron}$ refers to the number of muscle cells whose nearest neighbour cell are neurons) (Fig. 2h). This approach can reveal the interaction between cell types from a direct-interaction perspective.

Spatial-related quantitative analysis in stacked-3D HCC data involved the assessment of neighborhood enrichment, evaluating cell type distribution with respect to specific orientation, and examining cell type diversity across various Regions of Interest (ROIs) (Fig. 5e). The neighborhood enrichment analysis was primarily conducted utilizing the Squidpy framework[49]. Following the loading of three-dimensional cell centroid coordinates, we initially constructed a spatial neighbor graph with default parameters. Subsequently, we applied the Squidpy function 'gr.nhood_enrichment,' which executed a 1000-round permutation test to extract valuable enrichment insights. The results of this analysis were visually represented through the generation of an enrichment heatmap employing the 'pl.nhood_enrichment' function. To succinctly assess the 'barrier effect' of CAF on various cell types, we evaluated cell type distribution in specific orientations by projecting cells of each type onto a line based on their centroids and quantifying the cell density along this line (Fig. 5f). For the investigation of cell type variety in different ROIs, we systematically selected five ROIs and quantified the composition of immune and non-immune cell types within each of these regions. A Chi-squared test was subsequently applied to elucidate the extent of variation, and the ratio of observed cells to expected cells was calculated to provide a detailed depiction of cell enrichment and sparsity, with particular emphasis on immune cell subtypes (Fig. 5j).

Subcellular analysis. Given the coordinates of each RNA spot, the nucleus centroid, and the nucleus region, we profiled the RNA subcellular distribution inside and outside the nucleus, along with its polarity, which exhibited certain specificities on the tissue or cell type. The chi-square test was used to analyze the relation of different variables. The variables in our experiment are RNA distribution and tissue or cell type. To further represent the direction of the relation, we calculate the Ratio of Observed to Expected (Roe) of annotated RNA molecules. This allowed us to observe under which classification conditions RNA "clusters", thereby determining its distribution characteristics.

For the analysis of RNA distribution inside and outside the nucleus, our analytical approach relied on the classification of RNA molecules based on four parameters for each RNA molecule: tissue type (CTX, HP, TH, HY), cell type (including excitatory neurons, inhibitory neurons, glial cells, and their subtypes), gene type (among the thirty genes detected), and the RNA's cellular location (either inside or outside the nucleus). RNA was divided into several categories across four dimensions, and the Roe for each subdivided category was obtained by dividing the observed count of RNA molecules by the expected count of RNA molecules calculated for each category.

To focus on the differences in RNA distribution across different tissues and cell types, we reduced the dimensionality for the gene type dimension by averaging Roe values. Due to the enrichment effect of marker genes in corresponding cell types, marker genes contribute significantly when averaging across the thirty genes (Extended Data Fig. 30b). We then reduced the dimensionality for the nuclear/extranuclear dimension by using the extranuclear/nuclear ratio, ultimately obtaining the average relative abundance of RNA from thirty genes inside and outside the nucleus across different tissue and cell types (Fig. 6g).

In addition to the analysis of RNA distribution inside and outside the nucleus, we observed that the distribution of some RNAs within the cell exhibited polarity. To reflect the overall polarity of RNA, we first calculated the ratio of the distance between each RNA's centroid and the cell nucleus's

centroid to the average distance of each RNA to the cell nucleus's centroid for each cell. This ratio was used to indicate the overall polarity of RNA, and after compiling statistics for all cells, we obtained a quasi-normal distribution with a mean of 0.35. We designated the top 20% of this distribution as high polarity, the bottom 20% as low polarity, and the middle portion as moderate polarity, thus setting the thresholds for polarity.

Subsequently, to reflect the variability among different genes, we focused on genes with counts of more than 5 per cell for our analysis. We calculated their relative polarity using the aforementioned method and classified them into one of three categories (low, medium, high) based on the calculated thresholds.

Similar to the nuclear and extranuclear analysis, we classified each RNA molecule based on four parameters: tissue type (CTX, HP, TH, HY), cell type (excitatory neurons, inhibitory neurons, glial cells, and their subtypes), gene type (the thirty genes we detected), and the polarity of the gene within the cell (low, medium, high). RNA was categorized across four dimensions, and the Roe value for each category was obtained.

Likewise, since our main interest was in the differences in RNA polarity distribution across different tissue and cell types, we reduced the dimensionality for the gene type dimension by averaging different gene types. Due to the enrichment effect of marker genes in corresponding cell types, resulting in higher Roe values, thus making marker genes contribute significantly when averaging across the thirty genes (Extended Data Fig. 30b). We further reduced the dimensionality for the polarity dimension by using the high-to-low polarity ratio, ultimately obtaining the average RNA polarity distribution of thirty genes across different tissue and cell types (Fig. 6f).

It's worth noting that since we assigned a high, medium, or low polarity to each RNA molecule, the higher the RNA counts of a particular gene in a cell, the greater its impact on the polarity calculation result.

Acknowledgments

The authors thank Dr. Ye Liang of the State Key Laboratory of Membrane Biology and Dr. Liqin Fu of the School of Life Sciences and National Biomedical Imaging Center at Peking University for assistance with confocal microscopy imaging. The authors thank Lei Wang for experimental assistance. This work was supported by grants from National Natural Science Foundation of China Grants 21927802 (Y.H., J.W., P.F.), T2188102 (Y.H.), T2225005 (J.W.), Beijing Municipal Science and Technology Commission Grant Z221100007022003 (Y.H.), Ministry of Science and Technology Grant 2018YFA0800200 (J.W.), and Clinical Medicine Plus X - Young Scholars Project of Peking University, the Fundamental Research Funds for the Central Universities (C.Z.).

Author contributions

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

Website

A website is available to provide a clear understanding of PRISM's capabilities.
http://www.spatialprism.org

Competing interests

Y.H., T.C., S.Z., K.D., M.T., Z.L., M.J. and W.H. have applied for a patent related to this work.

Data and code availability

Single cell RNA-seq data were obtained from mousebrain.org, GEO accession number (GSE151530, GSE140228) and CNP0000650. Raw data of this study was deposited in zenodo including raw images (HCC: 10.5281/zenodo.12750711; MouseEmbryo: 10.5281/zenodo.12750725; MouseBrain: 10.5281/zenodo.12673246) and analysis related data (10.5281/zenodo.12755414). Source code is provided in Github repository at: https://github.com/huanglab111/PRISM_Code. A combined repository including analysis related data is also provided in github: https://github.com/huanglab111/PRISM ('dataset/processed' folder) for the ease of use.

References

1. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
2. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
3. Liu, Y. *et al.* High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **183**, 1665-1681.e18 (2020).
4. Liu, Y. *et al.* High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat Biotechnol* **41**, 1405–1409 (2023).
5. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat Methods* **18**, 9–14 (2021).
6. Cao, J. *et al.* Decoder-seq enhances mRNA capture efficiency in spatial RNA sequencing. *Nat Biotechnol* (2024) doi:10.1038/s41587-023-02086-y.
7. Liu, L. *et al.* Spatiotemporal omics for biology and medicine. *Cell* **187**, 4488–4519 (2024).
8. Schott, M. *et al.* Open-ST: High-resolution spatial transcriptomics in 3D. *Cell* **0**, (2024).
9. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Science* **381**, eabq4964 (2023).
10. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of Single RNA Transcripts in Situ. **280**, (1998).
11. Wang, F. *et al.* RNAscope. *The Journal of Molecular Diagnostics* **14**, 22–29 (2012).
12. 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 (2014).
13. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
14. Codeluppi, S. *et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat Methods* **15**, 932–935 (2018).
15. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
16. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19490–19499 (2019).
17. Wang, Y. *et al.* EASI-FISH for thick tissue defines lateral hypothalamus spatio-molecular organization. *Cell* **184**, 6361-6377.e24 (2021).

18. He, S. *et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat Biotechnol* **40**, 1794–1806 (2022).
19. Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. In situ detection and genotyping of individual mRNA molecules. *Nat Methods* **7**, 395–397 (2010).
20. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* **10**, 857–860 (2013).
21. Lee, J. H. *et al.* Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* **343**, 1360–1363 (2014).
22. Wang, X. *et al.* Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
23. Chen, X. *et al.* High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing. *Cell* **179**, 772-786.e19 (2019).
24. Gyllborg, D. *et al.* Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. *Nucleic Acids Research* **48**, (2020).
25. Sountoulidis, A. *et al.* SCRINSHOT enables spatial mapping of cell states in tissue sections with single-cell resolution. *PLoS Biol* **18**, e3000675 (2020).
26. Chang, T. *et al.* Rapid and signal crowdedness-robust in situ sequencing through hybrid block coding. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2309227120 (2023).
27. Shi, H. *et al.* Spatial atlas of the mouse central nervous system at molecular resolution. *Nature* **622**, 552–561 (2023).
28. Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat Methods* **9**, 743–748 (2012).
29. Vu, T. *et al.* Spatial transcriptomics using combinatorial fluorescence spectral and lifetime encoding, imaging and analysis. *Nat Commun* **13**, 169 (2022).
30. Wei, L. *et al.* Super-multiplex vibrational imaging. *Nature* **544**, 465–470 (2017).
31. Hong, F. *et al.* Thermal-plex: fluidic-free, rapid sequential multiplexed imaging with DNA-encoded thermal channels. *Nat Methods* **21**, 331–341 (2024).
32. Deng, R. *et al.* DNA-Sequence-Encoded Rolling Circle Amplicon for Single-Cell RNA Imaging. *Chem* **4**, 1373–1386 (2018).
33. Le, T. N. *et al.* GABAergic Interneuron Differentiation in the Basal Forebrain Is Mediated through Direct Regulation of Glutamic Acid Decarboxylase Isoforms by *Dlx* Homeobox Transcription Factors. *J. Neurosci.* **37**, 8816–8829 (2017).
34. Zhou, T. *et al.* Spatiotemporal Characterization of Human Early Intervertebral Disc Formation at Single-Cell Resolution. *Advanced Science* **10**, 2206296 (2023).
35. Nóbrega-Pereira, S. *et al.* Postmitotic Nkx2-1 controls the migration of telencephalic interneurons by direct repression of guidance receptors. *Neuron* **59**, 733 (2008).
36. Striedter, G. F. & Northcutt, R. G. The Independent Evolution of Dorsal Pallia in Multiple Vertebrate Lineages. *Brain Behavior and Evolution* **96**, 200–211 (2021).
37. Kaiser, K. *et al.* WNT5A is transported via lipoprotein particles in the cerebrospinal fluid to regulate hindbrain morphogenesis. *Nat Commun* **10**, 1498 (2019).
38. Zhang, Q. *et al.* Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. *Cell* **179**, 829-845.e20 (2019).
39. Sun, Y. *et al.* Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. *Cell* **184**, 404-421.e16 (2021).
40. Ma, L. *et al.* Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. *Journal of Hepatology* **75**, 1397–1408 (2021).

41. Zheng, C. *et al.* Transcriptomic profiles of neoantigen-reactive T cells in human gastrointestinal cancers. *Cancer Cell* **40**, 410-423.e7 (2022).
42. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-Convex Polygons. in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (eds. Frangi, A. F., Schnabel, J. A., Davatzikos, C., Alberola-López, C. & Fichtinger, G.) vol. 11071 265–273 (Springer International Publishing, Cham, 2018).
43. Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy. in *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)* 3655–3662 (IEEE, Snowmass Village, CO, USA, 2020). doi:10.1109/WACV45572.2020.9093435.
44. Smith, K. *et al.* CIDRE: an illumination-correction method for optical microscopy. *Nat Methods* **12**, 404–406 (2015).
45. Chalfoun, J. *et al.* MIST: Accurate and Scalable Microscopy Image Stitching Tool with Stage Modeling and Error Minimization. *Sci Rep* **7**, 4988 (2017).
46. Lionnet, T. *et al.* A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat Methods* **8**, 165–170 (2011).
47. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).
48. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *JOSS* **3**, 861 (2018).
49. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat Methods* **19**, 171–178 (2022).