Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+

Chee-Huat Linus Eng[1], **Michael Lawson**[2], **Qian Zhu**[3], **Ruben Dries**[3], **Noushin Koulena**[2], **Yodai Takei**[2], **Jina Yun**[2], **Christopher Cronin**[2], **Christoph Karp**[2], **Guo-Cheng Yuan**[3], and **Long Cai**[2]

[1.]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena USA 91125

[2.]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena USA 91125

[3.]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard T.H.Chan School of Public Health, Boston, MA 02215, USA

Abstract

Imaging the transcriptome *in situ* with high accuracy has been a major challenge in single cell biology, particularly hindered by the limits of optical resolution and the density of transcripts in single cells[1–5]. Here, we demonstrate seqFISH+, that can image the mRNAs for 10,000 genes in single cells with high accuracy and sub-diffraction-limit resolution, in the mouse brain cortex, subventricular zone, and the olfactory bulb, using a standard confocal microscope. The transcriptome level profiling of seqFISH+ allows unbiased identification of cell classes and their spatial organization in tissues. In addition, seqFISH+ reveals subcellular mRNA localization patterns in cells and ligand-receptor pairs across neighboring cells. This technology demonstrates the ability to generate spatial cell atlases and to perform discovery-driven studies of biological processes *in situ*.

Spatial genomics, the analysis of the transcriptome and other genomic information directly in the native context of tissues, is crucial to many fields in biology, including neuroscience and developmental biology. Pioneering work in single molecule Fluorescence *in situ*

Hybridization (smFISH) showed that individual mRNA molecules could be accurately detected in cells[6,7]. Development of sequential FISH (seqFISH) to impart a temporal barcode on RNAs through multiple rounds of hybridization allowed many molecules to be multiplexed[1–3]. Recently, we showed that seqFISH scales to the genome level *in vitro*[8] and for nascent transcription active sites[9].

However, the major challenge preventing global profiling mRNA in cells is the optical density of transcripts in cells: each mRNA occupies a diffraction limited spot in the image and there are tens to hundreds of thousands of mRNAs per cell depending on the cell type. Thus, optical crowding prevents mRNAs from being resolved and has bottlenecked all implementations of spatial profiling experiments[3–5]. For example, *in situ* sequencing methods, detected only ~500 transcripts per cell[4,5,10] because of the lower efficiency and larger dot size of rolling circle amplification, whereas seqFISH detected thousands of transcripts per cell[3]. We have previously proposed to combine super-resolution microscopy with FISH[11] to overcome this crowding problem. However, existing super-resolution localization microscopy[12,13] relies on detection of single dye molecules, which emit limited number of photons and only work robustly in optically thin (<1 μm) samples.

To enable discovery-driven approaches *in situ*, it is essential to scale up the spatial multiplexed methods to the genome level. To date, spatial methods have always relied on existing genomics methods, such as scRNAseq, to identify target genes, and serve to only map cell types identified from scRNAseq. At the level of hundreds and even a thousand genes, spatial methods cannot be used as *de novo* discovery-driven tool, which is a major drawback of the technology. In addition, many genes are expressed in a spatially dependent fashion independent of cell types[14] that is not recovered in the dissociated cell analysis.

Here, we demonstrate seqFISH+, which achieves super-resolution imaging and multiplexing of 10,000 genes in single cells using sequential hybridizations and imaging with a standard confocal microscope. The key to seqFISH+ is expanding the barcode base palette from 4–5 colors, as used in seqFISH[1,3] and *in situ* sequencing experiments[4,5], to a much larger palette of "pseudocolors" (Figure 1a) achieved by sequential hybridization. By using 60 pseudocolor channels, we effectively dilute mRNA molecules into 60 separate images and allows each mRNA dot to be localized below the diffraction limit[12,15,16] before recombining the images to reconstruct a super-resolution image. We separate the 60 pseudocolors into 3 fluorescent channels (Alexa 488, Cy3b and Alexa 647) and generate barcodes only within each channel to avoid chromatic aberrations between channels. 20^3=8000 genes can be barcoded in each channel for a total of 24,000 genes by repeating this pseudocolor imaging 4 times with one round used for error-correction[3].

As imaging time is the main bottleneck in spatial transcriptomics experiments, seqFISH+ is 8-fold faster in imaging time compared to implementing seqFISH with expansion microscopy[17] (Figure 1b). An equivalent 60-fold expansion of the sample would require 4 colors × 8 barcoding rounds × 60 volume expansion = 1920 images per field of view (FOV) to cover 4^7=16,384 genes. In contrast, seqFISH+ acquires 60 pseudocolors × 4 barcoding rounds = 240 images per FOV to cover 24,000 genes, an 8-fold reduction in imaging time.

Furthermore, a large number of pseudocolors and a shorter barcode (4 units) decreases the errors that accumulate over barcode rounds.

To demonstrate transcriptome level profiling in cells, we first applied seqFISH+ to cleared NIH3T3 fibroblast cells (Figure 1c, Extended Data Figure 1,2)[18–20]. We randomly selected 10,000 genes while avoiding highly abundant housekeeping genes, such as ribosomal proteins. These 10,000 genes add up to a total of >125,000 FPKM values with a wide range of expression levels from 0 to 995.1 FPKM. All 24,000 genes in the fibroblast transcriptome add up to ~420,000 FPKM[21], only a 3 fold higher density from the 10,000 gene experiment, which can be accommodated with the current scheme, or with more channels or pseudocolors.

Overall, 35,492±12,222 (mean±s.d.) transcripts are detected per cell (Figure 2a). The 10,000 seqFISH+ data are highly reproducible and strongly correlated with RNA-seq (R=0.80)[21], RNA SPOTs (R=0.80)[8], and smFISH (R=0.87) (Figure 2b–d, Extended Data Figure 3a,b). Each of the three fluorescent channels was decoded independently and correlated well with RNA-seq and smFISH (Extended Data Figure 3a,c). The false positive rate per cell is 0.22±0.07 (mean±s.d.) per barcode (Extended Data Figure 3d,e). Comparison with 60 genes from smFISH showed that the seqFISH+ detection efficiency is 49%, which is highly sensitive compared to single cell RNAseq.

seqFISH+ allows us to visualize the subcellular localization patterns for tens of thousands of RNA molecules *in situ* in single cells. Three major clusters were observed to be nuclear/peri-nuclear, cytoplasmic and protrusion enriched. Many new protrusion localized genes are found in addition to the ones identified previously[22,23]. We further observed three distinct subclusters in the perinuclear/nuclear localized transcripts with genes in each of these subclusters enriched in distinct functional roles (Extended Data Figure 3f–j).

To demonstrate seqFISH+ works robustly in tissues, we used the same 10,000 gene probe set to image cells in the mouse brain cortex, the sub-ventricular zone (SVZ) (Figure 3a), and the olfactory bulb in two separate brain sections. We collected 10,000-gene-profiles for 2963 cells (Figure 3b–e), covering an area of approximately 0.5 mm^2. In the cortex, cells contained on average 5615±3307 (mean±s.d.) transcripts from 3338±1489 (mean±s.d.) detected genes (Extended Data Figure 4a,b). We imaged only a single z optical plane (0.75 μm) to save imaging time. Full 3D imaging of cells with seqFISH+ is available for 5–10x "deeper" sampling of the transcriptome.

With an unsupervised clustering analysis[24], the seqFISH+ cell clusters show clear layer structures (Figure 3h) and are strongly correlated to the clusters in a scRNAseq[25] dataset (Methods, Extended Data Figure 4c–f, 5). Similar layer patterns are observed with Hidden Markov Random Field (HMRF) analysis[14] where the expression patterns of neighboring cells were taken into account (Extended Data Figure 4g–i, 6).

With the seqFISH+ data, we can explore the subcellular localization patterns of 10,000 mRNAs directly in the brain in a cell type specific fashion (Supplementary Table 3). In many cells types, the transcripts for *Snrnp70*, a small nuclear riboprotein, and *Nr4a1*, a nuclear receptor, are found in the nuclear/perinuclear regions. In contrast, *Atp1b2*, a Na+/K+

ATPase, and *Kif5a*, a kinesin, are observed to be near the cell peripheries in many cell types including excitatory, inhibitory neurons as well as glia cells. In addition, many transcripts in astrocytes, such as *Gja1* and *Htra1*, localize to the cell periphery and processes, which we confirmed by smFISH (Figure 3f,g, Extended Data Figure 7).

We next explored the spatial organization of the SVZ. We identified neural stem cells (NSCs, Clusters 8,16) expressing astrocyte markers *Gja1* and *Htra1*, transit-amplifying progenitors (TAPs, Cluster 15) expressing *Ascl1, Mcm5* and *Mki67*, and neuroblasts (NBs) expressing *Dlx1* and *Sp9*, consistent with previous studies[26]. We further quantified the spatial organization of the different cell types in the SVZ (Figure 4a, Extended Data Figure 8), and found that class 12 and 17 neuroblasts are preferentially in contact, whereas TAP cells tend to associate with other TAP cells. It would be exciting to further investigate the RNA velocity trajectories[27] of these cells *in situ* with intron seqFISH[9] as well as their lineage relationships with MEMOIR[28].

Next, we examined the spatial organization of the olfactory bulb. Our clustering analysis revealed distinct classes of GABAergic interneurons, olfactory ensheathing cells (OECs), astrocytes, microglia, and endothelial cells (Figure 4b,c), consistent with literature[29]. In the granule cell layer (GCL) at the center of the olfactory bulb, several cell classes are observed, with an interior core consisting of immature neuroblast-like cells expressing *Dlx1* and *Dlx2* encased by a distinct outer layer of the GCL composed of more mature interneurons (Figure 4b and Extended Data Figure 9,10). An excitatory cluster of cells expressing *Reln, Slc17a7* are observed in the mitral cell layer (MCL) as mitral cells and in the external plexiform layer (EPL) and glomerulus as tufted cells. We also found several clusters of *Th*+ dopaminergic neurons (Figure 4b–d, Supplementary Table 2) which were previously not known. For example, Cluster 1 cells express both *Vgf*, a neuropeptide, as well as tyrosine hydroxylase (*Th*), and are distributed both in the glomerulus and the GCL. Similarly, *Trh* is enriched in a distinct set of *Th*+ cells (Cluster 3), which are predominantly in the glomerulus, whereas Clusters 5 and 22 dopaminergic neurons are in the GCL. We validated these clusters by smFISH imaging (Figure 4d, Extended Data Figure 9,10).

Finally, we analyzed ligand-receptor pairs that are enriched in neighboring cells, which are not available in the dissociated cell analysis. These proposed potential cell-cell interactions are hypothesized on the basis of mRNA and not protein. In endothelial cells adjacent to microglia in the olfactory bulb, Endoglin (*Eng*, a type III TGF-β receptor) and Activin A-receptor (*Acvrl1 or Alk1*, a type I TGF-β receptor) mRNAs are expressed, with TGF-β ligand (*Tgfb1*) mRNA expressed by the microglia. Microglia-endothelial neighbor cells express, *Lrp1 (Tgfbr5)* and *Pdgfb*, in the cortex, indicating that signaling pathways may be used in a tissue specific fashion. Beyond ligand receptor interactions, we found broadly that gene expression patterns in a particular cell type is highly dependent on the local tissue context of neighboring cells (Figure 4e,f, Supplementary Table 4).

These experiments demonstrate that seqFISH+ can robustly profile transcriptomes in tissues, overcoming optical crowding and removing the last conceptual roadblock in generating spatial single cell atlases in tissues. seqFISH+ provides 10-fold or more improvement over existing methods in the number of mRNAs profiled and the total number of RNA barcodes

detected per cell. seqFISH+ also allows super-resolved imaging with commercial confocal microscopes and can be generalized to chromosome[30] and protein imaging.

With the genome coverage and spatial resolution of seqFISH+, it is now possible to perform discovery-driven studies directly *in situ*. In particular, elucidating signaling interactions between cells is a crucial first step towards understanding developmental processes and cell fate decisions, along with explorations of the combinatorial signaling logic[21]. Lastly, the genomics coverage of seqFISH+ will allow discovery of novel targets that are cell type specific in disease samples as well as enable precise spatial-genomics and single-cell based diagnostics test.

Methods

Data Reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments outcome assessment.

Experiment Design

Primary probe design.—Gene-specific primary probes were designed as previously described with some modifications[8]. To obtain probe sets for 10,000 different genes, 28-nt sequences of each gene were extracted first using the exons from within the CDS region. For genes that did not yield enough target sequences from the CDS region, exons from both the CDS and UTRs were used. The masked genome and annotation from UCSC were used to look up the gene sequences. Probe sequences were required to fall within the GC content in the range of 45–65%. Any probe sequences that contained five or more consecutive bases of the same kind were dropped. Any genes which do not achieve a minimum number of 24 probes were dropped. A local BLAST query was run on each probe against the mouse transcriptome to ensure specificity. BLAST hits on any sequences other than the target gene with a 15-nt match were considered off targets. ENCODE RNA-seq data across different mouse samples were used to generate an off-target copy number table. Any probe that hit an expected total off-target copy number exceeding 10,000 FPKM was dropped to remove housekeeping genes, ribosomal genes and very highly expressed genes. To minimize cross-hybridization between probe sets, a local BLAST database was constructed from the probe sequences and probes with hits of 17-nt or longer were removed by dropping the matched probe from the larger probe set.

Readout probe design.—15-nt readout probes were designed as previously described[9]. Briefly, a set of probe sequences was randomly generated with the combinations of A, T, G, or C nucleotides. Readout probe sequences within a CG content range of 40–60% were selected. We BLAST against the mouse transcriptome to ensure the specificity of the readout probes. To minimize cross-hybridization of the readout probes, any probes with 10-contiguously matching sequences between readout probes were removed. The reverse complements of these readout probe sequences were included in the primary probes according to the designed barcodes.

Primary probe construction.—Primary probes were ordered as oligoarray complex pools from Twist Bioscience and were constructed as previously described with some modifications[8]. Briefly, limited PCR cycles were used to amplify the designated probe sequences from the oligo complex pool. Then, the amplified PCR products were purified using QIAquick PCR Purification Kit (28104; Qiagen) according to the manufacturer's instructions. The PCR products were used as the template for in vitro transcription (E2040S; NEB) followed by reverse transcription (EP7051; Thermo Fisher) with the forward primer containing a uracil nucleotide[31]. After reverse transcription, the probes were subjected to 1:30 dilution of Uracil-Specific Excision Reagent (USER) Enzyme (N5505S; NEB) treatment to remove the forward primer by cleaving off the uracil nucleotide next to it for ~24 hours at 37°C. Since the reverse complement of T7 sequences was used as the reverse primer, the final probe length in this probe set was ~93-nt. Then, the ssDNA probes were alkaline hydrolyzed by 1 M NaOH at 65°C for 15 minutes to degrade the RNA templates, followed by 1 M acetic acid neutralization. Next, to clean up the probes, we performed ethanol precipitation to remove stray nucleotides, phenol-chloroform extraction to remove protein, and Zeba Spin Desalting Columns (7K MWCO) (89882, Thermo Fisher) to remove any residual nucleotides and phenol contaminants. Then, the probes were mixed with 2 μM of Locked Nucleic Acid (LNA) polyT15 and 2 μM of LNA polyT30 before speed-vac to dry powder and resuspended in primary probe hybridization buffer comprised of 40% formamide (F9027, Sigma), 2x SSC (15557036, Thermo Fisher), and 10% (w/v) Dextran Sulfate (D8906; Sigma). The probes were stored at −20°C until use.

Readout probe synthesis.—15-nt readout probes were ordered from Integrated DNA Technologies (IDT) as 5' amine modified[9]. The construction of readout probe was similar to previously described. Briefly, 5 nmoles of DNA probes were mixed with 25 μg of Alexa Fluor 647 NHS ester or Cy3B or Alexa Fluor 488 NHS ester in 0.5 M sodium bicarbonate buffer containing 10% DMF. The reaction was allowed to go for at least 6 hours at 37°C. Then, the DNA probes were subjected to ethanol precipitation, HPLC purification, and column purification to remove all contaminants. Once resuspended in water, the readout probes were quantified using Nanodrop and a 500 nM working stock was made. All the readout probes were kept at −20°C.

Coverslip functionalization.—For cell culture experiment, the coverslips were cleaned with a plasma cleaner at HIGH (PDC-001, Harrick Plasma) for 5 minutes followed by the immersion in 1% bind-silane solution (GE; 17-1330-01) made in pH3.5 10% (v/v) acidic ethanol solution for 30 minutes at room temperature. Then the coverslips were rinsed with 100% ethanol 3 times, and heat-dry in an oven for > 90°C for 30 minutes. Next, the coverslips were treated with 100 μg/uL of Poly-D-lysine (P6407; Sigma) in water for >1 hour at room temperature. Followed by 3 times rinsing with water, the coverslips were air-dried and kept at 4°C for no longer than 2 weeks. For mouse brain slices experiment, the coverslips were cleaned by 1M HCl at room temperature for 1 hour, rinsed with water once, and followed by 1M NaOH solution treatment at room temperature for 1 hour. Then, the coverslips were rinsed three times with water, before immersion in 1% bind-silane solution for 1 hour at room temperature. The remaining steps are the same as the coverslip functionalization for cell culture.

seqFISH+ encoding strategy.—We separate the 60 pseudocolors into 3 fluorescent channels (Alexa 488, Cy3b and Alexa 647) equally. In each channel, the 20- pseudocolor imaging was repeated 3 times hence achieving $20^3 = 8000$ genes barcoding capacity. We did an extra round of pseudocolor imaging to obtain error-correctable barcodes, an error-correction scheme which we had previously introduced[3]. Thus, we obtained 8000 error-correctable barcodes × 3 fluorescent channels = 24,000 error-correctable barcoding capacity in total. One can easily use more fluorescent channels and/or more pseudocolors to achieve greater dilution of the mRNA density per imaging round. In this experiment, we encoded 3333, 3333, and 3334 genes in each of the fluorescent channels. This pseudocolor scheme evolved from the one used in RNA SPOTs[8] and intron seqFISH[9] by eliminating chromatic aberration and dramatically diluting the density to achieve profiling of mRNA at the transcriptome level *in situ*.

To visualize the different transcripts, 24 "primary" probes were designed against each target mRNA. The primary probes contain overhang sequences that code for the 4-unit base-20 barcode unique to each gene. Hybridization with fluorophore labeled "readout" probes allows the readout of these barcodes and fluorescently labels the subset of genes that contain the corresponding sequences. All of the genes are sampled every 20 rounds of readout hybridization and collapsed into super-resolved images. A total of 80 rounds of hybridizations enumerate the 4-unit barcode for each gene. Each round of stripping and readout hybridization is fast and completed in minutes.

After primary probes hybridization, the samples were subjected to hydrogel embedding and clearing before seqFISH+ imaging. The details are available on *cell culture experiment*, *tissue slices experiment*, and *seqFISH+ imaging*.

Cell culture experiment.—NIH/3T3 cells (ATCC) were cultured as previously described[8] on the functionalized coverslips until ~80–90% confluence. Then the cells were washed with 1x PBS once, fixed with freshly made 4% formaldehyde (28906; Thermo Fisher) in 1x PBS (AM9624, Invitrogen) at room temperature for 10 minutes. The fixed cells were permeabilized with 70% ethanol for 1 hour at room temperature. The cell samples were dried and the 10,000 gene probes (~1 nM per probe for 24 probes per gene) were hybridized by spreading out using another coverslip. The hybridization was allowed to proceed for ~36–48 hours in a humid chamber at 37°C. We found hybridization for 48 hours yielded slightly brighter signals. After hybridization, the samples were washed with 40% formamide in 2x SSC at 37°C for 30 minutes, followed by 3 times rinsing with 1 mL 2x SSC. Next, the cell samples were incubated with 1:1000 dilution of Tetraspeck beads in 2x SSC at room temperature for 5–10 minutes. The density of the beads can be easily adjusted by varying the dilution factor or incubation time. Then, the samples were rinsed with 2x SSC and incubated with degassed 4% acrylamide (1610154; Bio-Rad) solution in 2x SSC for 5 minutes at room temperature. To initiate polymerization, the 4% acrylamide solution was aspirated, then 10 μL of 4% hydrogel solution containing 4% acrylamide (1:19), 2x SSC, 0.2% ammonium persulfate (APS) (A3078; Sigma) and 0.2% N,N,N′,N′-Tetramethylethylenediamine (TEMED) (T7024; Sigma) was dropped on the sample, and sandwiched by a coverslip functionalized by GelSlick (Lonza;50640). The polymerization step was allowed to happen at room temperature for 1 hour in a homemade nitrogen gas chamber. After that, the two

coverslips were gently separated, and the excess gel was cut away with a razor. A custom-made flow cell (RD478685-M; Grace Bio-labs) was attached to the coverslips covering the region of cells embedded in hydrogel. The hydrogel embedded cell samples were cleared as previously described for >1 hour at 37°C[19]. The digestion buffer consists of 1:100 Proteinase K (P8107S; NEB), 50 mM pH 8 Tris HCl (AM9856; Invitrogen), 1 mM EDTA (15575020; Invitrogen), 0.5% Triton-X 100, and 500 mM NaCl (S5150, Sigma). Then, the samples were rinsed with 2x SSC multiple times and subjected to Label-IT modification (1:10) (MIR 3900; Mirus Bio) at 37°C for 30 minutes. After that, the cell samples were post-fixed with 4% PFA in 1x PBS to stabilize the DNA, RNA, and the overall cell sample for 15 mins at room temperature. The reaction was quenched by 1 M pH8.0 Tris HCl at room temperature for 10 minutes. The cell samples were either imaged immediately or kept in 4x SSC supplemented with 2 U/μL of SUPERase In RNase Inhibitor (AM2696; Invitrogen) at 4°C for no longer than 6 hours.

Animals.—All animal care and experiments were carried out in accordance to Caltech Institutional Animal Care and Use Committee (IACUC) and NIH guidelines. Wild-type mice C57BL/6J P23 (male) and P40 (male) were used for the cortex and olfactory bulb seqFISH+ experiments respectively. For smFISH experiments, adult wild-type mice C57BL/6J aged 10 weeks (female) were used for the RNA localization experiment in the cortex and ligand-receptor interaction experiment in the olfactory bulb. For cell clusters validation in the olfactory bulb, section from P40 mice was used.

Tissue slices experiment.—Brain extraction was performed as previously described[3]. In brief, mice were perfused for 8 minutes with perfusion buffer (10 U/ml heparin, 0.5% NaNO2 (w/v) in 0.1 M PBS at 4°C). Mice were then perfused with fresh 4% PFA in 0.1 M PBS buffer at 4°C for 8 minutes. The mouse brain was dissected out of the skull and immediately placed in a 4% PFA buffer for 2 hours at room temperature under gentle mixing. The brain was then immersed in 4°C 30% RNAse-free Sucrose (Amresco 0335–2.5KG) in 1x PBS until the brain sank. After the brain sank, the brain was frozen in a dry ice of isopropanol bath in OCT media and stored at −80°C. 5 μm sections were cut using a cryotome and immediately placed on the functionalized coverslips. The thin tissue slices were stored at −80°C. To perform hybridization on the tissue slices, the tissue slices were first permeabilized in 70% ethanol at 4°C for >1 hour. Then, the tissue slices were cleared with 8% SDS (AM9822; Invitrogen) in 1x PBS for 30 minutes at room temperature. Primary probes were hybridized to the tissue slices by spreading out the hybridization buffer solution with a coverslip. The hybridization was allowed to proceed for ~60 hours at 37°C. After primary probe hybridization, the tissue slices were washed with 40% formamide at 37°C for 30 minutes. After rinsing with 2X SSC 3 times and 1X PBS once, the sample was subjected to 0.1mg/mL Acryoloyl-X SE (A20770; Thermo Fisher) in 1X PBS treatment for 30 minutes at room temperature. After that, the tissue slices were incubated with 4% acrylamide (1:19 crosslinking) hydrogel solution in 2X SSC for 30 minutes at room temperature. Then the hydrogel solution was aspirated and 20 μL of 4% hydrogel solution containing 0.05% APS and 0.05% TEMED in 2x SSC was dropped onto the tissue slice and sandwiched by Gel-Slick functionalized slide. The samples were transferred to 4°C in a homemade nitrogen gas chamber for 30 minutes before transferring to 37°C for 2.5 hours to

complete polymerization. After polymerization, the hydrogel embedded tissue slices were cleared with digestion buffer as mentioned above, except it includes 1% SDS, for >3 hours at 37°C. After digestion, the tissue slices were rinsed by 2X SSC multiple times and subjected to 0.1mg/mL Label-X modification for 45 minutes at 37°C. The preparation of Label-X stock was as previously described[19]. To further stabilize the DNA probes, RNA molecules, and the tissue slices overall structure, the tissue slices were re-embedded in hydrogel solution as the previous step, except the gelation time can be shortened to 2 hours. The tissue slice samples were either imaged immediately or kept in 4X SSC supplemented with 2 U/μL of SUPERase In RNase Inhibitor at 4°C for no longer than 6 hours.

seqFISH+ Imaging.—Imaging platform and automated fluidics delivery system were similar to those previously described with some modifications. In brief, the flow cell on the sample was first connected to the automated fluidics system. Then the region of interests(ROI) was registered using nuclei signals stained with 10 μg/mL of DAPI (D8417; Sigma). For cell culture experiments, blank images containing beads only were first imaged before the first round of serial hybridization. Each serial hybridization buffer contained three unique sequences with different concentrations of 15-nt readouts conjugated to either Alexa Fluor 647(50 nM), Cy3B(50 nM) or Alexa Fluor 488(100 nM) in EC buffer made from 10% Ethylene Carbonate (E26258; Sigma), 10% Dextran Sulfate (D4911; Sigma), 4X SSC and 1:100 dilution of SUPERase In RNase Inhibitor. The 100 μL of serial hybridization buffers for 80 rounds of seqFISH+ imaging with a repeat for round 1 (in total 81 rounds) were pipetted into a 96 well-plate. During each serial hybridization, the automated sampler will move to the well of the designated hyb buffer and flow the 100 μL hyb solution through a multichannel fluidic valves (EZ1213-820-4; IDEX Health & Science) to the flow cell (required ~25 μL) using a syringe pump (63133–01, Hamilton Company). The serial hyb solution was incubated for 17 minutes for cell culture experiments and 20 minutes for tissue slice experiments at room temperature. After serial hybridization, the sample was washed with ~300 μL of 10% formamide wash buffer (10% formamide and 0.1% Triton X-100 in 2X SSC) to remove excess readout probes and non-specific binding. Then, the sample was rinsed with ~200 μL of 4X SSC supplemented with 1:1000 dilution of SUPERase In RNase Inhibitor before stained with DAPI solution (10 μg/mL of DAPI, 4X SSC, and 1:1000 dilution of SUPERase In RNase Inhibitor) for ~15 seconds. Next, an anti-bleaching buffer solution made of 10% (w/v) glucose, 1:100 diluted catalase (Sigma C3155), 0.5 mg/mL Glucose oxidase (Sigma G2133), 0.02 U/μL SUPERase In RNase Inhibitor, 50 mM pH8 Tris-HCl in 4x SSC was flowed through the samples. Imaging was done with the microscope (Leica, DMi8) equipped with a confocal scanner unit (Yokogawa CSU-W1), a sCMOS camera (Andor Zyla 4.2 Plus), 63 × oil objective lens (Leica 1.40 NA), and a motorized stage (ASI MS2000). Lasers from CNI and filter sets from Semrock were used. Snapshots were acquired with 0.35 μm z steps for two z slices per FOV across 647-nm, 561-nm, 488-nm and 405-nm fluorescent channels. After imaging, stripping buffer made from 55% formamide and 0.1% Triton-X 100 in 2x SSC was flowed through for 1 minute, followed by an incubation time of 1 minute before rinsing with 4X SSC solution. In general, the 15-nt readouts were stripped off within seconds, and a 2-minute wash ensured the removal of any residual signal. The serial hybridization, imaging, and signal extinguishing steps were repeated for 80-rounds. Then, stainings buffer for segmentation purpose consists of 10

μg/mL of DAPI, 50nM LNA T20-Alexa 647, and 1: 100 dilution of Nissl stainings (N21480; Invitrogen) in 1x PBS was flowed in and allowed to incubate for 30 mins at room temperature before imaging. The integration of automated fluidics delivery system and imaging was controlled by a custom written script in Micro-Manager[32]

smFISH.—Single molecule FISH(smFISH) experiments were done as previously described[8]. In brief, 60 genes were randomly chosen from the 10,000 gene list across a broad range of expression levels. The same probe sequences were used for these 60 genes, except each primary probe contained two binding sites of the readout probes. The fixed cells were hybridized with the primary probes(10nM/probes) in 40% hyb buffer (40% formamide, 10% Dextran Sulfate and 2x SSC) at 37°C for overnight. The sample was washed with 40% wash buffer for 30 minutes at 37°C and subjected to the same hydrogel embedding and clearing as the cell culture experiment before imaging. The imaging platform is the same as the one in seqFISH+ experiment. A single z-slice across hundreds of cells was imaged and the sum of the gene counts per cell was analyzed by using a custom written Matlab script. For smFISH experiments in the tissue, sample was hybridized with 10nM/probe in 40% hyb buffer at 37°C for >16 hours. The sample was washed with 40% wash buffer for 30 minutes at 37°C and subjected to the same hydrogel embedding and clearing as the tissue experiment before imaging. Since the imaging time is short, the Acryoloyl-X functionalization and post hydrogel anchoring steps were omitted. 5 z-slices with z-step of 1μm were taken across multiple FOVs with the imaging platform in the seqFISH+ experiment, except a 40x oil objective was used (Leica 1.40 NA). Images were background subtracted and maximum z-projected for clearer display of RNA dots.

Image Analysis

All image analysis was performed in Matlab. Unless a specific Matlab function is referenced, custom code was used.

Image Registration.—Each round of imaging included imaging with the 405-nm channel which included the DAPI stain of the cell along with imaging in the 647-nm, 561-nm and 488-nm channels of TetraSpeck beads' (T7279, Thermo Fischer) and seqFISH+ probes. In addition, a pre-hybridization image was used to find all beads before the readouts were hybridized. Bead locations were fit to a 2D Gaussian. An initial estimate of the transformation matrix between the DAPI image for each serial hybridization round and the only beads image was found using imregcorr (Matlab). Using this estimate transformation, the bead coordinates were transformed to each serial hybridization image, where the location of the bead was again fit to a 2D gaussian. A final transformation matrix between each hybridization image and the only beads image was then found by applying fitgeotrans (Matlab) to the sets of Gaussian fit bead locations. For the tissue samples no beads were used and registration was based on DAPI alone.

Image processing.—Each image was deconvolved, using a bead (7×7pixels) as an estimate for the point spread function. Cell segmentation was performed manually using ImageJ's ROI tool.

Barcode Calling.—The potential RNA signals were then found by finding local maxima in the image above with a predetermined pixel threshold in the registered and deconvolved images. Dot locations were then further resolved using radialcenter.m[33]. Once all potential points in all serial hybridizations of one fluorescent channel were obtained, they were organized by pseudocolor and barcoding round. Dots were matched to potential barcode partners in all other pseudo channels of all other barcoding rounds using a 1-pixel search radius (or for the tissue samples a 1.4-pixel search radius) to find symmetric nearest neighbors. Point combinations that constructed only a single barcode were immediately matched to the on-target barcode set. For points that matched to construct multiple barcodes, first the point sets were filtered by calculating the residual spatial distance of each potential barcode point set and only the point sets giving the minimum residuals were used to match to a barcode. If multiple barcodes were still possible, the point was matched to its closest on-target barcode with a hamming distance of 1. If multiple on target barcodes were still possible, then the point was dropped from the analysis as an ambiguous barcode. This procedure was repeated using each barcoding round as a seed for barcode finding and only barcodes that were called similarly in at least 3 out of 4 rounds were used in the analysis. The number of each barcode was then counted in each of the assigned cell areas and transcript numbers were assigned based on the number of on-target barcodes present in the cell. Centroids for each called barcode were also recorded and assigned to cells. The same procedure was repeated for 647, 561 and 488 channels. The remaining unused barcodes were used as an off-target evaluation by repeating the same procedure as described.

Data Analysis

RNA-seq/RNA SPOTs.—Pearson's r correlation was performed to compare seqFISH+ data to RNA-seq (GEO: GSE98674), RNA SPOTs[8], and smFISH measurement using Matlab or Python function.

Spatial clustering of genes for NIH3T3 cells.—The same barcode calling procedure described above was repeated without cell segmentation to remove the possibility of clipping potentially interesting regions of the cell. RNA locations were coarse grained to 10×10 pixels, resulting in a matrix of dimension total number of coarse grained pixels by the number of genes. Coarse pixels with no RNA were removed from the analysis. RNA with fewer than 10 copies per field of view were dropped. Genes were then correlated with Pearson's r correlation and hierarchical clustering was performed on the resulting correlation matrix. Clusters of less than 10 genes were dropped.

Hierarchical clustering of brain seqFISH+ data.—The 10,000 genes were divided into 3 approximately equal subsets (with 3334, 3333, and 3333 genes, respectively) based on the group in which genes are barcoded. Genes were normalized separately within each subset, by dividing the gene counts in per cell by the total counts per cell within each subset. We then multiplied the result by the scaling factor of 2,000 which is approximately the median count. Next, we selected the subset of cells that were in the cortex. We computed log (1+normalized counts).

To select genes for clustering, we first computed statistics for the following criteria for each gene: 1) number of cells with nonzero expression, 2) average gene expression of all cells, 3) average expression of top 5% cells with highest expression, 4) average of top 10% cells with highest expression, 5) average of top 2% cells with highest expression, 6) average gene expression of all nonzero cells. For each criterion, we selected the top 25% of genes that were ranked based on the criterion. We next obtained the union of all 6 gene lists forming an initial 3877 gene-set. The reasoning is that the union of genes would contain both genes needed to cluster common cell types (which would be expressed in a large population of cells, captured by criterion 2) and rare cell types (which would be expressed in a small population, captured by criteria 3, 4, 5). The 3877-gene expression data matrix was next transformed by z-scoring per cell and per gene. Principle component analysis (PCA) was performed and jackstraw procedure was adopted in order to further select the most relevant genes for clustering. Specifically, the jackstraw procedure[34] permutes the expression of a small number of genes in order to identify significant genes with significantly higher loading than permuted case (P<0.001). Using the top 9 components, we found a total of 1916 significant genes to be used for final clustering.

To this 1916-gene matrix we applied hierarchical clustering with Ward's linkage and with (1 - Pearson correlation) as the distance measure. Using the sigClust R package[35], which evaluates the significance of each branching in the dendrogram, we found significant tree splits and produced 10-cluster and 16-cluster annotations corresponding to different cluster granularity. Each split was significant according to sigClust FWER corrected P < 0.05. We further performed an additional round of clustering within the interneuron annotated clusters, repeated gene-selection procedure, and replaced the broad interneuron cluster with the subclusters. All together, we derived 13-cluster and 18-cluster annotations.

Unsupervised comparison with scRNAseq data.—Mouse visual cortex scRNAseq data was obtained from Tasic et al[25]. We used the cell-type annotations from the original study, representing 9 major, 22 fine, and 49 minor cell-types. For comparison, we focused on the 1857 genes that were commonly profiled by scRNAseq and seqFISH+ and processed the scRNAseq data in the same way as seqFISH+. The degree of similarity was evaluated by using the Pearson correlation (Extended Data Figure 4a).

Supervised mapping of cell types from scRNAseq to seqFISH+.—Cell-type mapping was done as described before[14]. Briefly, MAST[36] was used to identify differentially expressed genes across annotated cell types in Tasic *et. al.* scRNA-seq dataset, using P=0.005 as the cutoff. 1253 of the differentially expressed genes were also profiled by seqFISH+ therefore retained for cell-type mapping. Then, we performed a quantile-normalization on the expression vectors of each gene in both the seqFISH+, scRNA-seq data to normalize cross-platform differences[14]. Multi-class support-vector machine models were trained on the scRNAseq cell types using linear kernels, and setting the tuning parameter C to 1e-5, shown in Figure 3g. The cross-validation accuracy of prediction of the 22 annotated cell types was 91% with these 1253 differentially expressed genes.

Spatial gene identification.—Briefly, we computed a spatial score per gene as previously described[14]. Cells were divided into two sets based on gene g: L1, contains cells

with highest 90th percentile by expression, and L0, the remaining cells. The spatial score measures whether the cells in L1 are spatially adjacent to each other and is quantified by the silhouette coefficient. The silhouette coefficient was computed using the calc_silhouette_per_gene() function in the smfish Hmrf Python package[14] (https://bitbucket.org/qzhudfci/smfishhmrf-py), setting dissimilarity matrix to rank-transformed Euclidean distance, examine_top=0.1, permutation_test=True, and permutations=1000. Rank-transformed distance was computed with rank_transform_matrix() function with reverse=False, rbp_p=0.99 where rbp_p is a rank-weighting parameter. We select all spatial genes with significant silhouette coefficient (P<0.01 permutation test). To further enrich for spatial signals within these genes, we performed a PCA analysis, and then jackstraw procedure[34] to arrive at a set of 988 spatial genes significantly correlated to the principle components. We performed HMRF analysis on the top 100, 200, 400 of 988 genes.

Spatial domain identification via HMRF procedure.—HMRF is a probabilistic spatial clustering method that we developed previously to identify spatial domains based on spatial relationships and gene expression per cell. We constructed a neighborhood graph by adopting a fixed radius corresponding to top 1-percentile of pairwise physical distances between cells, resulting in an average of 5 neighbors per cell. HMRF was run with the following parameters: tolerance=1e-10, k=9, and convergence_error=1e-8. To search for an optimal value of beta, we scanned through all integer values between 2 and 100 and ran the HMRF model for each setting. The value that resulted in minimal change of log-likelihood was selected as the final beta.

Louvain clustering.—Unless specified, all functions of pre-processing and Louvain clustering was performed in Python using the package SCANPY[37]. We followed a standard procedure as suggested in the SCANPY reimplementation of Seurat's tutorial to analyze seqFISH+ data with some modifications. For clustering all cells from mouse cortex, subventricular zone(SVZ), choroid plexus, and olfactory bulb, we first normalize the counts per cell, then we choose highly variable genes with >0.4 min_dispersion, 0.01 min_mean, with max_mean =3. This yields 3509 genes. Then we take the logarithm of the data, regress out the total count effect per cell and scale the data to unit variance. We compute the PCAs and using top PCs to compute the neighborhood graph before performing Louvain clustering. We use the rank_gene_groups function with raw data and the top 20 genes enrichment in each cluster were used to identify the clusters based on marker genes annotation from single cell RNA-seq / DropSeq data[29,38]. We found that both Hierarchical clustering and Louvain clustering yield similar results despite different methods.

To spatially map back the clusters on the raw image, we perform Louvain clustering on cortex, SVZ, and choroid plexus data, and olfactory bulb data separately. Genes with max count greater than 4 across all cells were chosen for cortex and SVZ (include choroid plexus cells) data. Next, we filtered out cells with less than 200 genes expressed from analysis. The counts were normalized per cell and a minimum dispersion of greater than 0 with min_mean of 0.05 were chosen to filter out the variable genes. This yield 1813 genes for subsequent analysis. For the olfactory bulb, genes with max count greater than 2 across all cells were first chosen. Then the counts were normalized per cells. To obtain the highly variable genes,

a threshold of min_mean=0.05, and min_dispersion of 0.2 were chosen. This yields 1972 genes for subsequent analysis. After choosing the highly variable genes, the data was subjected to PCA reduction, computed neighborhood graph with top PCs, and Louvain clustering. The top 20 enrichment genes were obtained using rank_genes_groups function and the clusters were identified according to published literature. Sub-clustering of the main cluster was performed by repeating the process described above. The visualization of these clusters to two dimensions using Uniform Manifold Approximation and Projection (UMAP) was done with SCANPY function. These cluster numbers were mapped back to the original data to visualize the spatial heterogeneity of different cell types across different part of the tissues.

Calculation of the time acceleration of seqFISH+ vs expansion seqFISH.—For expansion seqFISH, we assume that to code ~20,000 genes, the coding scheme is with 4 colors and 8 rounds of hybridization (4^7=16,384 genes) with 1 round of error correction. Thus, the total number of effective imaging per field of view (FOV) is equal to the expansion factor $\times$ 4 $\times$ 8. For 60-fold expansion, this is 60×4×8=1920 images. For seqFISH+, we assume a coding scheme with 3 separate fluorescent channels, with 8000 genes coded in each channel for a total of 24000 genes. Pseudocolors are used to code for 8000 genes. For example, if the number of pseudocolors is 20 per fluorescent channel, then 4 rounds of barcoding (including 1 round of error correction) is need. The effective imaging per FOV is then $20 \times 4 \times 3$ =240 images, an 8-fold acceleration compared to expansion seqFISH. As another example, if the pseudocolor per channel is 10, then 5 rounds of barcoding is need to cover 8000 genes per channel. Then a total of $10 \times 5 \times 3$ =150 images. However, this coding scheme only provides 10×3= 30-fold decrease in the RNA density. If an equivalent of 30-fold expansion was implemented, then $30 \times 4 \times 8 = 960$ images are needed per FOV for an acceleration rate of 960/150 = 6.4 fold.

Bootstrap analysis.—We calculate the cell-to-cell correlation matrix with the number of genes were downsampled from the 2511 genes that expressed at least 5 copies in a cell. For each downsampled dataset, 100, 250, 500, 1000, 1500, and 2000 genes were selected randomly. The Pearson's correlation coefficient of each of the cell-to-cell correlation matrix is computed with the cell-to-cell correlation matrix for the 2511 gene dataset. 5 trials are simulated for each downsampled gene level. Error bars denote standard deviation.

Neighbor cell analysis.—The spatial coordinates for the cell centroids were used to create a nearest neighbor network (k = 4), whereby nodes represent individual cells and edges are observed proximities between 2 cells. Edges between identical or different annotated cell types were respectively labeled as homo- and heterotypic. To identify enriched or depleted proximities between two identical or different cell types the observed number of edges between any two cell types was compared to a random permutation (n = 100) distribution by reshuffling the cell labels. Associated p-values were calculated by observing how often the simulated values were higher or lower as the observed value for respectively enriched or depleted proximities.

Gene expression enrichment for cell types in close proximity was calculated as the average expression for that gene in all the cells of these two cell types that were in close proximity

according to the spatial network. Number of observed edges between two cell types and z-scores for each gene were further used to filter and identify enriched gene expression in any combination of two proximal cell types.

To determine the ligand-receptor pairs in neighboring cells, we extracted genes that have z-scores of 1 or greater, are expressed in at least 25% of the cells in the interacting pairs, and have at least 4 or more instances of being neighbors. We then match up the ligand-receptor pairs from literature[39], which is shown in Supplementary Table 4. To identify statistically enriched ligand-receptor pairs we compared the calculated ligand-receptor scores with that of a random permutation (n = 1000) distribution by reshuffling the cell labels. p-value < 0.05 is deemed to be significant.

RNA localization analysis.—To determine the subcellular localization patterns of mRNAs in the cortex, all cells are first separated into the 26 cell clusters (Extended Data Table 2). Within each cell class, the top 200 highly expressed genes are selected for localization analysis. In each cell, the average distance of all of the transcripts for each of the 200 genes from the center of the mass of all of the transcripts for all the genes are calculated. This metric corresponds to whether the gene is likely to be found close or far from the cell center. Only cells with 4 or more copies of that RNA are included in the calculation. The average distance from the center for each cell is normalized by the size of the cell, determined as the square root of the area span by the convex hull of all the mRNA dots in that cell. To select the genes that are localized far from the center of the cell, a threshold of 0.45 for the localization score is used and the average expression level is set at greater than 2.5 copies detected per cell. We selected genes that are close to the cell center using a localization score of 0.35 or lower and the expression level of greater than 2.5 copies per cell. The results are shown in Supplementary Table 3.

Contact maps.—The minimum distance between the pixels defining the edge of all pairs of cells in a field of view were tabulated. To count the number of times cells of each type were in contact with cells of each other type, the following procedure was followed. Cells within 15 pixels of a given cell were considered in contact, and the appropriate entry in a square matrix of length equal to the number of cell types was incremented. The counts were then normalized such that each row sums to 1. Hierarchical clustering was then performed to cluster cell types.
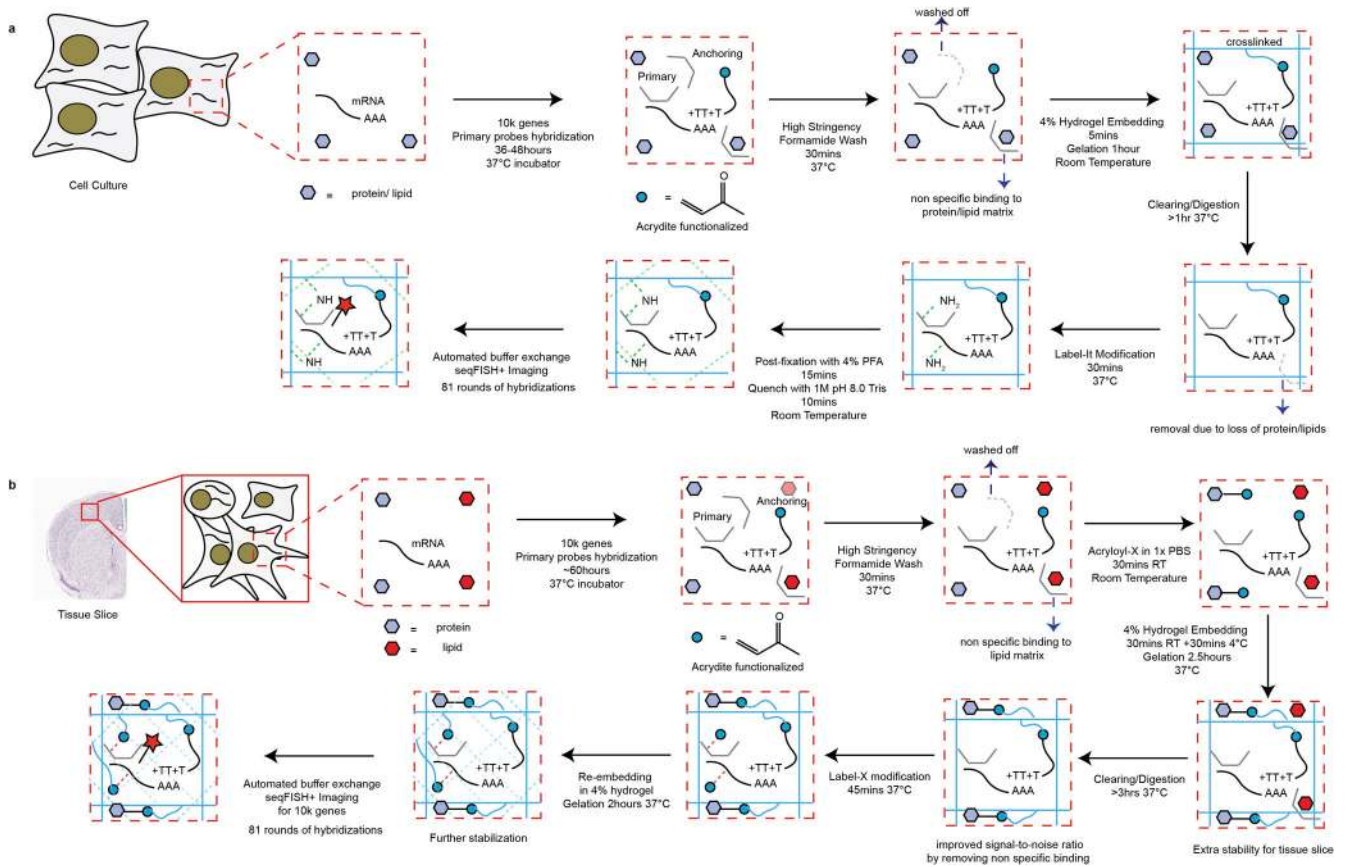
Code Availability.—The custom written scripts used in this study are available at https://github.com/CaiGroup/seqFISH-PLUS

Step-by-step protocol.—A detailed protocol for RNA seqFISH+ sample preparation is available at the *Protocol Exchange*[40].
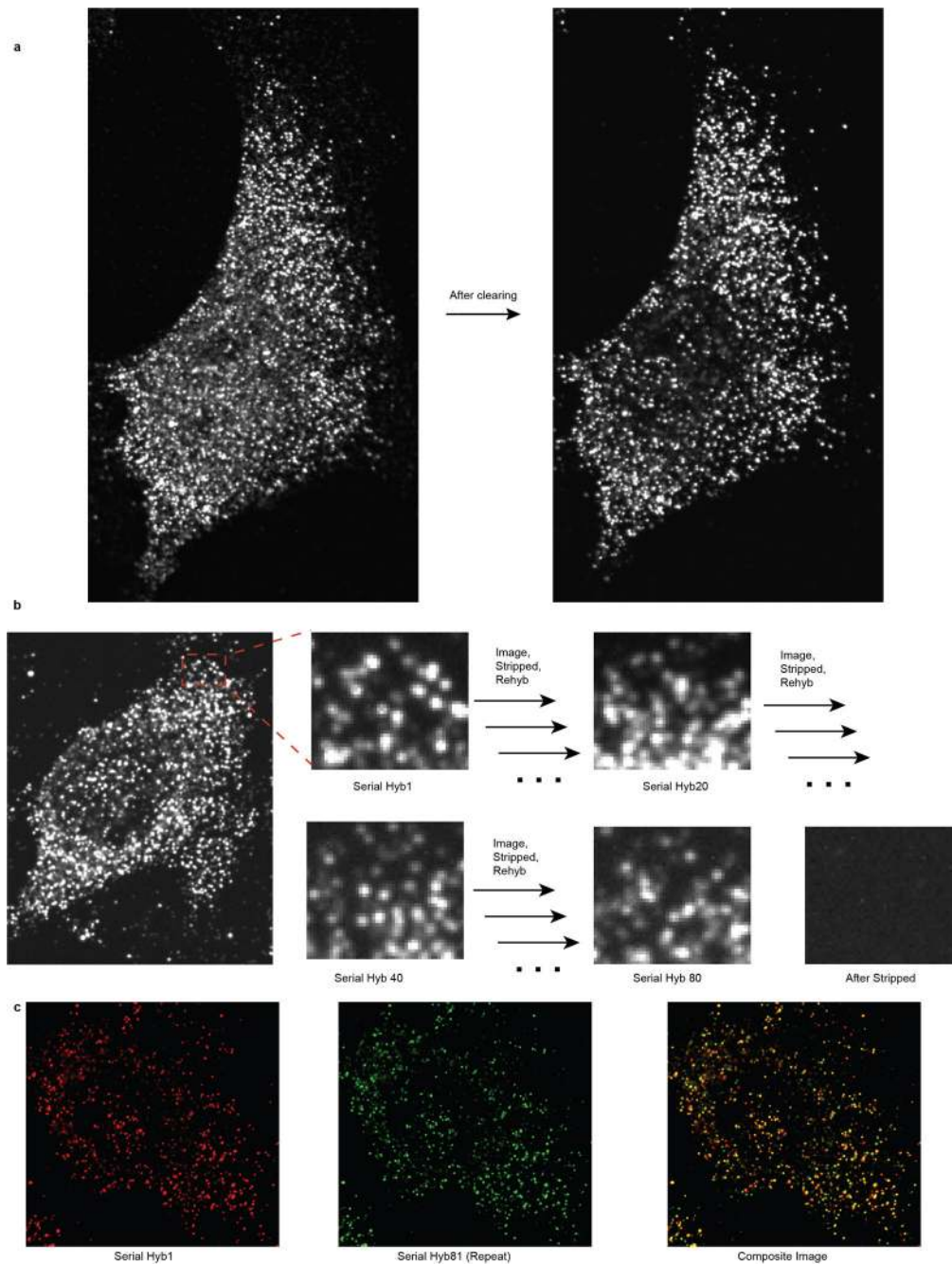
Supplementary Material

Refer to Web version on PubMed Central for supplementary material.
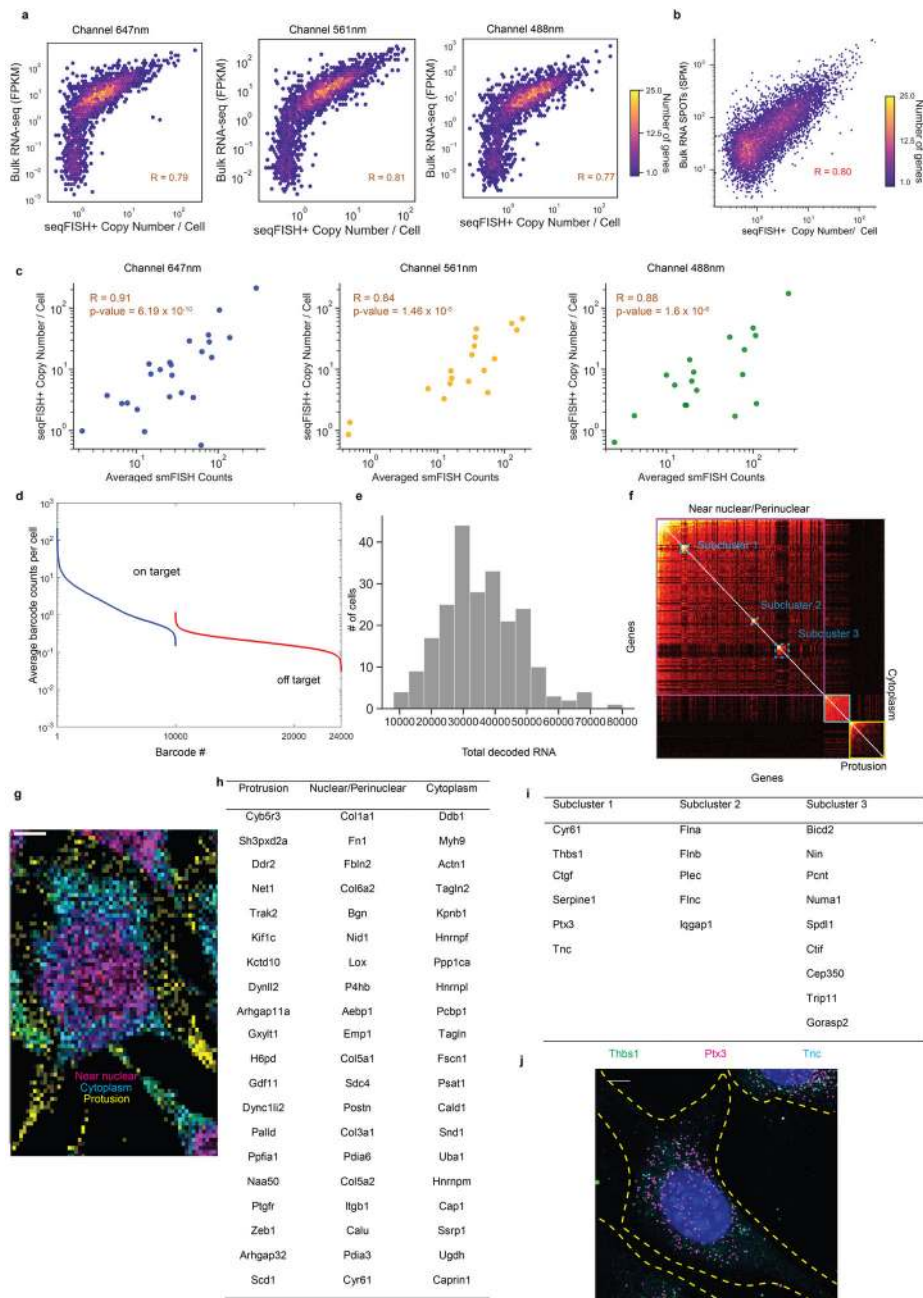
Extended Data

Extended Data Figure 1.
Clearing and probe anchoring protocols for the seqFISH+ experiments in (**a**) NIH3T3 cells and (**b**) the mouse brain slices.

Extended Data Figure 2.

Clearing removes background nonspecific bound dots. **a,** Raw images of a NIH3T3 cell before and after clearing. Significant decrease in background is observed in cleared sample. Image is acquired on a spinning disk confocal microscope. **b,** In each round of hybridization for the 10,000 gene experiment, diffraction limited dots are clearly separated, indicating the pseudocolor scheme effectively dilutes the density of the sample. Signal is completely removed between different rounds of hybridization, with no "cross-talk" between the pseudocolors. Stripping is accomplished by 55% formamide wash, which is highly efficient. **c,** After the completion of each seqFISH+ experiment, readout probes used in hyb1 is re-

hybridized in round 81. The colocalization rates between Hyb1 and 81 are 76% (647 channel), 73% (561 channel) and 80% (488 channel) within a 2 pixel radius. The colocalization between the two images indicates that most of the primary probes remain bound through 80 rounds of hybridization and imaging, although some loss of RNA and signal is seen across 80 rounds of hybridization (**a-c,** n = 227 cells).

Extended Data Figure 3.

seqFISH+ works efficiently across all three fluorescent channels and identifies localization patterns of transcripts in NIH3T3 cells. **a,** Correlation plots between seqFISH+ and bulk RNAseq in three fluorescent channels. Barcodes are coded entirely within each channel, with n = 3334, 3333, and 3333 barcodes in each channel respectively. Barcodes in all channels are decoded and called out efficiently. **b,** seqFISH+ result correlates strongly with RNA SPOTs measurement in NIH3T3 cells. SPM= Spots Per Million. **c,** Correlation between seqFISH+ and smFISH for each fluorescent channel (from left to right: n = 24, 18, 18 genes). All correlations were computed by Pearson's r coefficient correlation with two-

tailed p values reported. **d,** The callout frequency of on-target 10,000 barcodes versus the remaining 14,000 off target barcodes. Off target barcodes are called out at a rate of 0.22±0.07 (mean±s.d) per barcode. **e,** Histogram of the total number of mRNAs detected per NIH3T3 cell. On average, 35,492±12,222 transcripts are detected per cell. **f,** Genes are clustered based on their co-occurrence in 10×10 pixel window. Three major clusters are nuclear/perinuclear, cytoplasmic and protrusions. **g**, mRNAs show preferential spatial localization patterns: nuclear, cytoplasm and protrusion (n = 227 cells). The image is binned into 1 μm × 1 μm windows and colored based on the genes enriched in each bin (scale bar = 10 μm). **h,** Example of genes enriched in each spatial cluster. **i,** Genes in the subclusters within the nuclear localized group. Subcluster 1 contains genes encode for extracellular matrix proteins. Subcluster 2 genes are involved in actin cytoskeleton while subcluster 3 genes are involved in microtubule networks. **j,** Representative smFISH image (single z-slice) of three genes in subcluster 1 shows nuclear/perinuclear localization (n = 20 FOVs, 40x objective). Scale bar: 10μm.
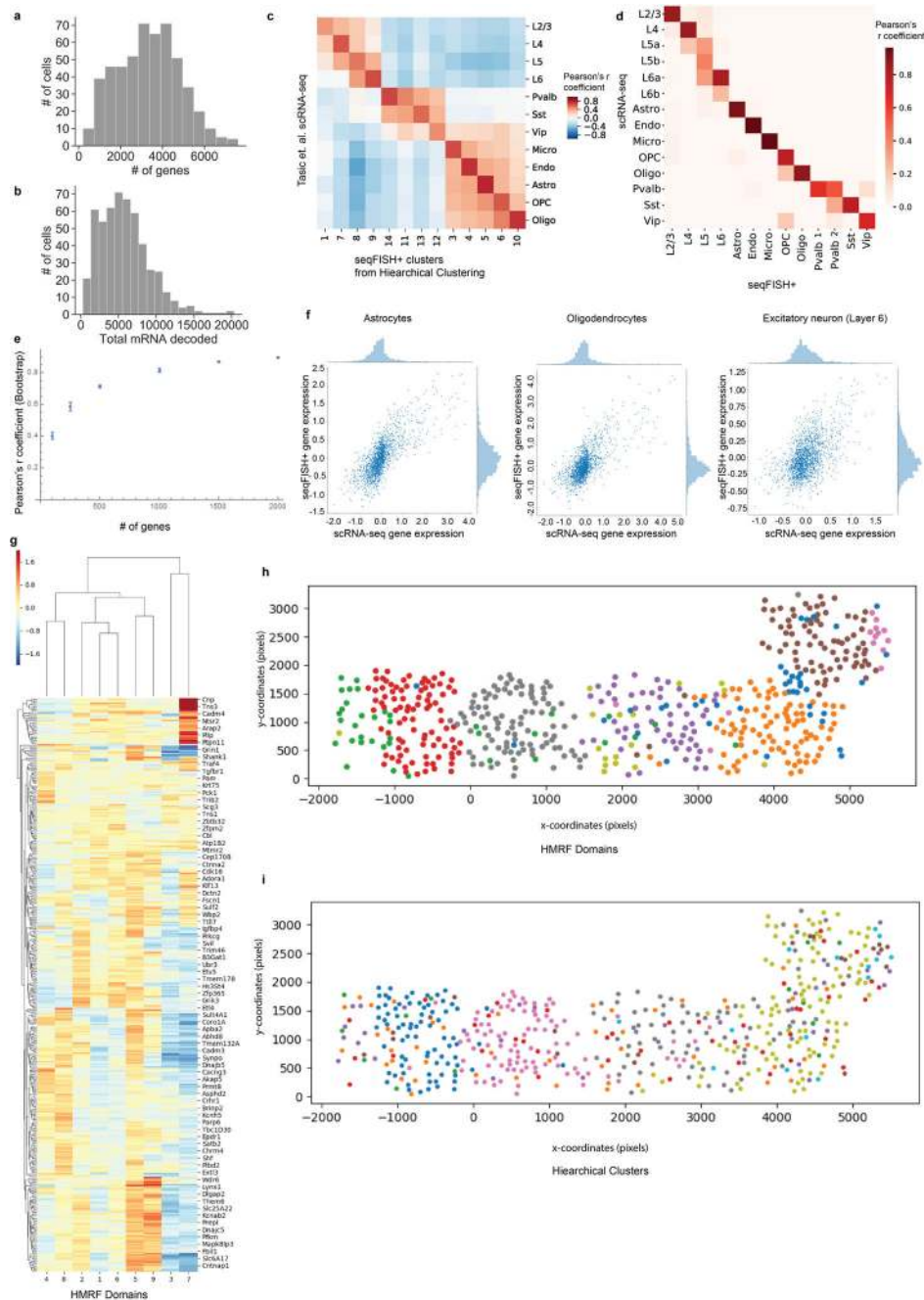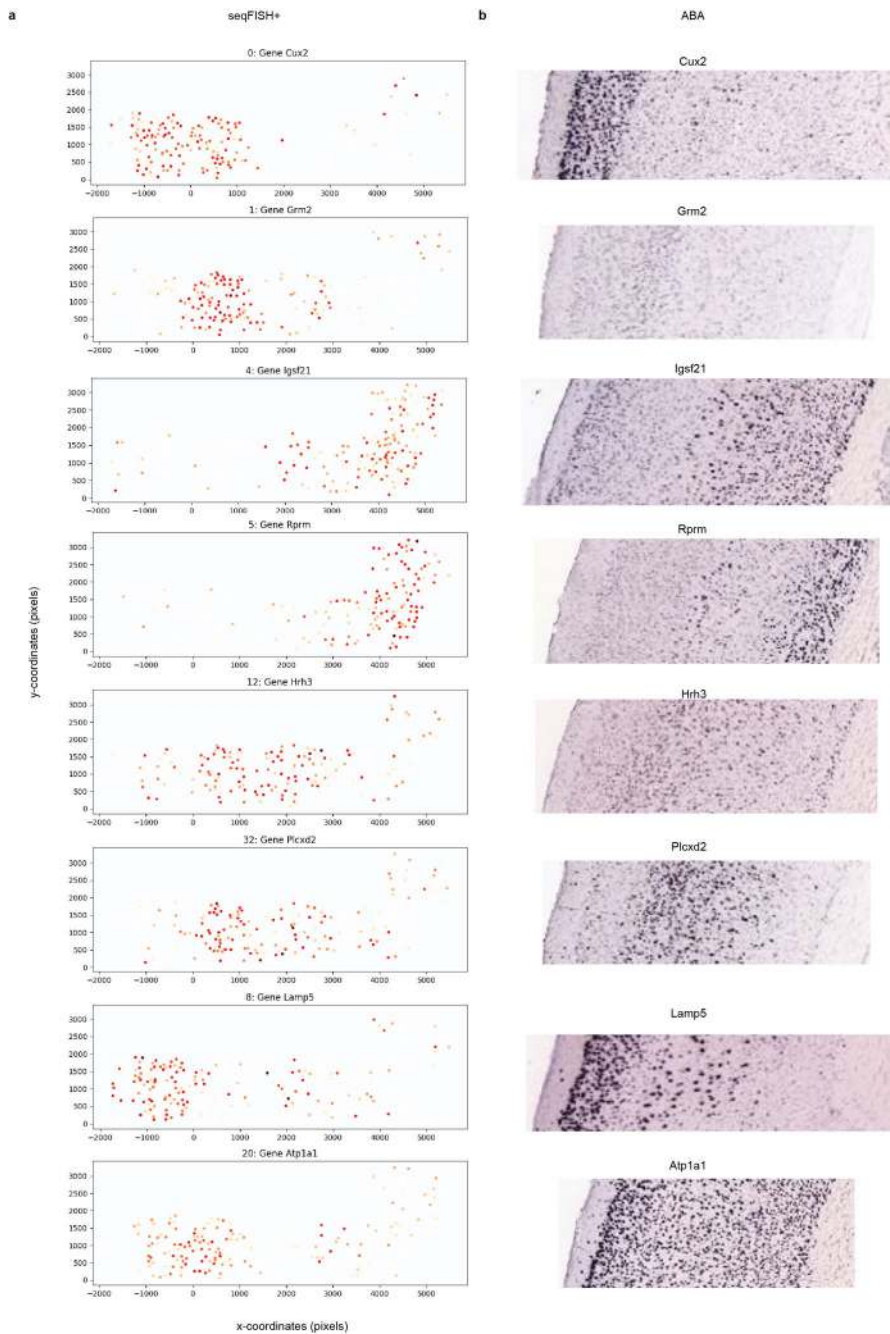
Extended Data Figure 4.

scRNAseq comparison with seqFISH+, bootstrap and HMRF analysis. **a,** Histogram of the
number of genes and **b,** total RNA barcodes detected per cell by seqFISH+ in the cortex. **c,**
Unsupervised clustering of seqFISH+ correlates well with scRNAseq. (n = 1857 genes;
Pearson's r coefficient correlation) **d,** Supervised mapping of seqFISH+ analyzed cortex cell
clusters with those from single cell RNA-seq clusters. (n = 1253 genes; p-value < 0.005). **e,**
The number of genes were downsampled from the 2511 genes that expressed at least 5
copies in a cell. For each downsampled dataset, the cell-to-cell correlation matrix is
calculated and correlated with the cell-to-cell correlation matrix for the 2511 gene dataset. 5

trials are simulated for each downsampled gene level. Error bars denote mean +/− standard deviation. Even when downsampled to 100 genes, about 40% of the cell to cell correlation is retained, because the expression pattern of many genes are correlated. **f,** Scatterplots of seqFISH+ with scRNAseq in different cell types. Each dot represents a gene and their mean expression z-score values in either seqFISH+ or scRNAseq in astrocytes, oligodendrocytes and excitatory neurons. In general, seqFISH+ and scRNAseq are in good agreement (n = 598 genes each). **g,** HMRF detects spatial domains that contain cells with similar expression patterns regardless of cell type. Domain specific genes are shown. **h,** Spatial domains in the cortex. **i,** Mapping of the hierarchical clusters onto the cortex. X-Y coordinates are in pixels (103 nm per pixel). Each camera field of view is 2000 pixels.
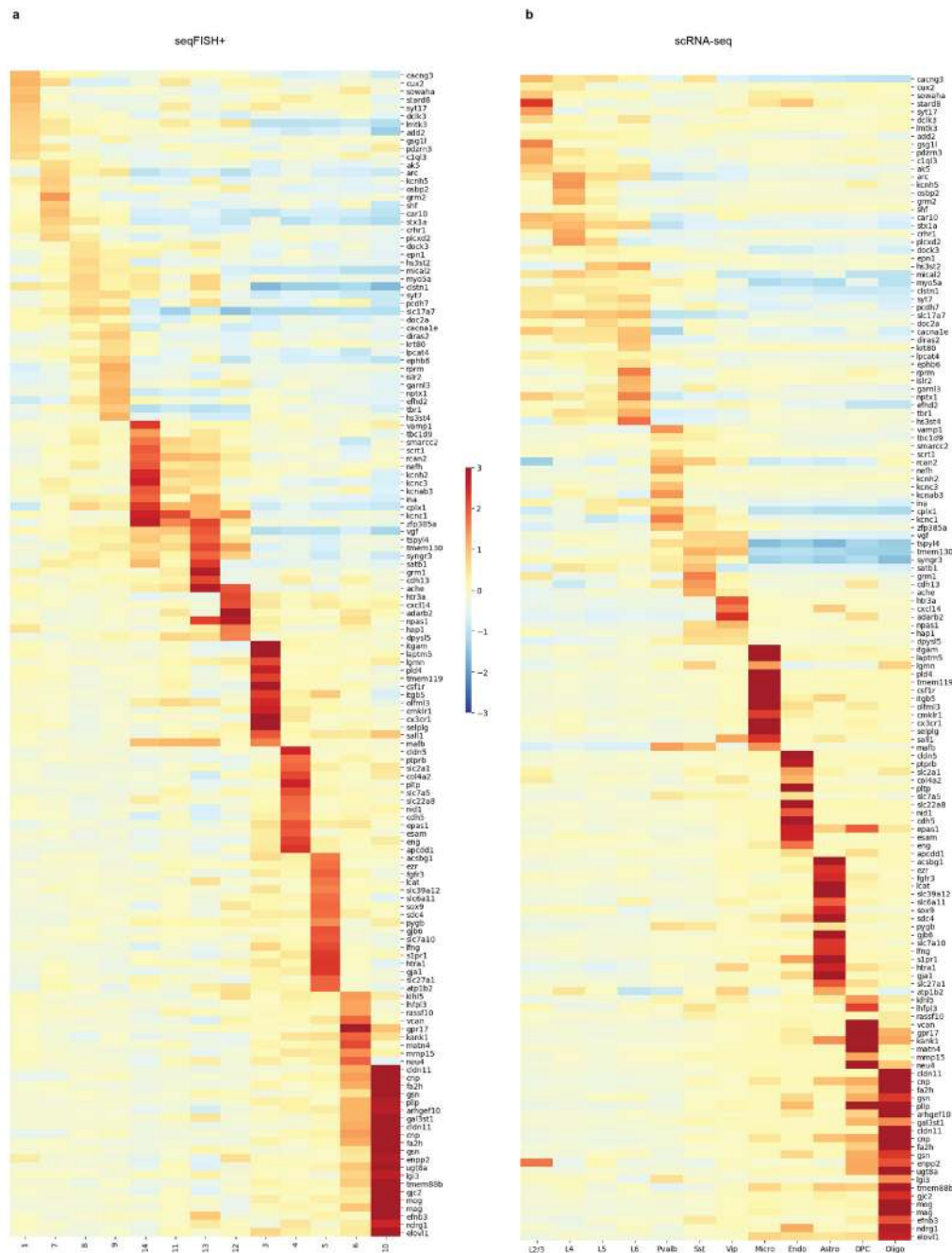
Extended Data Figure 5.
Differential gene expressions between the cell type clusters in both (**a**) seqFISH+ and (**b**) scRNA-seq. The expression patterns of seqFISH+ clusters are similar with scRNA-seq clusters (n = 143 genes)
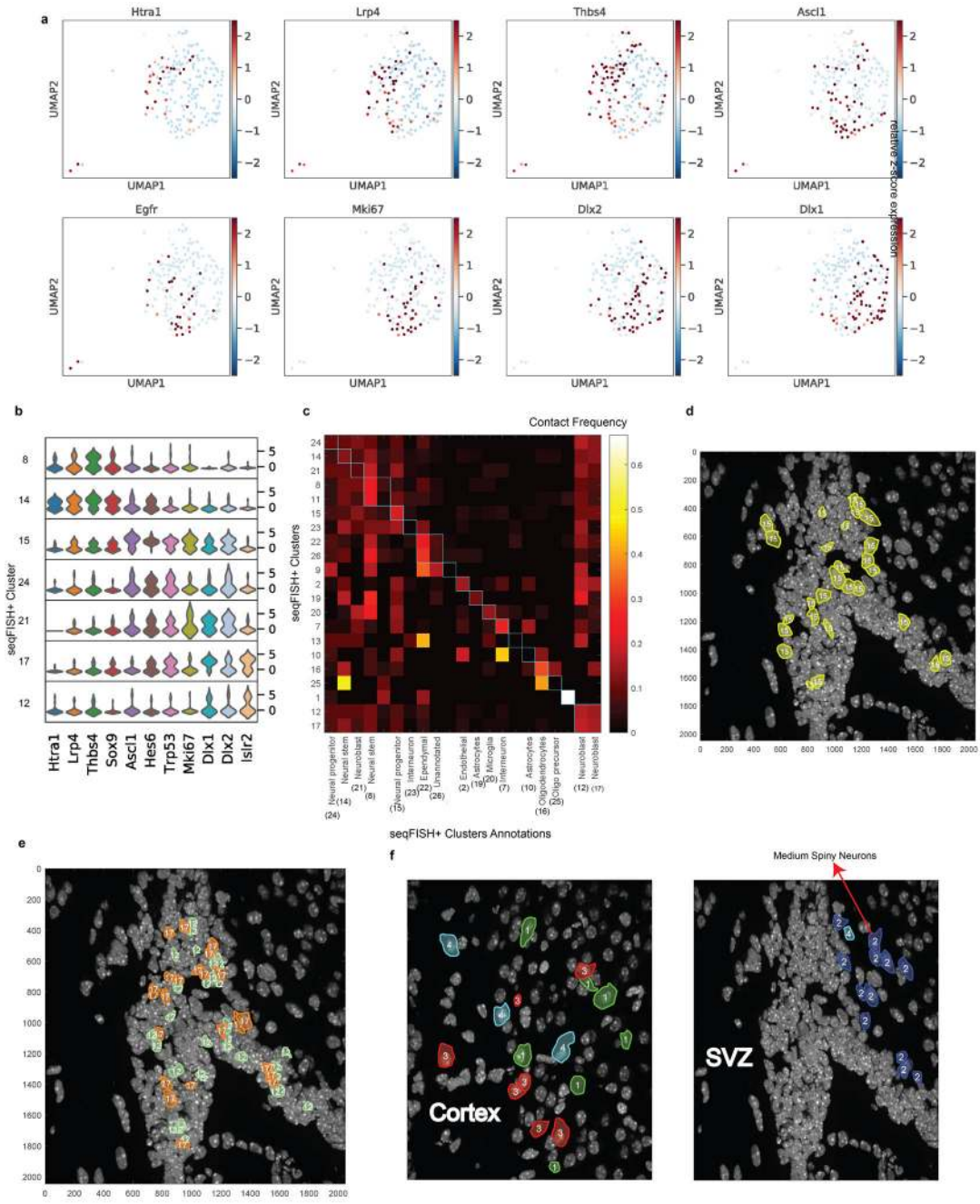
Extended Data Figure 6.
Comparison of the spatial expression patterns across the cortex in the (**a**) seqFISH+ data versus the (**b**) Allen Brain Atlas. X-Y coordinates are in pixels (103 nm per pixel). Layers I-VI are shown from left to right.
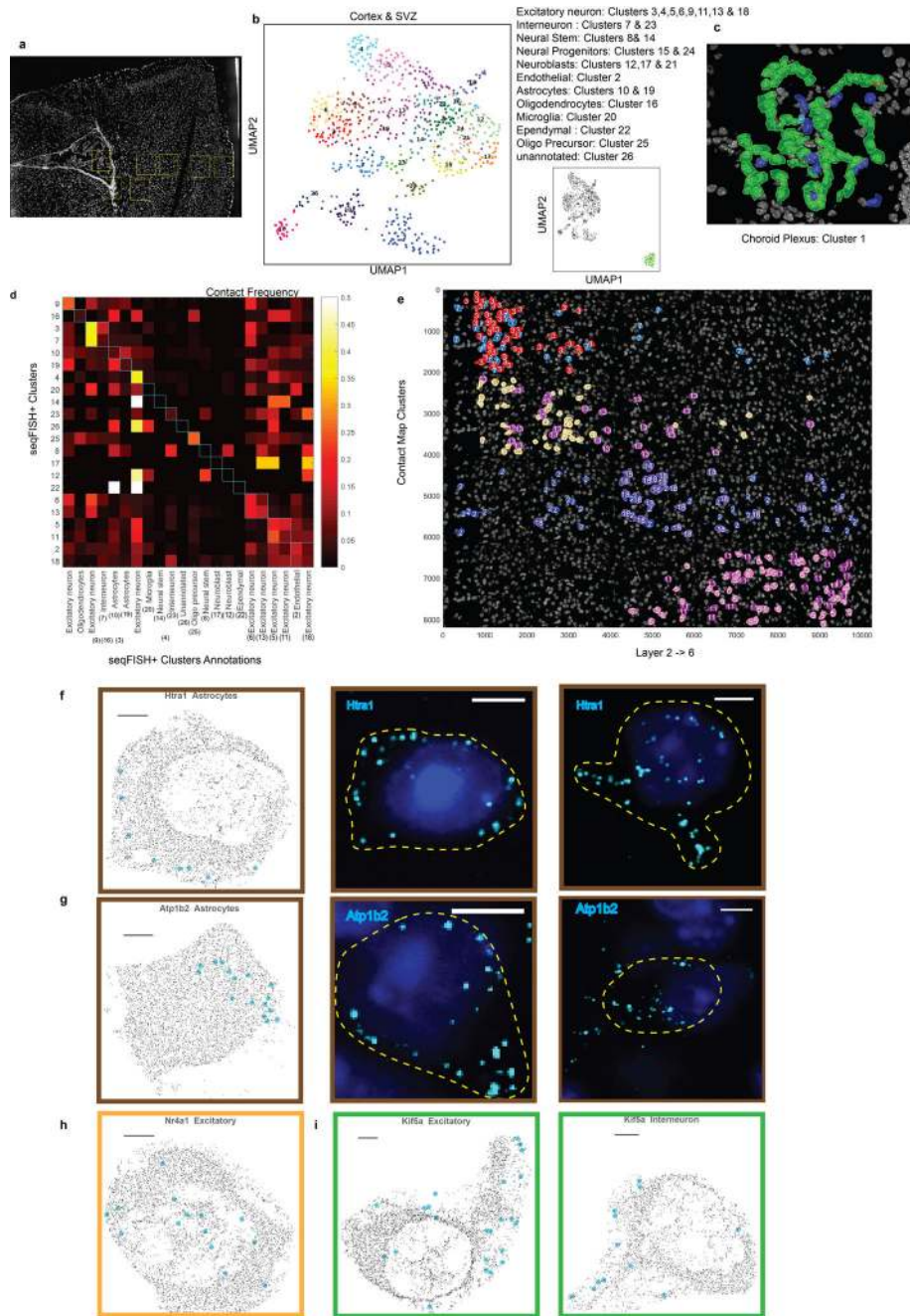
Extended Data Figure 7.

Additional analysis of cortex and subcellular localization patterns in different cell types. **a,** Slide explorer image of the cortex and SVZ FOVs imaged in the first brain slice (n=913 cells). Schematic is shown in Fig 3a. **b,** UMAP representation of cortex and SVZ cells. **c,** Mapping of the choroid plexus cells, which are exclusively present in the ventricle (n =109 cells). **d,** Frequency of contacts between the different cell class in the cortex, normalized for the abundances of cells in each clusters. **e,** Each strip represents cells that cluster together, which breaks into layers in the cortex, consistent with expectation, as cells within a layer preferential interact with each other (n = 523 cells). **f,** *Htra1* transcripts are preferentially

localized to the periphery of the astrocytes in the cortex. Left panel shows a reconstructed image from the 10,000 gene seqFISH+ experiment. *Htra1* transcripts are shown in cyan, and all other transcripts are shown in black. Scale bar is 2μm. Middle and right panels show single z-slice of smFISH images of *Htra1* in cortical astrocytes (Scale bar: 5μm). **g,** *Atp1b2* localization in seqFISH+ (left; scale bar: 2μm) and single z-slice smFISH images (middle and right; scale bars: 5μm). Many *Htra1* and *Atp1b2* transcripts are localized to astrocytic processes (**f,g**, n= 62 astrocytes). SmFISH images were background subtracted for better display of RNA molecules (n= 10 FOVs, 40x objective). **h,** *Nr4a1* localization patterns are distinct from *Htra1* and *Atp1b2* and are more nuclear localized across different cell types. An excitatory neuron is shown from the seqFISH+ reconstructions (n = 337 excitatory neurons; scale bars: 2μm). **i,** *Kif5a*, a kinesin, also exhibits periphery and process localizations in different cell types (n = 60 interneurons; scale bar: 2μm).

Extended Data Figure 8.

Additional analysis of the subventricular zone (SVZ). **a,** Expression of individual genes in the SVZ in the UMAP representation (n = 281 cells). **b,** Violin plots denotes z-scored gene expression patterns for Louvain clusters corresponding to NSC to neuroblasts in the SVZ, (n = 281 cells). **c,** Spatial proximity analysis of the cell clusters in the mouse subventricular zone(SVZ). Frequency of contacts between the different cell class in the SVZ, normalized for the abundances of cells in each clusters. **d,** Neural progenitors appear to be in spatial proximity with each other. **e,** Two neuroblasts cell clusters are found to be in spatial proximity in the SVZ (**c-d**, n = 281 cells). **f,** Subclusters of type 7 cells in the cortex (left).

Medium spiny neurons that expressed *Adora2, Pde10a,* and *Rasd2* marker genes form a separate cluster that are detected only in the striatum (right) (n = 42 cells in cluster 7).
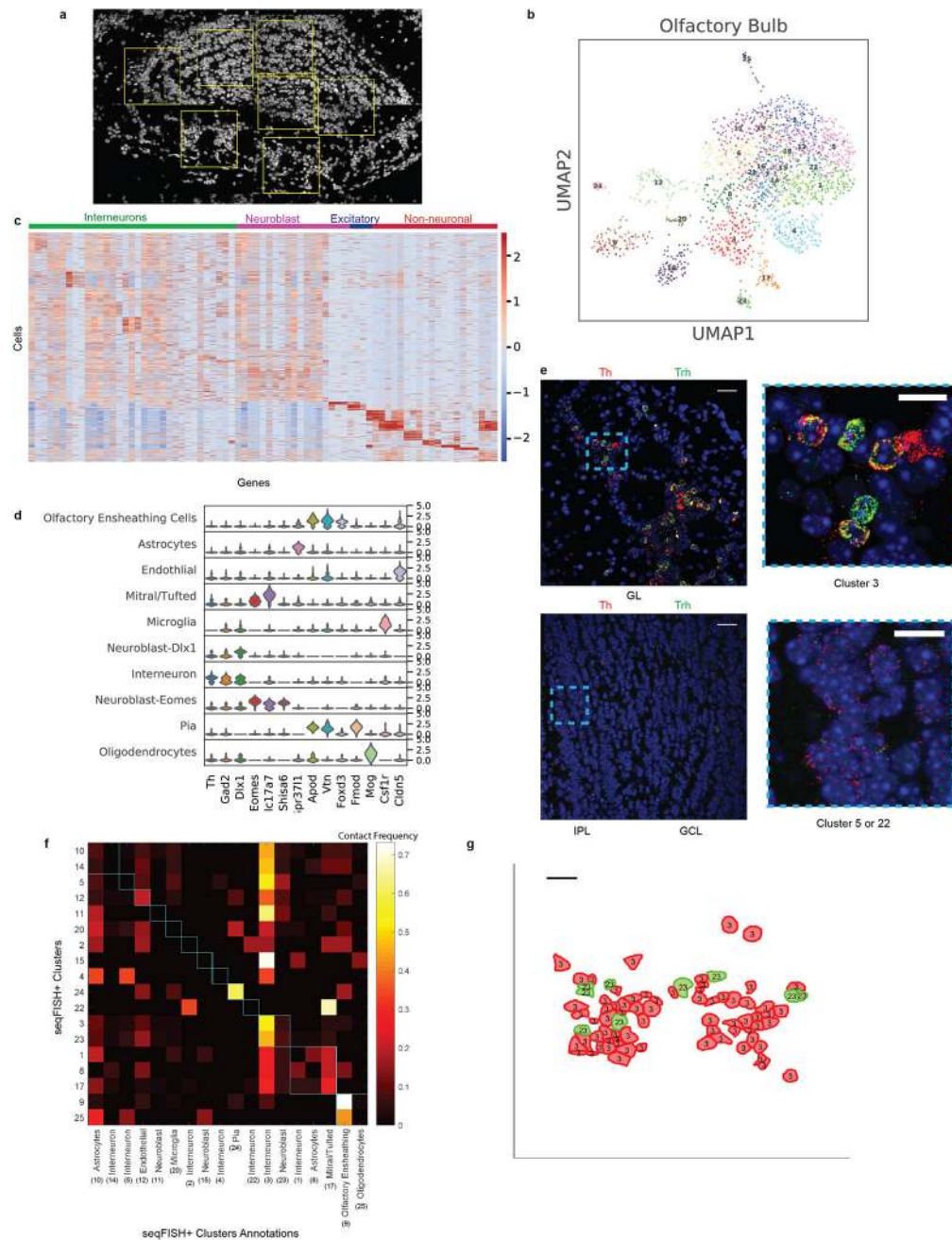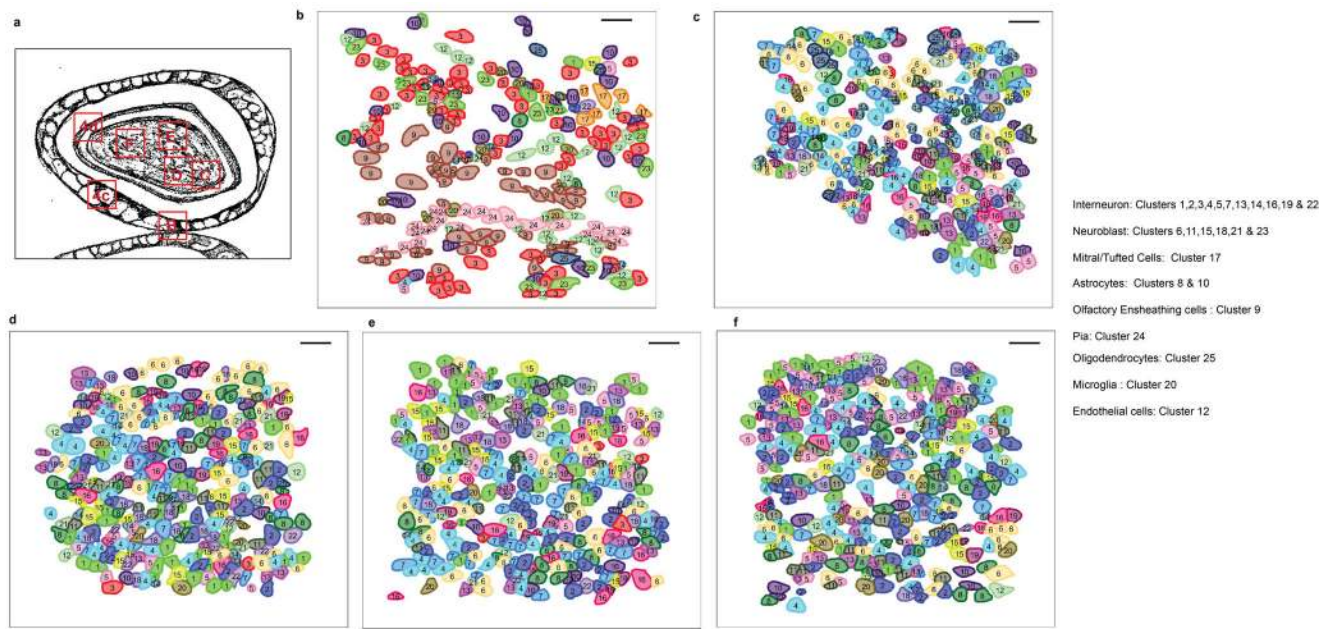
Extended Data Figure 9.

Additional analysis of the olfactory bulb (OB). **a,** Slide explorer image of the OB FOVs imaged in the second brain slice. **b,** UMAP analysis of OB cells. **c,** Z-scored gene expression patterns heatmap of cells in the olfactory bulb. **d,** Violin plots show z-scored marker genes expression patterns in the different classes of cells detected in the OB. (**a-d**, n = 2050 cells) **e,** Representative smFISH images of *Th* and *Trh*. Images were maximum z-projected. In the glomeruli layer (GL), cluster 3 cells express both *Th* and *Trh*, whereas in the GCL, only *Th* are expressed (cluster 5 and 22 cells). (n= 10 FOVs, 40x objective). Scale bars: 13μm (left image); 6.5μm (right image). **f,** Frequency of contacts between the different

cell class in the glomerulus, normalized for the abundances of cells in each cluster. **g,** Cell clusters #3 (*Th*+ interneurons) and #23 (neuroblast) are in close proximity in the mapped image (**f-g,** scale bars: 20μm).

Extended Data Figure 10.

Spatial organization of the olfactory bulb. **a,** Schematics of the field of views imaged in the OB. Spatial mapping of the cell clusters in the Glomerulus Layer (**b**) and Granule Cell Layer (**c-f**) in the OB. Note the neuroblast cells tend to reside in the interior of the GCL (upper parts of **c** and **d** and lower parts of **e** and **f**), whereas more mature interneurons are present in the outer layer. This is consistent with the migration of neuroblasts from the SVZ through the rostral migratory stream into the granule cell layer.Scale bars : 20μm.

Acknowledgment:

References

1. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M & Cai L Single-cell in situ RNA profiling by sequential hybridization. Nat. Methods 11, 360–361 (2014). [PubMed: 24681720]

2. Chen KH, Boettiger AN, Moffitt JR, Wang S & Zhuang X Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, aaa6090 (2015). [PubMed: 25858977]

3. Shah S, Lubeck E, Zhou W & Cai L In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron 92, 342–357 (2016). [PubMed: 27764670]

4. Lee JH et al. Highly multiplexed subcellular RNA sequencing in situ. Science 343, 1360–1363 (2014). [PubMed: 24578530]

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

6. Femino AM, Fay FS, Fogarty K & Singer RH Visualization of single RNA transcripts in situ. Science 280, 585–590 (1998). [PubMed: 9554849]

7. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A & Tyagi S Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods 5, 877–879 (2008). [PubMed: 18806792]

8. Eng C-HL, Shah S, Thomassie J & Cai L Profiling the transcriptome with RNA SPOTs. Nat. Methods 14, 1153–1155 (2017). [PubMed: 29131163]

9. Shah S et al. Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. Cell 174, 363–376.e16 (2018). [PubMed: 29887381]

10. Ke R et al. In situ sequencing for RNA analysis in preserved tissue and cells. Nat. Methods 10, 857 (2013). [PubMed: 23852452]

11. Lubeck E & Cai L Single-cell systems biology by super-resolution imaging and combinatorial labeling. Nat. Methods 9, 743–748 (2012). [PubMed: 22660740]

12. Betzig E et al. Imaging intracellular fluorescent proteins at nanometer resolution. Science 313, 1642–1645 (2006). [PubMed: 16902090]

13. Rust MJ, Bates M & Zhuang X Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat. Methods 3, 793–795 (2006). [PubMed: 16896339]

14. Zhu Q, Shah S, Dries R, Cai L & Yuan G-C Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. Nat. Biotechnol (2018). doi:10.1038/nbt.4260

15. Thompson RE, Larson DR & Webb WW Precise nanometer localization analysis for individual fluorescent probes. Biophys. J 82, 2775–2783 (2002). [PubMed: 11964263]

16. Yildiz A, Tomishige M, Vale RD & Selvin PR Kinesin walks hand-over-hand. Science 303, 676–678 (2004). [PubMed: 14684828]

17. Chen F, Tillberg PW & Boyden ES Optical imaging. Expansion microscopy. Science 347, 543–548 (2015). [PubMed: 25592419]

18. Yang B et al. Single-cell phenotyping within transparent intact tissue through whole-body clearing. Cell 158, 945–958 (2014). [PubMed: 25088144]

19. Chen F et al. Nanoscale imaging of RNA with expansion microscopy. Nat. Methods 13, 679–684 (2016). [PubMed: 27376770]

20. Moffitt JR et al. High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. Proc. Natl. Acad. Sci. U. S. A 113, 14456–14461 (2016). [PubMed: 27911841]

21. Antebi YE et al. Combinatorial Signal Perception in the BMP Pathway. Cell 170, 1184–1196.e24 (2017). [PubMed: 28886385]

22. Mili S, Moissoglu K & Macara IG Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. Nature 453, 115–119 (2008). [PubMed: 18451862]

23. Wang T, Hamilla S, Cam M, Aranda-Espinoza H & Mili S Extracellular matrix stiffness and cell contractility control RNA localization to promote cell migration. Nat. Commun 8, 896 (2017). [PubMed: 29026081]

24. McInnes L, Healy J, Saul N & Großberger L UMAP: Uniform Manifold Approximation and Projection. The Journal of Open Source Software 3, 861 (2018).

25. Tasic B et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci 19, 335–346 (2016). [PubMed: 26727548]

26. Shah PT et al. Single-Cell Transcriptomics and Fate Mapping of Ependymal Cells Reveals an Absence of Neural Stem Cell Function. Cell 173, 1045–1057.e9 (2018). [PubMed: 29727663]

27. La Manno G et al. RNA velocity of single cells. Nature 560, 494–498 (2018). [PubMed: 30089906]

28. Frieda KL et al. Synthetic recording and in situ readout of lineage information in single cells. Nature 541, 107–111 (2017). [PubMed: 27869821]

29. Zeisel A et al. Molecular Architecture of the Mouse Nervous System. Cell 174, 999–1014.e22 (2018). [PubMed: 30096314]

30. Takei Y, Shah S, Harvey S, Qi LS & Cai L Multiplexed Dynamic Imaging of Genomic Loci by Combined CRISPR Imaging and DNA Sequential FISH. Biophys. J 112, 1773–1776 (2017). [PubMed: 28427715]

31. Wang G, Moffitt JR & Zhuang X Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. Sci. Rep 8, 4847 (2018). [PubMed: 29555914]

32. Edelstein A, Amodaj N, Hoover K, Vale R & Stuurman N Computer control of microscopes using µManager. Curr. Protoc. Mol. Biol Chapter 14, Unit14.20 (2010).

33. Parthasarathy R Rapid, accurate particle tracking by calculation of radial symmetry centers. Nat. Methods 9, 724–726 (2012). [PubMed: 22688415]

34. Chung NC & Storey JD Statistical significance of variables driving systematic variation in high-dimensional data. Bioinformatics 31, 545–554 (2015). [PubMed: 25336500]

35. Huang H, Liu Y, Yuan M & Marron JS Statistical Significance of Clustering using Soft Thresholding. J. Comput. Graph. Stat 24, 975–993 (2015). [PubMed: 26755893]

36. Finak G et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 16, 278 (2015). [PubMed: 26653891]

37. Wolf FA, Angerer P & Theis FJ SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018). [PubMed: 29409532]

38. Saunders A et al. Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell 174, 1015–1030.e16 (2018). [PubMed: 30096299]

39. Ramilowski JA et al. A draft network of ligand-receptor-mediated multicellular signalling in human. Nat. Commun 6, 7866 (2015). [PubMed: 26198319]

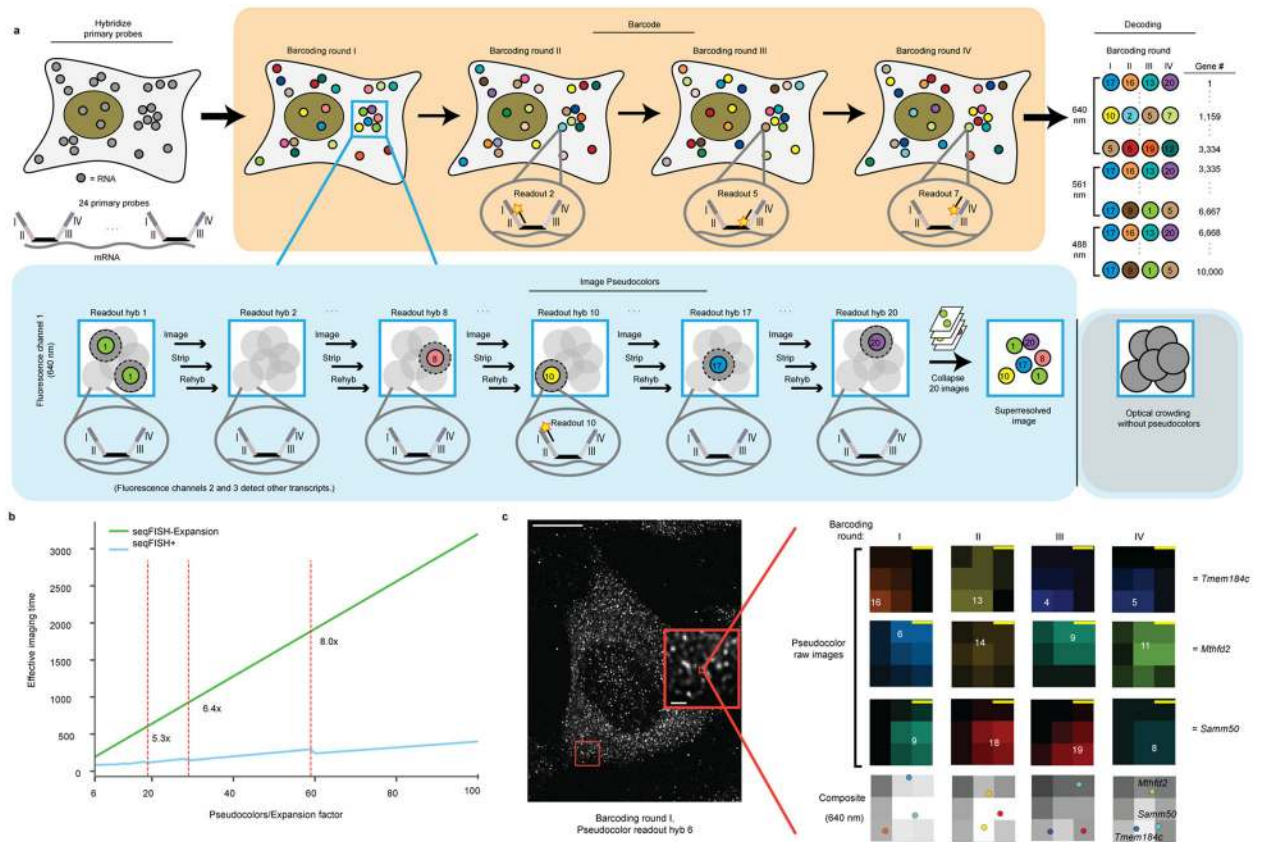40. Eng C-HL & Cai L Protoc. Exch 10.1038/protex.2019.019 (2019).

Figure 1. seqFISH+ resolves optical crowding and enables transcriptome profiling *in situ*.
a, Schematics of seqFISH+. Primary probes (24 per gene) against 10,000 genes are hybridized in cells. Overhang sequences (I-IV) on the primary probes correspond to 4 barcoding rounds (orange panel). Only 1/20th of the total genes in each fluorescent channel are labeled by readout probes in each pseudocolor readout round, lowering the density of transcripts in each image. mRNA dots in each pseudocolor can then be localized by Gaussian fitting and collapsed into a super-resolved image (blue panel). Each gene is barcoded within only one fluorescent channel (Methods). **b,** Compared to seqFISH with expansion microscopy (seqFISH-Expansion, green line) in covering 24,000 genes, seqFISH + with 60 pseudocolors (blue line) is 8 fold faster in imaging time. (Methods). **c,** Image of a NIH3T3 cell from one round of hybridization (n = 227 cells; scale bar = 10 μm). Zoomed in inset shows individual mRNAs (scale bar = 1 μm). Different mRNAs are decoded within a diffraction limited region, magnified from the inset (scale bar = 100 nm). The number in each panel corresponds to the pseudocolor round that each mRNA was detected, with no dots detected during the other pseudocolor rounds in this channel (640 nm).
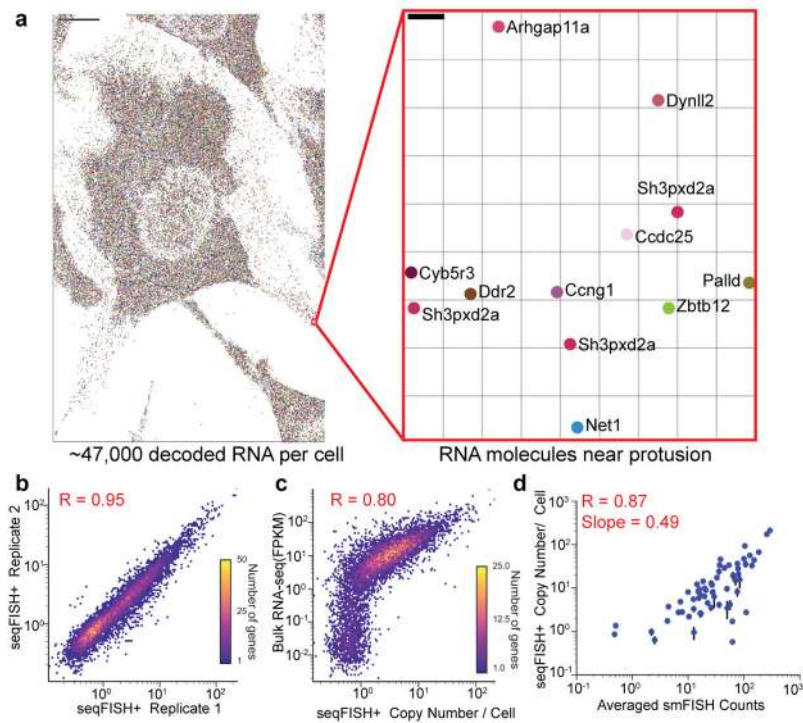
Figure 2. seqFISH+ profiles 10,000 genes in cells with high efficiency.
a, Approximately 47,000 mRNAs (colored dots) were identified in a NIH3T3 cell from a single z-section (scale bar = 10 μm). Inset shows the transcripts decoded in cell protrusions (n = 227 cells; scale bar = 100 nm). **b,** seqFISH+ replicates in NIH3T3 cells are highly reproducible (n$_1$ = 103 cells; n$_2$ = 124 cells). seqFISH+ correlates well with (**c**) RNA-seq (n = 9875 genes) and (**d**) single molecule FISH (n= 60 genes; p-value = 2.26×10^{-19}). The efficiency of seqFISH+ is about 49% compared to smFISH. Error bars in (**d**) represents standard error of the mean. (**b-d,** p-values < 0.0001, Pearson's r, two-tailed p values).
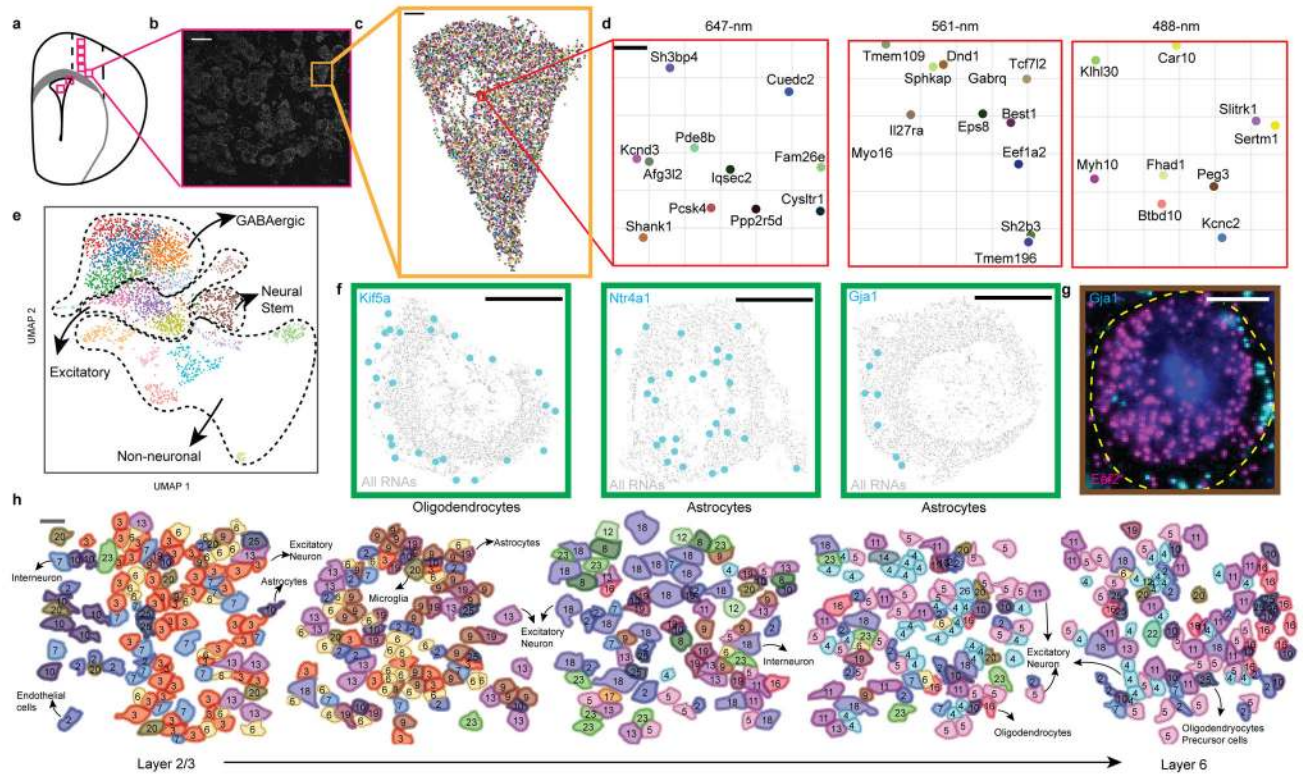
Figure 3. seqFISH+ robustly characterize cell classes and subcellular RNA localization in brain slices.

a, Schematic of the regions (red boxes) imaged. **b,** Cells in a single FOV of the cortex (scale bar = 20 μm). **c,** Reconstruction of the 9,418 mRNAs (colored dots) detected in a cell (scale bar = 2 μm). **d,** Decoded transcripts for a magnified region (n= 523 cells, scale bar= 100nm). **e,** Uniform Manifold Approximation and Projection (UMAP) representation of the seqFISH + data in the cortex, SVZ, and olfactory bulb (n=2963 cells). **f,** Reconstructed seqFISH+ images show subcellular localization patterns for mRNAs (Cyan) in different cell types. (n = 62 astrocytes and 28 oligodendrocytes; scale bar = 2 μm). **g,** smFISH of *Gja1* in cortical astrocytes shows periphery localization compared to the uniform distribution of *Eef2* mRNAs. (n=10 FOVs,40x objective;scale bar = 5μm). **h**. Each cortex layer consists of a distinct cell class composition (see annotations, Supplementary Table 2). (scale bar = 20 μm).
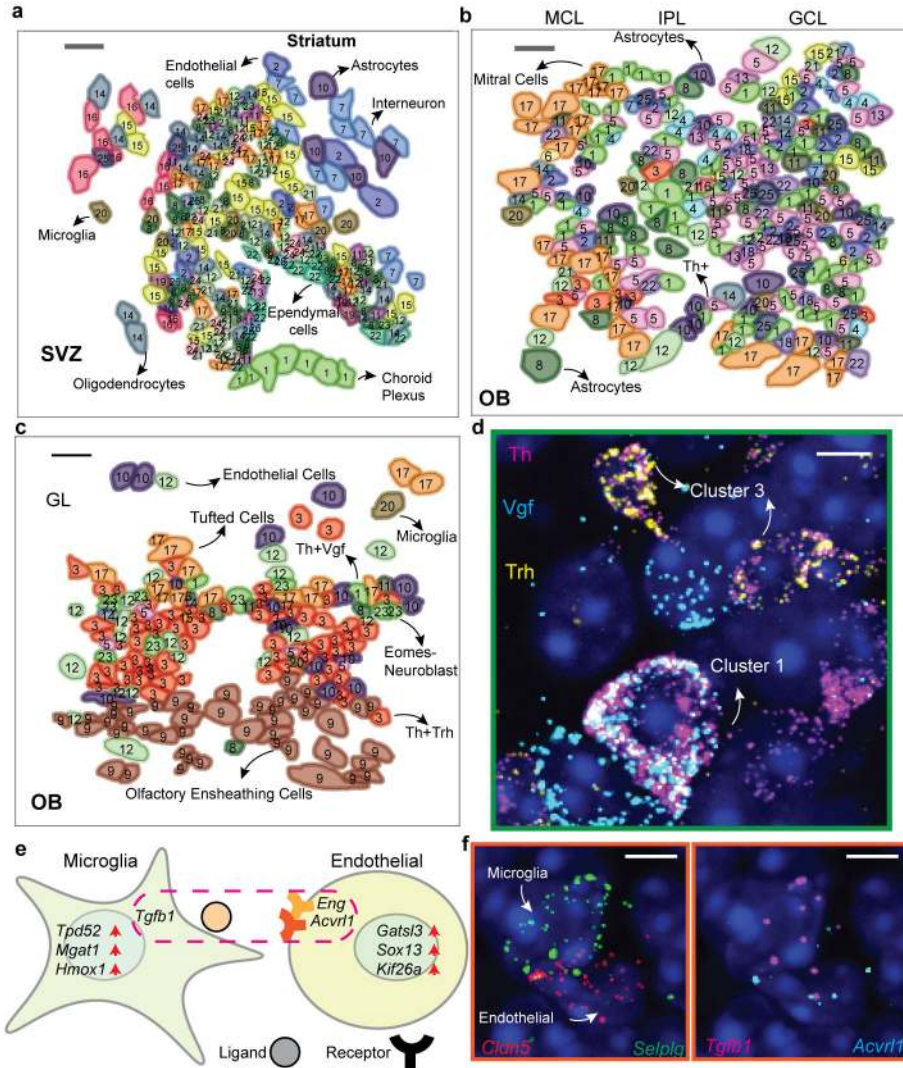
Figure 4. seqFISH+ reveals ligand receptor repertoires in neighboring cells and spatial organization in tissues.

a, Spatial organization of distinct cell clusters in the SVZ. **b,** Spatially-resolved cell cluster maps of the mitral cell layer(MCL), granule cell layer(GCL), and **c,** glomerular layer(GL) (scale bars: 20 μm). Remaining FOVs are shown in Extended Data Figure 10. The cluster numbers in the SVZ and OB are different (Supplementary Table 2). **d,** Distinct populations of *Th*+ dopaminergic neurons in the OB with differential expression of *Vgf* and *Trh*, shown with smFISH, confirming seqFISH+ clustering analysis. **e,** Schematic showing ligand-receptor pairs in neighboring microglia-endothelial cells. In microglia next to endothelial cells, certain genes, such as *Tpd52*, are enriched compared to microglia neighboring other cell types. **f,** mRNAs of *Tgfb1* ligand and *Acvrl1* receptor are visualized in adjacent microglia-endothelial cells by smFISH. (**d&f,** n = 10 FOVs, 40x objective; scale bars = 5 μm)