LETTER

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] & Long Cai[2]*

Imaging the transcriptome in situ with high accuracy has been a major challenge in single-cell biology, which is particularly hindered by the limits of optical resolution and the density of transcripts in single cells[1–5]. Here we demonstrate an evolution of sequential fluorescence in situ hybridization (seqFISH+). We show that seqFISH+ can image mRNAs for 10,000 genes in single cells—with high accuracy and sub-diffraction-limit resolution—in the cortex, subventricular zone and olfactory bulb of mouse brain, 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 neighbouring 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 fluorescence in situ hybridization (seqFISH) to impart a temporal barcode on RNAs through multiple rounds of hybridization allowed many molecules to be multiplexed[1–3]. seqFISH was recently shown to scale to the genome level in vitro[8], and could be applied for nascent transcription active sites[9].

Nevertheless, global profiling of mRNA in cells is hindered by 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 impeded all implementations of spatial profiling experiments[3–5]. For example, in situ sequencing methods detected only around 500 transcripts per cell[4,5,10] because of the low efficiency and large dot size of rolling-circle amplification, whereas seqFISH detected thousands of transcripts per cell[3]. We previously proposed combining super-resolution microscopy with FISH[11] to overcome this crowding problem. However, existing super-resolution localization microscopy[12,13] relies on the detection of single dye molecules, which emit a limited number of photons and only work robustly in optically thin (less than 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 relied on existing genomics methods, such as single-cell RNA sequencing (scRNA-seq) analysis, to identify target genes, and they serve only to map cell types identified by scRNA-seq. At the level of hundreds and even a thousand genes, spatial methods cannot be used as a de novo discovery-driven tool; this is a major disadvantage of the technology. In addition, many genes are expressed in a spatially dependent fashion independent of cell types[14], and this information is lost when analysing dissociated cells.

Here we demonstrate seqFISH+, which achieves super-resolution imaging and multiplexing of 10,000 genes in a single cell using sequential hybridizations and imaging with a standard confocal microscope. The key to seqFISH+ is expansion of the barcode base palette from four or five colours—as used in seqFISH[1,3] and in situ sequencing experiments[4,5]—to a much larger palette of 'pseudocolours' (Fig. 1a) by sequential hybridization. By using 60 pseudocolour channels, we effectively dilute mRNA molecules into 60 separate images and enable 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 pseudocolours into three fluorescent channels (Alexa Fluor 488, Cy3b and Alexa Fluor 647) and generate barcodes only within each channel to avoid chromatic aberrations between channels. Each channel can contain 20^3 (8,000) barcoded genes, giving a total of 24,000 genes when repeating this pseudocolour imaging four times, with one round used for error correction[3].

As imaging time is the main bottleneck in spatial transcriptomics experiments, imaging for seqFISH+ is eightfold faster compared to implementing seqFISH with expansion microscopy[17] (Fig. 1b). Using seqFISH, an equivalent 60-fold expansion of the sample would require four colours with eight barcoding rounds and 60-fold volume expansion, resulting in 1,920 images per field of view (FOV) to cover 4^7 (that is, 16,384) genes. By contrast, seqFISH+ acquires 60 pseudocolours with four barcoding rounds, giving 240 images per FOV to cover 24,000 genes—an eightfold reduction in imaging time. Furthermore, a large number of pseudocolours and a shorter four-unit barcode decrease the errors that accumulate over barcode rounds.

To demonstrate transcriptome-level profiling in cells, we first applied seqFISH+ to cleared NIH/3T3 fibroblast cells[18–20] (Fig. 1c, Extended Data Figs. 1, 2). We randomly selected 10,000 genes, avoiding highly abundant housekeeping genes, such as ribosomal proteins. These 10,000 genes sum up to more than 125,000 fragments per kilobase of transcript per million mapped reads (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 around 420,000 FPKM[21], only a threefold higher density than the 10,000-gene experiment, and this can be accommodated with the current scheme, or with more channels or pseudocolours.

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

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA. [2]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [3]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard T.H.Chan School of Public Health, Boston, MA, USA. *e-mail: lcai@caltech.edu
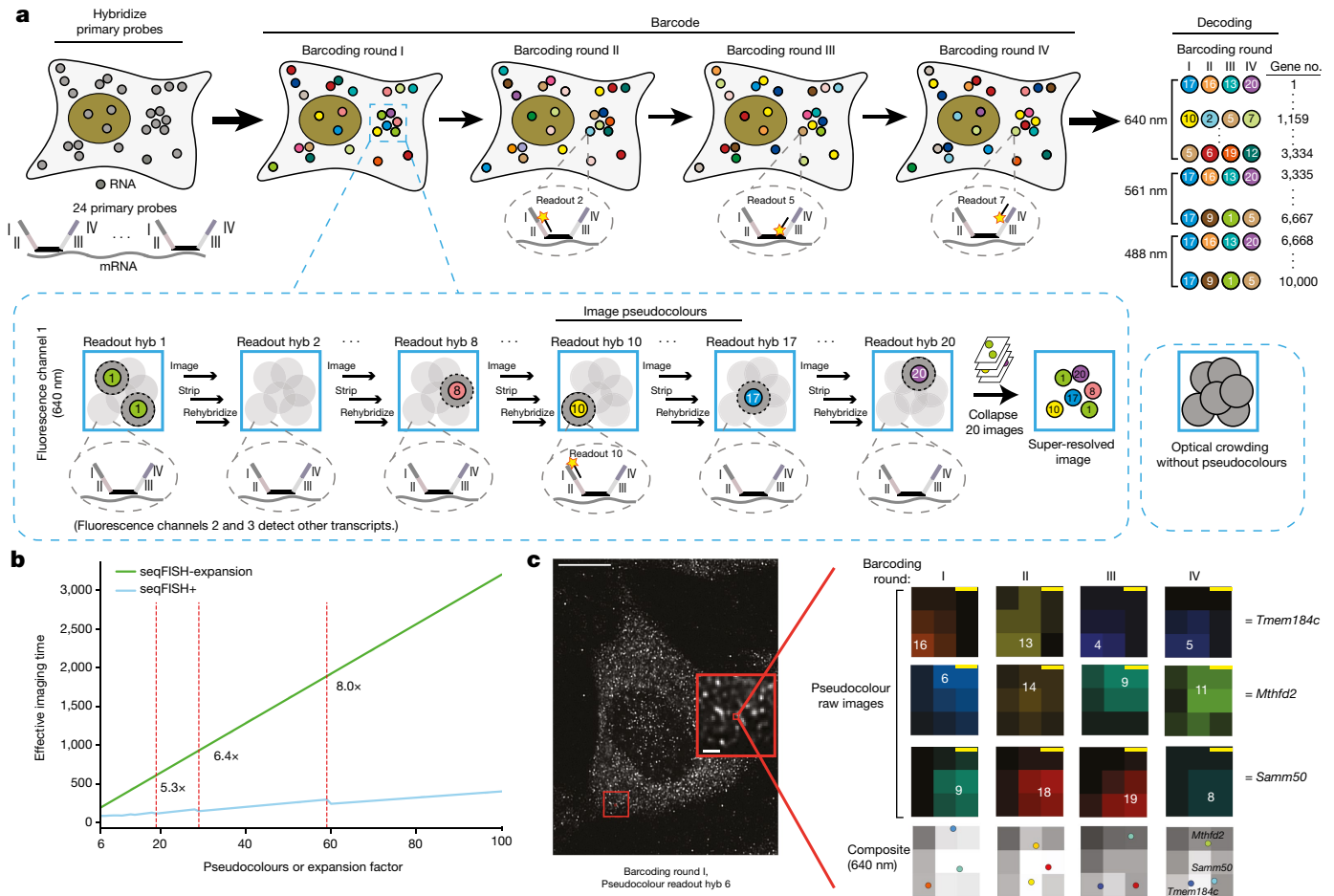
Fig. 1 | seqFISH+ resolves optical crowding and enables transcriptome profiling in situ. a, Schematic of the seqFISH+ protocol. Primary probes (24 per gene) against 10,000 genes are hybridized in cells. Overhang sequences (I–IV) on the primary probes correspond to four rounds of barcoding. Only one twentieth of the total genes in each fluorescent channel are labelled by readout probes in each pseudocolour hybridization readout round (hyb), lowering the density of transcripts in each image. mRNA dots in each pseudocolour can then be localized by Gaussian fitting and collapsed into a super-resolved image (dashed blue outline). Each gene is barcoded within only one fluorescent channel (Methods). **b**, Comparison with seqFISH implemented with expansion microscopy

(seqFISH-expansion, green line), covering 24,000 genes. Imaging is eightfold faster using seqFISH+ with 60 pseudocolours (blue line) than with seqFISH-expansion (see Methods). **c**, Image of a NIH/3T3 cell from one round of hybridization ($n = 227$ cells; scale bar, 10 μm). Expanded inset shows individual mRNAs (scale bar, 1 μm). Different mRNAs are decoded within a diffraction-limited region, magnified from the inset (scale bars, 100 nm). The number in each panel corresponds to the pseudocolour round in which each mRNA was detected, with no dots being detected during the other pseudocolour rounds in the channel (640 nm).

seqFISH+ enables visualization of the subcellular localization patterns of tens of thousands of RNA molecules in situ in single cells. Three major clusters of localization were observed, with enrichment in either nucleus–perinucleus, cytoplasm or protrusions. We found many mRNAs that had not been previously localized to protrusions, in addition to those that had been identified before[22,23]. We further observed three distinct subclusters in the perinuclear–nuclear-localized transcripts, and genes in each of these subclusters were enriched in distinct functional roles (Extended Data Fig. 3f–j).

To demonstrate that seqFISH+ works robustly in tissues, we used the same 10,000-gene probe set to image cells in the cortex, subventricular zone (SVZ) (Fig. 3a) and olfactory bulb in two separate sections of mouse brain. We collected 10,000-gene profiles for 2,963 cells (Fig. 3b–e), covering an area of approximately 0.5 mm[2]. In the cortex, cells contained on average $5,615 \pm 3,307$ (mean $\pm$ s.d.) transcripts from $3,338 \pm 1,489$ (mean $\pm$ s.d.) detected genes (Extended Data Fig. 4a, b). We imaged only a single optical plane (0.75-μm thickness) to save imaging time. Full 3D imaging of cells with seqFISH+ can detect five-to-tenfold more transcripts per cell.

With an unsupervised clustering analysis[24], the seqFISH+ cell clusters show clear layer structures (Fig. 3h) and are strongly

correlated to the clusters in an scRNA-seq[25] dataset (Methods, Extended Data Figs. 4c–f, 5). Similar layer patterns are observed with hidden-Markov random-field (HMRF) analysis[14], in which the expression patterns of neighbouring cells were taken into account (Extended Data Figs. 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 cell types, the transcripts for *Snrnp70*, a small nuclear riboprotein, and *Nr4a1*, a nuclear receptor, are found in the nuclear–perinuclear regions. By contrast, transcripts of the Na$^+$/K$^+$-ATPase *Atp1b2* and the kinesin *Kif5a* are observed near the cell peripheries in many cell types, including excitatory and inhibitory neurons as well as glial cells. In addition, many transcripts in astrocytes, such as *Gja1* and *Htra1*, localize to the cell periphery and processes; these results were confirmed by smFISH (Fig. 3f, g, Extended Data Fig. 7).

We next explored the spatial organization of the SVZ (Fig. 4a). 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 expressing *Dlx1* and *Sp9*, consistent with previous studies[26]. We further quantified the spatial organization of the different cell types in the SVZ (Fig. 4a,
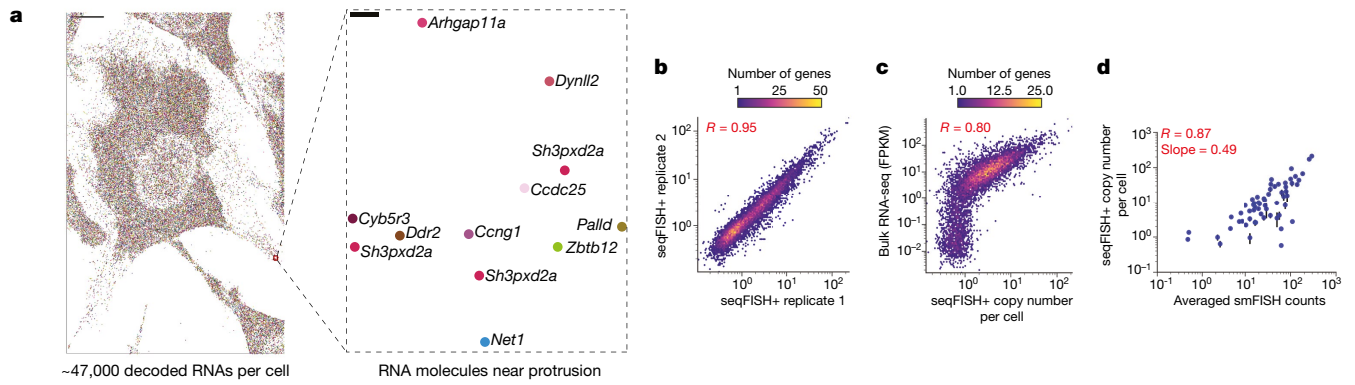
Fig. 2 | seqFISH+ profiles 10,000 genes in cells with high efficiency.
a, Approximately 47,000 mRNAs (coloured dots) were identified in a NIH/3T3 cell from a single z-section (scale bar, 10 μm). Inset shows the transcripts decoded in cell protrusions ($n = 227$ cells; scale bar, 100 nm). **b**, seqFISH+ replicates in NIH/3T3 cells are highly reproducible

(replicate 1, $n = 103$ cells; replicate 2, $n = 124$ cells). **c, d**, seqFISH+ correlates well with RNA-seq ($n = 9,875$ genes) (**c**) and smFISH ($n = 60$ genes; $P = 2.26 \times 10^{-19}$) (**d**). The efficiency of seqFISH+ is about 49% compared to smFISH. Error bars in **d** represent s.e.m. In **b–d**, $P < 0.0001$ (Pearson's R, two-tailed P values).

Extended Data Fig. 8), and found that class 12 and 17 neuroblasts are preferentially in contact, whereas TAP cells tend to associate with other TAP cells. Further investigation of the RNA velocity trajectories[27] of these cells in situ with intron seqFISH[9] and of their lineage relationships

using MEMOIR[28] would reveal further information regarding their temporal and spatial relationships.

Next, we examined the spatial organization of the olfactory bulb. Our clustering analysis revealed distinct classes of GABAergic interneurons,
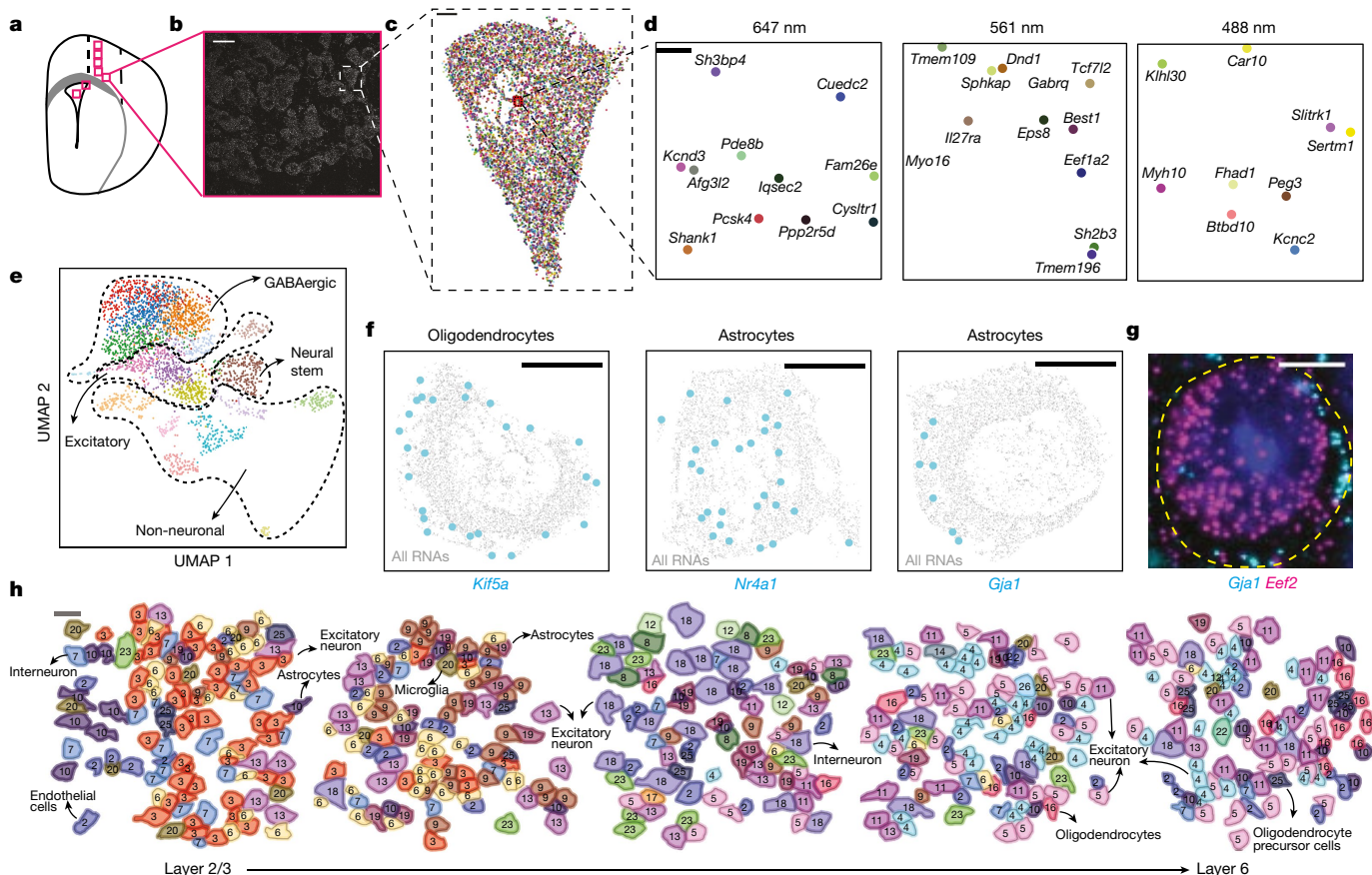

Fig. 3 | seqFISH+ robustly characterizes cell classes and subcellular RNA localization in brain slices. a, Schematic of the imaged regions (magenta boxes). **b**, Cells in a single FOV of the cortex (scale bar, 20 μm). **c**, Reconstruction of the 9,418 mRNAs (coloured dots) detected in a cell (scale bar, 2 μm). **d**, Decoded transcripts for a magnified region ($n = 523$ cells; scale bar, 100 nm). *Fam26e* is also known as *Calhm5*. **e**, Uniform manifold approximation and projection (UMAP) representation of the seqFISH+ data in the cortex, SVZ and olfactory bulb ($n = 2,963$ cells).

f, Reconstructed seqFISH+ images show subcellular localization patterns for mRNAs (cyan) in different cell types ($n = 62$ astrocytes and 28 oligodendrocytes; scale bars, 2 μm). **g**, smFISH of *Gja1* in cortical astrocytes shows localization at the periphery, compared with the uniform distribution of *Eef2* mRNAs. $n = 10$ FOVs, 40× objective; scale bar, 5 μm. **h**, Each cortex layer has a distinct cell-class composition (see annotations, Supplementary Table 2). Scale bar, 20 μm.
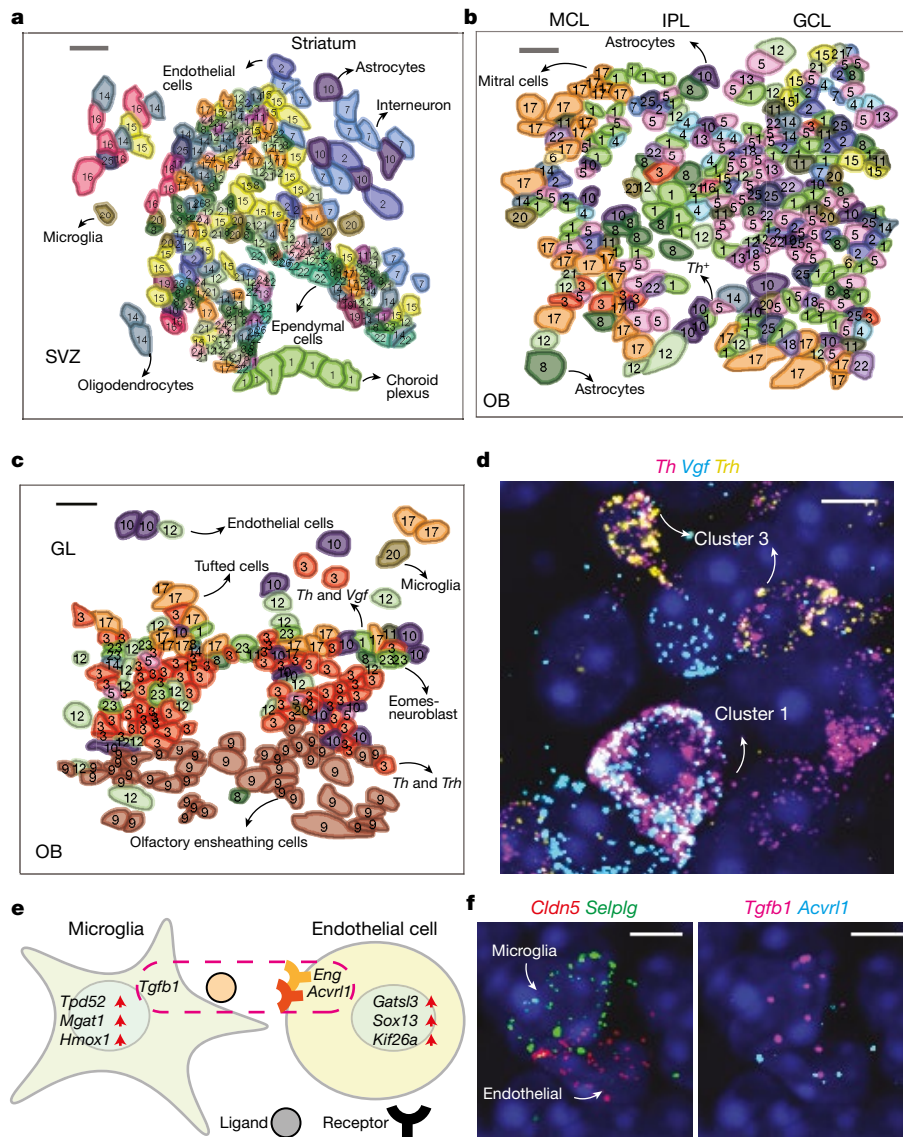
Fig. 4 | seqFISH+ reveals ligand–receptor repertoires in neighbouring cells and spatial organization in tissues. a, Spatial organization of distinct cell clusters in the SVZ. **b, c**, Spatially resolved cell-cluster maps of the mitral cell layer (MCL), inner plexiform layer (IPL), granule cell layer (GCL) (**b**) and glomerular layer (GL) (**c**) (scale bars, 20 μm). Remaining FOVs are shown in Extended Data Fig. 10. Note that different numbering is used for clusters in the SVZ and olfactory bulb (OB) (Supplementary Table 2). **d**, Distinct populations of Th^+ dopaminergic neurons in the olfactory bulb with differential expression of *Vgf* and *Trh*, shown with smFISH, confirming seqFISH+ clustering analysis. **e**, Schematic showing ligand–receptor pairs in neighbouring microglia and endothelial cells. In microglia next to endothelial cells, certain genes, such as *Tpd52*, are enriched compared with microglia that neighbour other cell types. **f**, mRNAs of *Tgfb1* ligand and *Acvrl1* receptor are visualized in adjacent microglia-endothelial cells by smFISH. In **d, f**, $n = 10$ FOVs, 40× objective; scale bars, 5 μm.

olfactory ensheathing cells (OECs), astrocytes, microglia and endo-thelial cells (Fig. 4b, c), consistent with previous studies[29]. In the granule cell layer (GCL) at the centre 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 comprising more mature interneurons (Fig. 4b, Extended Data Figs. 9, 10). Excitatory clusters of cells expressing *Reln* and *Slc17a7* are observed in the mitral cell layer as mitral cells and in the external plexiform layer and the glomerulus as tufted cells. We also found several clusters of dopaminergic neurons expressing tyrosine hydroxylase (*Th*) (Fig. 4b–d, Supplementary Table 2). For example, cluster 1 cells express *Th* and the neuropeptide *Vgf*, and are distributed in both the glomerulus and the GCL. Similarly, *Trh* is enriched in a distinct set of Th^+ cells (cluster 3), which are predominantly found in the glomerulus, whereas dopaminergic neurons in clusters 5 and 22 are located in the GCL. We validated these clusters by smFISH imaging (Fig. 4d, Extended Data Figs. 9, 10).

Finally, we analysed ligand–receptor pairs that are enriched in neigh-bouring cells—information that is not available in a dissociated-cell analysis. These potential cell–cell interactions are proposed on the basis of mRNA and not protein. Endothelial cells adjacent to microglia in the olfactory bulb express endoglin (*Eng*, a type III TGFβ receptor) and activin A receptor (*Acvrl1* (also known as *Alk1*), a type I TGFβ recep-tor) mRNAs, and the microglia express TGFβ ligand (*Tgfb1*) mRNA. By contrast, endothelial cells adjacent to microglia in the cortex express *Lrp1* (also known as *Tgfbr5*) and *Pdgfb*, indicating that signalling 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 are highly dependent on the local tissue context of neighbouring cells (Fig. 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 a tenfold or greater improvement over

existing methods in the number of mRNAs profiled and the total number of RNA barcodes detected per cell. seqFISH+ also enables 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 signalling interactions between cells is a crucial first step towards understanding developmental processes and cell-fate decisions, along with explorations of the combinatorial signalling logic[21]. Finally, the genomics coverage of seqFISH+ will enable the discovery of targets that are cell-type specific in disease samples, as well as enabling precise spatial genomics and single-cell-based diagnostics testing.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1049-y.

1. 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).
2. 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).
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).
4. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
5. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
6. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).
7. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
8. Eng, C. L., Shah, S., Thomassie, J. & Cai, L. Profiling the transcriptome with RNA SPOTs. *Nat. Methods* **14**, 1153–1155 (2017).
9. Shah, S. et al. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* **174**, 363–376.e16 (2018).
10. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
11. Lubeck, E. & Cai, L. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* **9**, 743–748 (2012).
12. Betzig, E. et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
13. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006).
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.* **36**, 1183–1190 (2018).
15. Thompson, R. E., Larson, D. R. & Webb, W. W. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* **82**, 2775–2783 (2002).
16. Yildiz, A., Tomishige, M., Vale, R. D. & Selvin, P. R. Kinesin walks hand-over-hand. *Science* **303**, 676–678 (2004).
17. Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science* **347**, 543–548 (2015).
18. Yang, B. et al. Single-cell phenotyping within transparent intact tissue through whole-body clearing. *Cell* **158**, 945–958 (2014).
19. Chen, F. et al. Nanoscale imaging of RNA with expansion microscopy. *Nat. Methods* **13**, 679–684 (2016).
20. Moffitt, J. R. et al. High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proc. Natl Acad. Sci. USA* **113**, 14456–14461 (2016).
21. Antebi, Y. E. et al. Combinatorial signal perception in the BMP pathway. *Cell* **170**, 1184–1196 (2017).
22. Mili, S., Moissoglu, K. & Macara, I. G. Genome-wide screen reveals APC-associated RNAs enriched in cell protrusions. *Nature* **453**, 115–119 (2008).
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).
24. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).
25. Tasic, B. et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
26. Shah, P. T. et al. Single-cell transcriptomics and fate mapping of ependymal cells reveals an absence of neural stem cell function. *Cell* **173**, 1045–1057 (2018).
27. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
28. Frieda, K. L. et al. Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
29. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).
30. Takei, Y., Shah, S., Harvey, S., Qi, L. S. & Cai, L. Multiplexed dynamic imaging of genomic loci by combined CRISPR imaging and DNA sequential FISH. *Biophys. J.* **112**, 1773–1776 (2017).

Reviewer information *Nature* thanks Samantha Morris, Arjun Raj and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions C.-H.L.E. and L.C. conceived the idea and designed experiments. C.-H.L.E. performed all the experiments. M.L. performed image analysis. C.-H.L.E., M.L., Q.Z., R.D. and L.C. performed data analysis. L.C. and G.-C.Y. supervised the analysis process. C.-H.L.E. and N.K. performed cell segmentation and generated the primary probes. Y.T. designed the readout probes. C.-H.L.E., Y.T. and J.Y. validated the readout probes. C.C. and C.K. built the automated fluidic delivery system. C.-H.L.E., M.L, Q.Z., R.D. and Y.T. provided input for L.C. when writing the manuscript. L.C. supervised all aspects of the project.

Competing interests C.-H.L.E and L.C. filed a patent on the pseudocolour-encoding scheme in seqFISH+.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1049-y.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1049-y.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to L.C.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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 and 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-nucleotide (nt) sequences of each gene were extracted, first using the exons from within the coding region. For genes that did not yield enough target sequences from the coding region, exons from both the coding and untranslated regions were used. The masked genome and annotation from the University of California Santa Cruz (UCSC) were used to look up the gene sequences. Probe sequences were required to have GC content within the range 45–65%. Any probe sequences that contained five or more consecutive bases of the same kind were dropped. Any genes that did 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. Readout probes of 15 nt in length were designed as previously described[9]. In brief, a set of probe sequences was randomly generated with combinations of A, T, G or C nucleotides. Readout-probe sequences within a GC-content range of 40–60% were selected. We performed a BLAST search against the mouse transcriptome to ensure the specificity of the readout probes. To minimize cross-hybridization of the readout probes, any probes with ten 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]. In brief, limited PCR cycles were used to amplify the designated probe sequences from the oligo complex pool. Then, the amplified PCR products were purified using the 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 a 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 h 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 single-stranded DNA (ssDNA) probes were alkaline hydrolysed with 1 M NaOH at 65 °C for 15 min 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 used 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 drying by speed-vac to powder, and resuspended in primary-probe hybridization buffer comprising 40% formamide (F9027, Sigma), 2× SSC (15557036, Thermo Fisher) and 10% (w/v) dextran sulfate (D8906; Sigma). The probes were stored at −20 °C until use.

Readout-probe synthesis. Readout probes of 15 nt in length were ordered from Integrated DNA Technologies as 5′-amine modified. The construction of readout probes was similar to that previously described[9]. In brief, 5 nmol DNA probes was 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% dimethyl fumarate. The reaction was allowed to go on for at least 6 h at 37 °C. Then, the DNA probes were subjected to ethanol precipitation, high pressure liquid chromatography (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 stored at −20 °C.

Coverslip functionalization. For cell culture experiments, coverslips were cleaned with a plasma cleaner on a high setting (PDC-001, Harrick Plasma) for 5 min, followed by immersion in 1% bind-silane solution (GE; 17-1330-01) made in pH 3.5 10% (v/v) acidic ethanol solution for 30 min at room temperature. Then, the coverslips were rinsed with 100% ethanol three times, and heat-dried in an oven at >90 °C for 30 min. Next, the coverslips were treated with 100 μg μl⁻¹ of poly-D-lysine (P6407; Sigma) in water for >1 h at room temperature, followed by

three rinses with water. The coverslips were then air-dried and kept at 4 °C for no longer than 2 weeks. For mouse brain-slice experiments, coverslips were cleaned with 1 M HCl at room temperature for 1 h, rinsed with water once, and treated with 1 M NaOH solution at room temperature for 1 h. Then, the coverslips were rinsed three times with water, before immersion in 1% bind-silane solution for 1 h at room temperature. The remaining steps were the same as the coverslip functionalization for cell culture.

seqFISH+ encoding strategy. We separated the 60 pseudocolours into three fluorescent channels (Alexa Fluor 488, Cy3b and Alexa Fluor 647) equally. In each channel, the 20-pseudocolour imaging was repeated three times, hence achieving 20^3 (that is, 8,000) genes barcoding capacity. We performed an extra round of pseudocolour imaging to obtain error-correctable barcodes, an error-correction scheme that we had previously introduced[3]. Thus, we obtained 8,000 error-correctable barcodes in three fluorescent channels—a 24,000 error-correctable barcoding capacity in total. More fluorescent channels and/or more pseudocolours can be added easily to achieve greater dilution of the mRNA density per imaging round. In this experiment, we encoded 3,333, 3,333 and 3,334 genes in each of the fluorescent channels. This pseudocolour scheme evolved from the one used in RNA sequential probing of targets (RNA SPOTs)[8] and intron seqFISH[9], by eliminating chromatic aberration and drastically 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 four-unit base-20 barcode unique to each gene. Hybridization with fluorophore-labelled 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 four-unit barcode for each gene. Each round of stripping and readout hybridization is fast and is completed in minutes.

After hybridization of primary probes, the samples were subjected to hydrogel embedding and clearing before seqFISH+ imaging. The details are available in the 'Cell culture experiment', 'Tissue slices experiment' and 'seqFISH+ imaging' sections below.

Cell culture experiment. NIH/3T3 cells (ATCC) were cultured as previously described[8] on the functionalized coverslips to ~80–90% confluence. Then, the cells were washed with 1× PBS once and fixed with freshly made 4% formaldehyde (28906; Thermo Fisher) in 1× PBS (AM9624, Invitrogen) at room temperature for 10 min. The fixed cells were permeabilized with 70% ethanol for 1 h 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 h in a humid chamber at 37 °C. We found hybridization for 48 h yielded slightly brighter signals. After hybridization, the samples were washed with 40% formamide in 2× SSC at 37 °C for 30 min, followed by three rinses with 1 ml 2× SSC. Next, the cell samples were incubated with a 1:1,000 dilution of Tetraspeck beads in 2× SSC at room temperature for 5–10 min. The density of the beads could be easily adjusted by varying the dilution factor or incubation time. Then, the samples were rinsed with 2× SSC and incubated with degassed 4% acrylamide (1610154; Bio-Rad) solution in 2× SSC for 5 min at room temperature. To initiate polymerization, the 4% acrylamide solution was aspirated, then 10 μl of 4% hydrogel solution containing 4% acrylamide (1:19), 2× 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 occur at room temperature for 1 h in a home-made 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 h at 37 °C[19]. The digestion buffer consisted 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 2× SSC multiple times and subjected to Label-IT modification (1:10) (MIR 3900; Mirus Bio) at 37 °C for 30 min. After that, the cell samples were post-fixed with 4% PFA in 1× PBS to stabilize the DNA, RNA and the overall cell sample for 15 min at room temperature. The reaction was quenched by 1 M pH 8.0 Tris HCl at room temperature for 10 min. The cell samples were either imaged immediately or kept in 4× SSC supplemented with 2 U μl⁻¹ of SUPERase In RNase inhibitor (AM2696; Invitrogen) at 4 °C for no longer than 6 h.

Mice. All animal care and experiments were carried out in accordance with Caltech Institutional Animal Care and Use Committee (IACUC) and NIH guidelines. Wild-type mice C57BL/6J of post-natal day (P)23 (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 cluster validation in the olfactory bulb, a section from P40 mice was used.

Tissue slices experiment. Brain extraction was performed as previously described[3]. In brief, mice were perfused for 8 min with perfusion buffer (10 U ml^{-1} heparin, 0.5% NaNO$_2$ (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 min. The mouse brain was removed from the skull and immediately placed in a 4% PFA buffer for 2 h at room temperature under gentle mixing. The brain was then immersed in 4 °C 30% RNase-free sucrose (Amresco 0335) in 1× PBS until the brain sank. After the brain sank, the brain was frozen in a dry ice–isopropanol bath in OCT medium and stored at −80 °C. Five-micrometre 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 h. Then, the tissue slices were cleared with 8% SDS (AM9822; Invitrogen) in 1× PBS for 30 min 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 h at 37 °C. After primary probe hybridization, the tissue slices were washed with 40% formamide at 37 °C for 30 min. After rinsing with 2× SSC three times and 1× PBS once, the sample was subjected to treatment with 0.1 mg ml^{-1} Acryloyl-X succinimidyl ester (A20770; Thermo Fisher) in 1× PBS for 30 min at room temperature. Then, the tissue slices were incubated with 4% acrylamide (1:19 crosslinking) hydrogel solution in 2× SSC for 30 min at room temperature. The hydrogel solution was aspirated and 20 μl of 4% hydrogel solution containing 0.05% APS and 0.05% TEMED in 2× SSC was dropped onto the tissue slice and sandwiched by Gel-Slick functionalized slides. The samples were transferred to 4 °C in a home-made nitrogen gas chamber for 30 min before transferring to 37 °C for 2.5 h to complete polymerization. After polymerization, the hydrogel-embedded tissue slices were cleared with digestion buffer as mentioned above, except that it included 1% SDS, for >3 h at 37 °C. After digestion, the tissue slices were rinsed with 2× SSC multiple times and subjected to 0.1 mg ml^{-1} label-X modification[19] for 45 min at 37 °C. The preparation of label-X stock was as previously described[19]. To further stabilize the DNA probes, RNA molecules, and the overall structure of the tissue slices, the tissue slices were re-embedded in hydrogel solution as in the previous step, except that the gelation time could be shortened to 2 h. The tissue slice samples were either imaged immediately or kept in 4× SSC supplemented with 2 U μl^{-1} of SUPERase In RNase Inhibitor at 4 °C for no longer than 6 h.

seqFISH+ imaging. The imaging platform and automated fluidics delivery system were similar to those previously described[9] with some modifications. In brief, the flow cell on the sample was first connected to the automated fluidics system. Then the region of interest (ROI) was registered using nuclei signals stained with 10 μg ml^{-1} 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 (10% ethylene carbonate (E26258; Sigma), 10% dextran sulfate (D4911; Sigma), 4× 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) was pipetted into a 96-well plate. During each serial hybridization, the automated sampler moves to the well of the designated hybridization buffer and moves the 100 μl hybridization buffer through a multichannel fluidic valve (EZ1213-820-4; IDEX Health & Science) to the flow cell (requires ∼25 μl) using a syringe pump (63133-01, Hamilton Company). The serial hybridization solution was incubated for 17 min for cell culture experiments and 20 min 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 2× SSC) to remove excess readout probes and non-specific binding. Then, the sample was rinsed with ∼200 μl of 4× SSC supplemented with a 1:1,000 dilution of SUPERase In RNase Inhibitor, before being stained with DAPI solution (10 μg ml^{-1} of DAPI, 4× SSC, and a 1:1,000 dilution of SUPERase In RNase inhibitor) for ∼15 s. Next, an anti-bleaching buffer solution made of 10% (w/v) glucose, 1:100 diluted catalase (Sigma C3155), 0.5 mg ml^{-1} glucose oxidase (Sigma G2133), 0.02 U μl^{-1} SUPERase In RNase inhibitor and 50 mM pH 8 Tris-HCl in 4× SSC was flowed through the samples. Imaging was done with a microscope (Leica DMi8) equipped with a confocal scanner unit (Yokogawa CSU-W1), a sCMOS camera (Andor Zyla 4.2 Plus), a 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 (55% formamide and 0.1% Triton-X 100 in 2× SSC) was flowed through for 1 min, followed by an incubation time of 1 min before rinsing with

4× SSC solution. In general, the 15-nt readouts were stripped off within seconds, and a 2-min wash ensured the removal of any residual signal. The serial hybridization, imaging and signal extinguishing steps were repeated for 80 rounds. Then, staining buffer for segmentation (10 μg ml^{-1} DAPI, 50 nM LNA T20-Alexa Fluor 647 and a 1:100 dilution of Nissl stain (N21480; Invitrogen)) in 1× PBS was flowed in and allowed to incubate for 30 min at room temperature before imaging. The integration of automated fluidics delivery system and imaging was controlled by a custom written script in μManager[32].

smFISH. smFISH experiments were performed 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 (10 nM per probe) in 40% hybridization buffer (40% formamide, 10% dextran sulfate and 2× SSC) at 37 °C overnight. The sample was washed with 40% wash buffer for 30 min at 37 °C and subjected to the same hydrogel embedding and clearing as the cell culture experiment before imaging. The imaging platform was the same as the one in the seqFISH+ experiment. A single z-slice across hundreds of cells was imaged and the sum of the gene counts per cell was analysed using a custom written Matlab script. For smFISH experiments in the tissue, the sample was hybridized with 10 nM per probe in 40% hybridization buffer at 37 °C for >16 h. The sample was washed with 40% wash buffer for 30 min at 37 °C and subjected to the same hydrogel embedding and clearing as the tissue experiment before imaging. As the imaging time is short, the Acryloyl-X functionalization and post-hydrogel anchoring steps were omitted. Five z-slices with z-steps of 1 μm were taken across multiple FOVs with the imaging platform as in the seqFISH+ experiment, except that a 40× 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 Fisher) 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 function. An initial estimate of the transformation matrix between the DAPI image for each serial hybridization round and the beads-only 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 function. A final transformation matrix between each hybridization image and the beads-only 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 × 7 pixels) as an estimate for the point spread function. Cell segmentation was performed manually using the ROI tool in ImageJ.

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 radialcentre.m[33]. Once all potential points in all serial hybridizations of one fluorescent channel were obtained, they were organized by pseudocolour 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 neighbours. Point combinations that constructed only a single barcode were immediately matched to the on-target barcode set. For point combinations that matched to construct multiple barcodes, the point sets were first 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 on the basis of 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-nm, 561-nm and 488-nm channels. The remaining unused barcodes were used as an off-target evaluation by repeating the same procedure as described.

Data analysis. *RNA-seq and RNA SPOTs.* Pearson's *R* correlation was performed to compare seqFISH+ data to RNA-seq (Gene Expression Omnibus (GEO) accession number GSE98674), RNA SPOTs[8] and smFISH measurement using Matlab or Python functions.

Spatial clustering of genes for NIH/3T3 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 dimensions of the 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 ten copies per FOV 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 ten genes were dropped.

Hierarchical clustering of brain seqFISH+ data. The 10,000 genes were divided into three approximately equal subsets (with 3,334, 3,333 and 3,333 genes, respectively) on the basis of the group in which genes were barcoded. Genes were normalized separately within each subset, by dividing the gene counts per cell by the total counts per cell within each subset. We then multiplied the result by the scaling factor 2,000, which is approximately the median count. Next, we selected the subset of cells that were in the cortex. We computed $\log(1 + \text{normalized counts})$.

To select genes for clustering, we first computed statistics for the following criteria for each gene: (1) number of cells with non-zero expression; (2) average gene expression of all cells; (3) average expression of top 5% of cells with highest expression; (4) average of top 10% of cells with highest expression; (5) average of top 2% of cells with highest expression; and (6) average gene expression of all non-zero 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 six gene lists, forming an initial 3,877-gene set. The reasoning is that the union of genes would contain genes needed to cluster both 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 and 5). The 3,877-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 to further select the most relevant genes for clustering. Specifically, the jackstraw procedure[34] permutes the expression of a small number of genes to identify significant genes (that is, genes with significantly higher loading than the permuted case ($P < 0.001$)). Using the top nine components, we found a total of 1,916 significant genes to be used for final clustering.

To this 1,916-gene matrix we applied hierarchical clustering with Ward's linkage and with $(1 - \text{Pearson correlation})$ 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 the 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 scRNA-seq data. Mouse visual cortex scRNA-seq data were obtained from ref. [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 1,857 genes that were commonly profiled by scRNA-seq and seqFISH+, and we processed the scRNA-seq data in the same way as for seqFISH+. The degree of similarity was evaluated by using the Pearson correlation (Extended Data Fig. 4c).

Supervised mapping of cell types from scRNA-seq to seqFISH+. Cell-type mapping was performed as described[14]. In brief, MAST[36] was used to identify differentially expressed genes across annotated cell types from the scRNA-seq dataset in ref. [25], using $P = 0.005$ as the cut-off. A total of 1,253 of the differentially expressed genes were also profiled by seqFISH+, and therefore were retained for cell-type mapping. Then, we performed a quantile normalization on the expression vectors of each gene in both the seqFISH+ and scRNA-seq data to normalize cross-platform differences[14]. Multi-class support-vector machine models were trained on the scRNA-seq cell types using linear kernels, and the tuning parameter C was set to 1×10^{-5} (Extended Data Fig. 4d). The cross-validation accuracy of prediction of the 22 annotated cell types was 91% with these 1,253 differentially expressed genes.

Spatial gene identification. In brief, we computed a spatial score per gene, as previously described[14]. Cells were divided into two sets based on the basis of each gene: L1 contains cells with highest 90th percentile by expression, and L0 contains 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 the dissimilarity matrix to rank-transformed Euclidean distance, examine_top = 0.1, permutation_test = True, and permutations = 1,000. 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 selected all spatial genes with a significant silhouette coefficient ($P < 0.01$ permutation test). To further enrich for spatial signals within these genes, we performed

a PCA analysis, and then the jackstraw procedure[34], to arrive at a set of 988 spatial genes that were significantly correlated to the principle components. We performed HMRF analysis on the top 100, 200 and 400 of the 988 genes.

Spatial domain identification via HMRF procedure. HMRF is a probabilistic spatial-clustering method that we developed previously to identify spatial domains on the basis of spatial relationships and gene expression per cell. We constructed a neighbourhood graph by adopting a fixed radius corresponding to the top percentile of pairwise physical distances between cells, resulting in an average of five neighbours per cell. HMRF was run with the following parameters: tolerance $= 1 \times 10^{-10}$, $k = 9$, and convergence_error $= 1 \times 10^{-8}$. 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 were performed in Python using the package SCANPY[37]. We followed a standard procedure as suggested in the SCANPY reimplementation of Seurat's tutorial to analyse seqFISH+ data, with some modifications. For clustering all cells from mouse cortex, subventricular zone (SVZ), choroid plexus and olfactory bulb, we first normalized the counts per cell, then chose highly variable genes with >0.4 min_dispersion and 0.01 min_mean, with max_mean = 3. This yielded 3,509 genes. Then we took the logarithm of the data, regressed out the total count effect per cell and scaled the data to unit variance. We computed the PCAs and used the top principal components to compute the neighbourhood graph, before performing Louvain clustering. We used the rank_gene_groups function with raw data, and the top 20 enriched genes in each cluster were used to identify the clusters on the basis of marker gene annotation from scRNA-seq or DropSeq data[29,38]. We found that both hierarchical clustering and Louvain clustering yielded similar results despite using different methods.

To spatially map back the clusters on the raw image, we performed Louvain clustering on cortex, SVZ and choroid plexus data, and olfactory bulb data separately. Genes with max count greater than four 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 the analysis. The counts were normalized per cell and a minimum dispersion of greater than 0 with min_mean of 0.05 was chosen to filter out the variable genes. This yielded 1,813 genes for subsequent analysis. For the olfactory bulb, genes with max count greater than two across all cells were first chosen. Then the counts were normalized per cell. To obtain the highly variable genes, a threshold of min_mean of 0.05 and min_dispersion of 0.2 was chosen. This yields 1,972 genes for subsequent analysis. After choosing the highly variable genes, the data were subjected to PCA reduction, computed neighbourhood 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 UMAP was done with the SCANPY function. These cluster numbers were mapped back to the original data to visualize the spatial heterogeneity of different cell types across different parts of the tissues.

Calculation of the time acceleration of seqFISH+ versus seqFISH-expansion. For seqFISH-expansion, we assume that to code ~20,000 genes, the coding scheme uses four colours and eight rounds of hybridization ($4^7 = 16,384$ genes) with one round of error correction. Thus, the total number of effective imaging per FOV is equal to the expansion factor $\times 4 \times 8$. For 60-fold expansion, this is $60 \times 4 \times 8 = 1,920$ images. For seqFISH+, we assume a coding scheme with three separate fluorescent channels, with 8,000 genes coded in each channel for a total of 24,000 genes. Pseudocolours are used to code for 8,000 genes. For example, if the number of pseudocolours is 20 per fluorescent channel, then four rounds of barcoding (including one round of error correction) are needed. The effective imaging per FOV is then $20 \times 4 \times 3 = 240$ images—an eightfold acceleration compared to seqFISH-expansion. As another example, if the pseudocolour per channel is ten, then five rounds of barcoding are needed to cover 8,000 genes per channel. Then, there is a total of $10 \times 5 \times 3 = 150$ images. However, this coding scheme only provides a $10 \times 3 = 30$-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 calculated the cell-to-cell correlation matrix with the number of genes downsampled from the 2,511 genes that expressed at least 5 copies in a cell. For each downsampled dataset, 100, 250, 500, 1,000, 1,500 and 2,000 genes were selected randomly. The Pearson's correlation coefficient of each of the cell-to-cell correlation matrices was computed with the cell-to-cell correlation matrix for the 2,511-gene dataset. Five trials were simulated for each downsampled gene level. Error bars denote standard deviation.

Neighbour-cell analysis. The spatial coordinates for the cell centroids were used to create a nearest-neighbour network ($k = 4$), in which nodes represent individual cells and edges are observed proximities between two cells. Edges between

identical or different annotated cell types were labelled as homo- or heterotypic, respectively. 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 than 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 the two cell types that were in close proximity according to the spatial network. The number of observed edges between two cell types and the 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 neighbouring cells, we extracted genes that had z-scores of 1 or higher, were expressed in at least 25% of the cells in the interacting pairs, and had at least 4 instances of being neighbours. We then matched up the ligand–receptor pairs from the literature[39], as 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 ($n = 1,000$) distribution by reshuffling the cell labels. $P < 0.05$ was deemed to be significant.

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

Contact maps. The minimum distances between the pixels defining the edge of all pairs of cells in a FOV 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 as being 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.

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

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Code availability. The custom written scripts used in this study are available at https://github.com/CaiGroup/seqFISH-PLUS.

Data availability

RNA-seq data were obtained from GEO accession number GSE98674. RNA SPOTs data were obtained from a previous study[8]. Source data from this study are available at https://github.com/CaiGroup/seqFISH-PLUS. All data obtained during this study are available from the corresponding author upon reasonable request.

31. Wang, G., Moffitt, J. R. & Zhuang, X. Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci. Rep.* **8**, 4847 (2018).
32. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using µManager. *Curr. Protoc. Mol. Biol.* **92**, 14.20.1–14.20.17 (2010).
33. Parthasarathy, R. Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods* **9**, 724–726 (2012).
34. Chung, N. C. & Storey, J. D. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* **31**, 545–554 (2015).
35. Huang, H., Liu, Y., Yuan, M. & Marron, J. S. Statistical significance of clustering using soft thresholding. *J. Comput. Graph. Stat.* **24**, 975–993 (2015).
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).
37. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
38. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 (2018).
39. Ramilowski, J. A. et al. A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).
40. Eng, C.-H. L. & Cai, L. RNA seqFISH+ supplementary protocol. *Protoc. Exch.* https://doi.org/10.1038/protex.2019.019 (2019).
41. Lein, E.S. et al. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
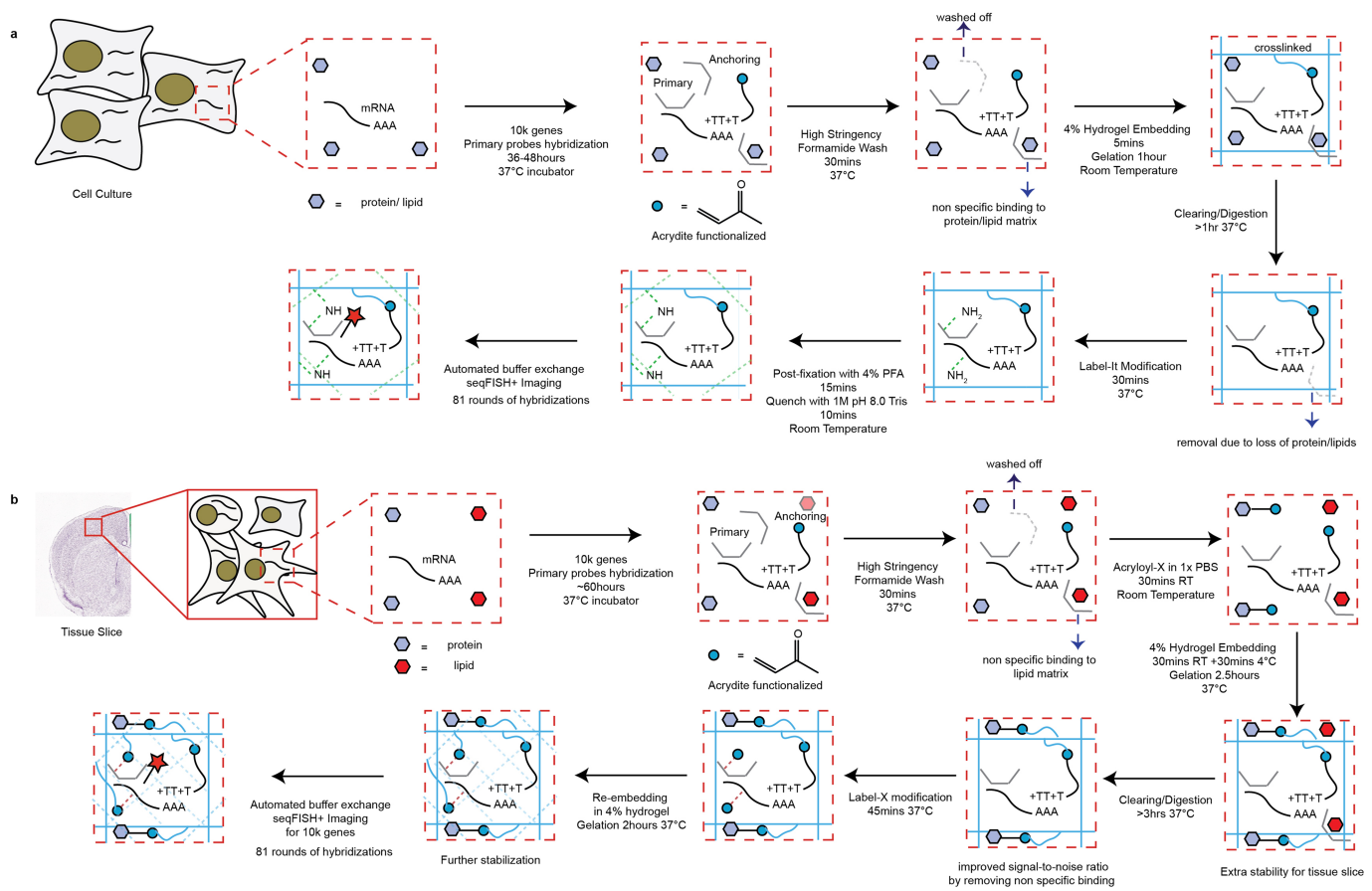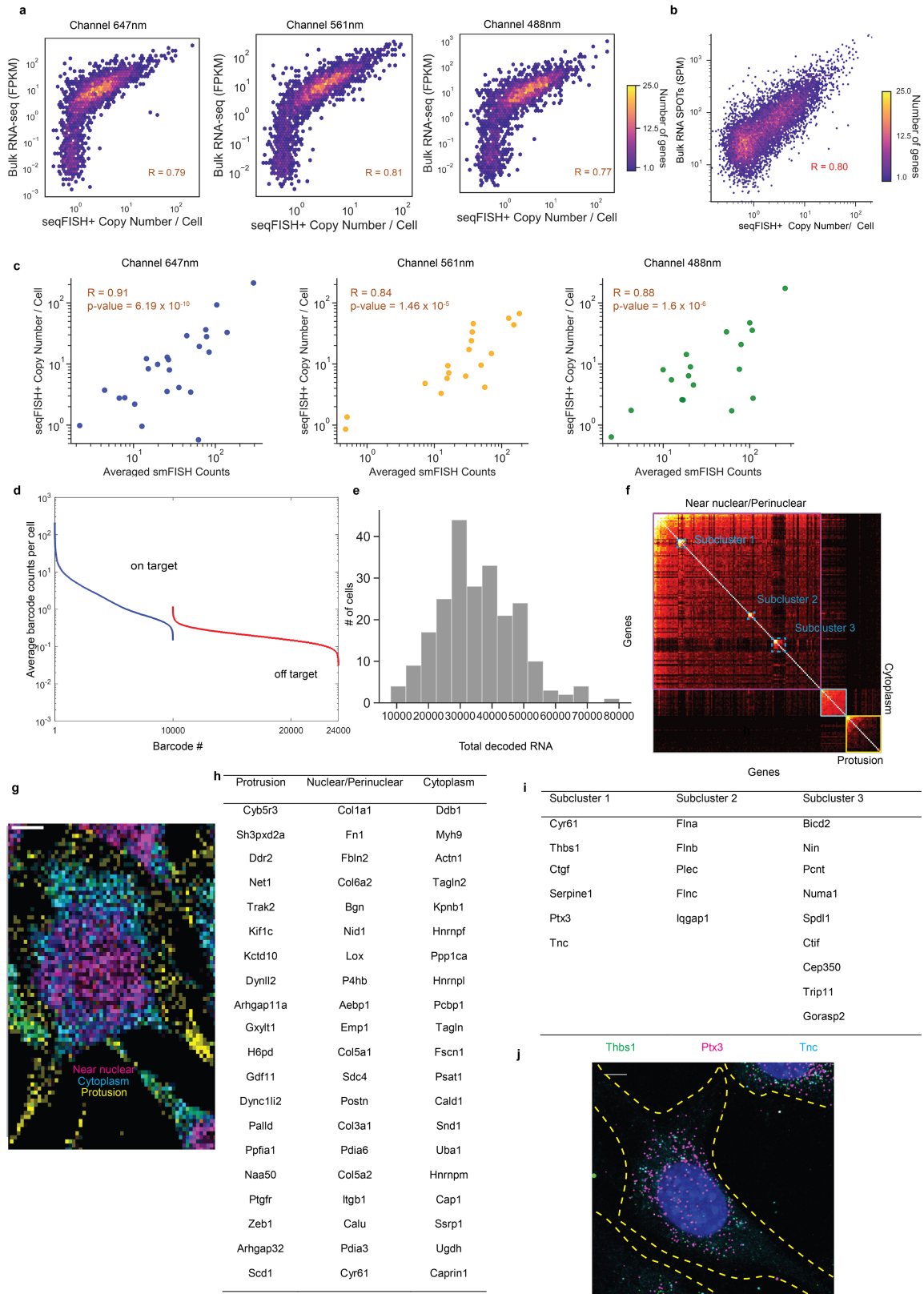
Extended Data Fig. 1 | Clearing and probe-anchoring protocols. a, **b**, seqFISH+ experiments in NIH/3T3 cells (**a**) and the mouse brain slices (**b**).
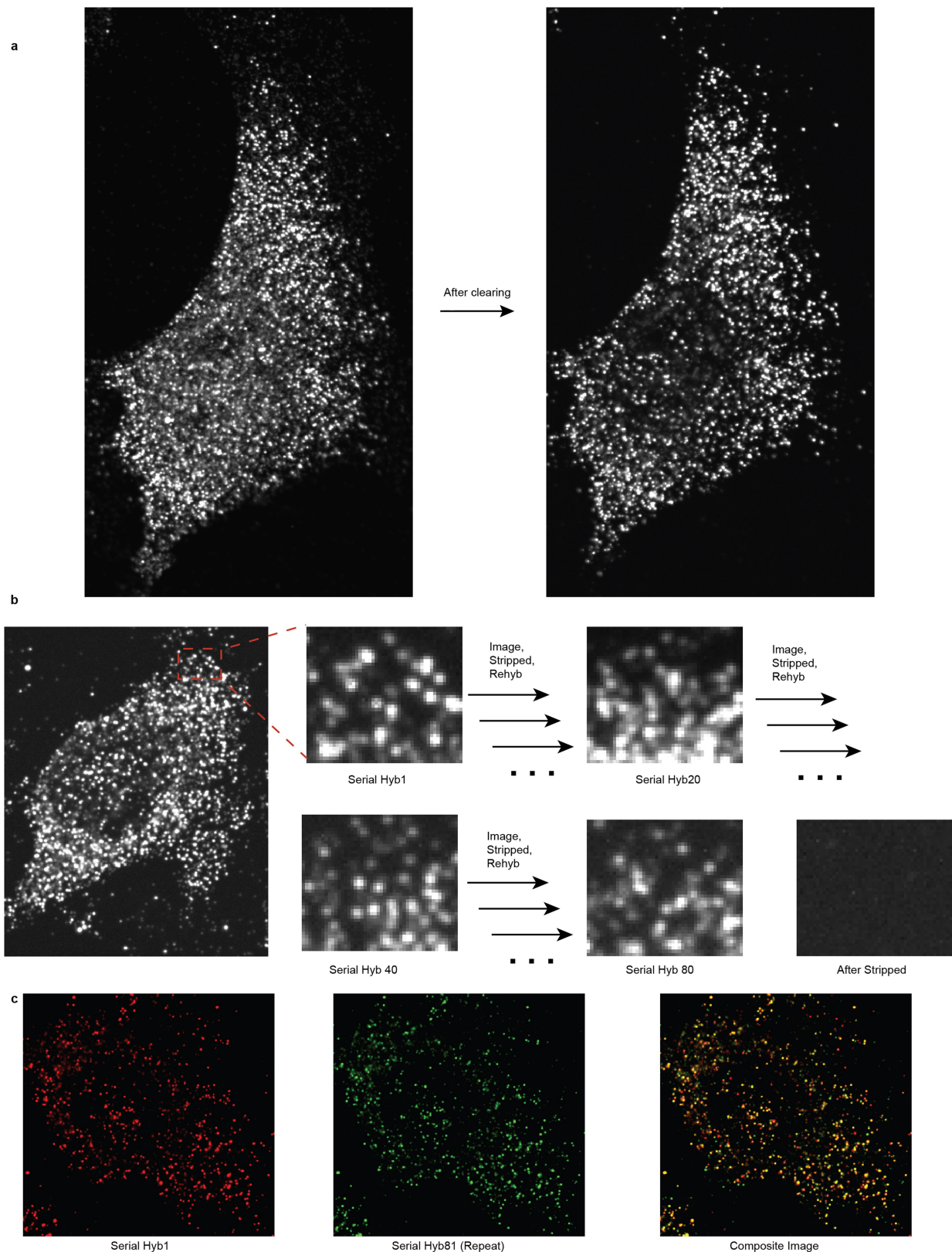
Extended Data Fig. 2 | Clearing removes background non-specific bound dots. a, Raw images of a NIH/3T3 cell before and after clearing. A marked decrease in background is observed in the cleared sample. The image was acquired on a spinning disk confocal microscope. **b**, In each round of hybridization for the 10,000-gene experiment, diffraction-limited dots are clearly separated, indicating that the pseudocolour scheme effectively dilutes the density of the sample. The signal is completely removed between different rounds of hybridization, with no 'cross-talk' between the pseudocolours. Stripping is accomplished by a 55% formamide wash, which is highly efficient. **c**, After the completion of each seqFISH+ experiment, the readout probes used in hybridization 1 are rehybridized in round 81. The colocalization rates between rounds 1 and 81 are 76% (647-nm channel), 73% (561-nm channel) and 80% (488-nm channel) within a two-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 ($n = 227$ cells).

a Channel 647nm / Channel 561nm / Channel 488nm — Bulk RNA-seq (FPKM) vs seqFISH+ Copy Number / Cell (R = 0.79, R = 0.81, R = 0.77)

b Bulk RNA SPOTs (SPM) vs seqFISH+ Copy Number / Cell (R = 0.80)

c Channel 647nm (R = 0.91, p-value = 6.19×10^{-10}), Channel 561nm (R = 0.84, p-value = 1.46×10^{-5}), Channel 488nm (R = 0.88, p-value = 1.6×10^{-6}) — seqFISH+ Copy Number / Cell vs Averaged smFISH Counts

d Average barcode counts per cell vs Barcode # (on target / off target)

e # of cells vs Total decoded RNA

f Genes vs Genes — Near nuclear/Perinuclear, Cytoplasm, Protrusion (Subcluster 1, Subcluster 2, Subcluster 3)

g Near nuclear / Cytoplasm / Protrusion

h

Protrusion	Nuclear/Perinuclear	Cytoplasm
Cyb5r3	Col1a1	Ddb1
Sh3pxd2a	Fn1	Myh9
Ddr2	Fbln2	Actn1
Net1	Col6a2	Tagln2
Trak2	Bgn	Kpnb1
Kif1c	Nid1	Hnrnpf
Kctd10	Lox	Ppp1ca
Dynll2	P4hb	Hnrnpl
Arhgap11a	Aebp1	Pcbp1
Gxylt1	Emp1	Tagln
H6pd	Col5a1	Fscn1
Gdf11	Sdc4	Psat1
Dync1li2	Postn	Cald1
Palld	Col3a1	Snd1
Ppfia1	Pdia6	Uba1
Naa50	Col5a2	Hnrnpm
Ptgfr	Itgb1	Cap1
Zeb1	Calu	Ssrp1
Arhgap32	Pdia3	Ugdh
Scd1	Cyr61	Caprin1

i

Subcluster 1	Subcluster 2	Subcluster 3
Cyr61	Flna	Bicd2
Thbs1	Flnb	Nin
Ctgf	Plec	Pcnt
Serpine1	Flnc	Numa1
Ptx3	Iqgap1	Spdl1
Tnc		Ctif
		Cep350
		Trip11
		Gorasp2

j Thbs1 / Ptx3 / Tnc

Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | seqFISH+ works efficiently across all three fluorescent channels and identifies localization patterns of transcripts in NIH/3T3 cells. a, Correlation plots between seqFISH+ and bulk RNA-seq in three fluorescent channels. Barcodes are coded entirely within each channel, with $n = 3,334$, 3,333 and 3,333 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 NIH/3T3 cells. SPM, spots per million. **c**, Correlation between seqFISH+ and smFISH for each fluorescent channel (from left to right, $n = 24$, 18 and 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 $\pm$ s.d.) per barcode. **e**, Histogram of the total number of mRNAs detected per NIH/3T3 cell. On average, $35,492 \pm 12,222$ transcripts are detected per cell. **f**, Genes are clustered on the basis of co-occurrence in a 10×10-pixel window. Three major clusters are nuclear–perinuclear, cytoplasmic and protrusions. **g**, mRNAs show preferential spatial localization patterns: nuclear, cytoplasm and protrusions ($n = 227$ cells). The image is binned into $1\,\mu m \times 1\,\mu m$ windows and coloured on the basis of the genes enriched in each bin (scale bar, $10\,\mu m$). **h**, Examples of genes enriched in each spatial cluster. **i**, Genes in the subclusters within the nuclear-localized group. Subcluster 1 contains genes that encode for extracellular matrix proteins. Genes in subcluster 2 are involved in the actin cytoskeleton, whereas genes in subcluster 3 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, $40\times$ objective). Scale bar, $10\,\mu m$.
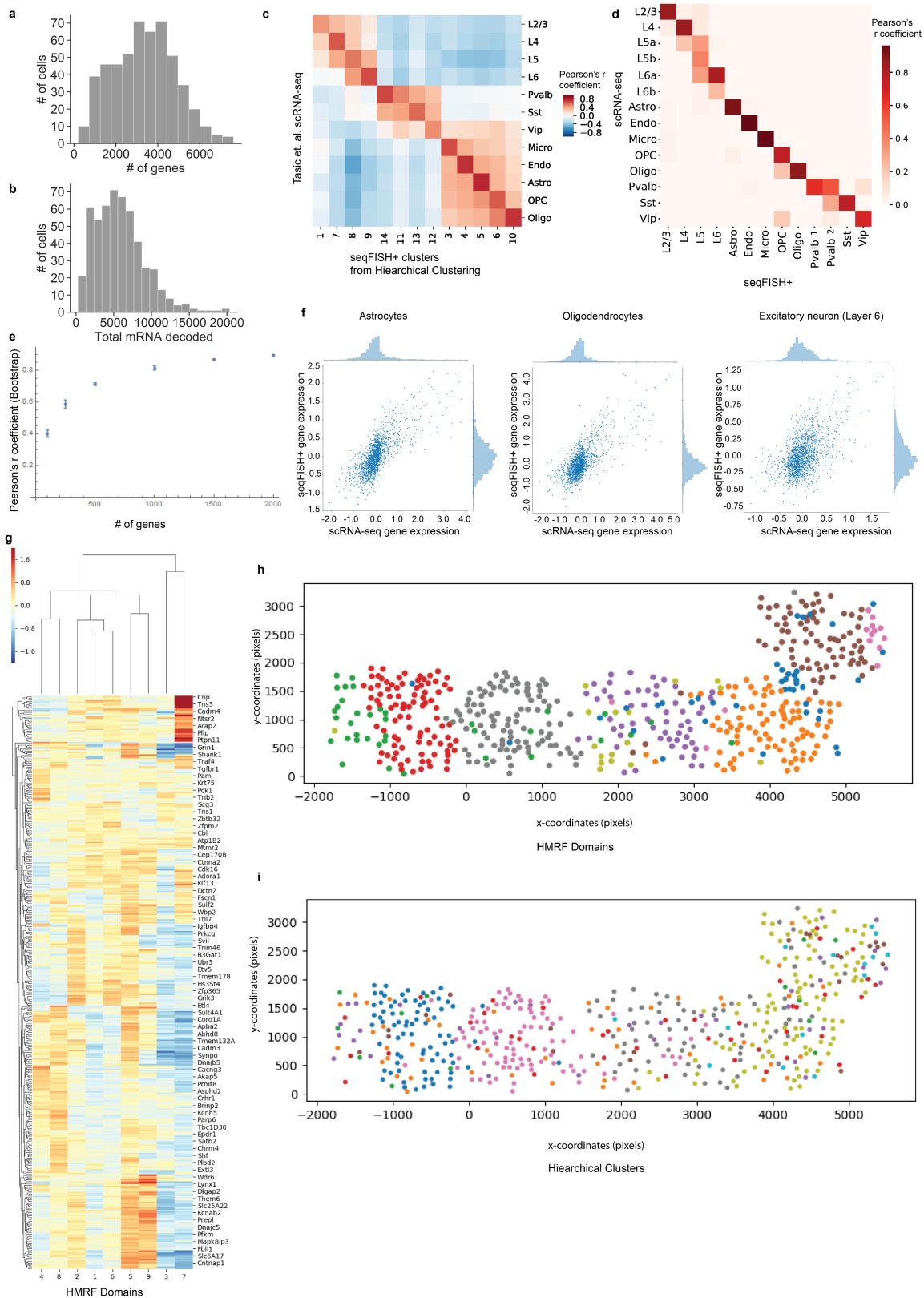
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | scRNA-seq comparison with seqFISH+, bootstrap and HMRF analysis. a, **b**, Histogram of the number of genes (**a**) and total RNA barcodes (**b**) detected per cell by seqFISH+ in the cortex. **c**, Unsupervised clustering of seqFISH+ correlates well with scRNA-seq ($n = 1,857$ genes; Pearson's R coefficient correlation)[25]. **d**, Supervised mapping of seqFISH+-analysed cortex cell clusters with those from scRNA-seq clusters ($n = 1,253$ genes; $P < 0.005$). **e**, The number of genes was downsampled from the 2,511 genes that expressed at least five 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 2,511-gene dataset. Five trials are simulated for e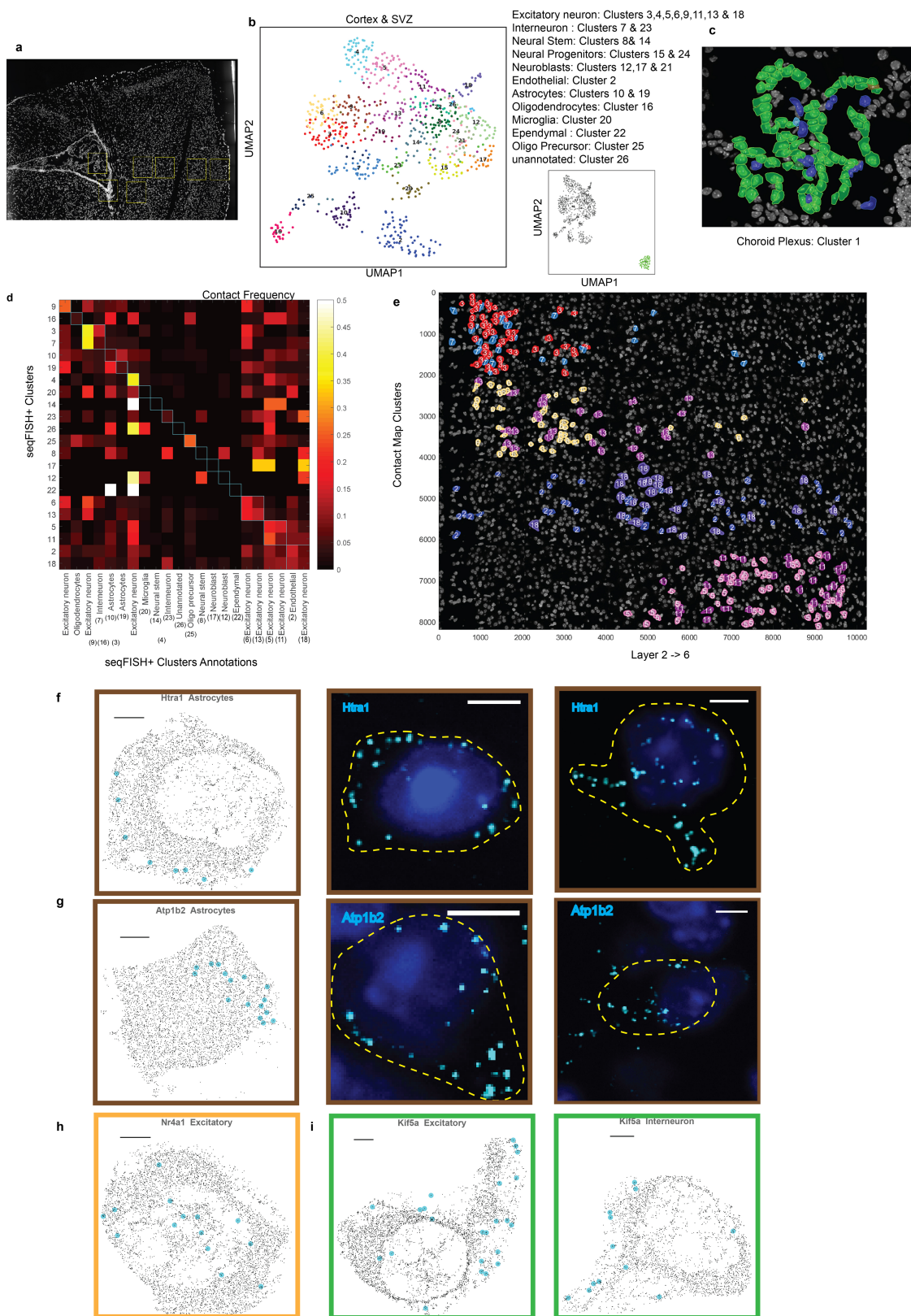ach downsampled gene level. Data are mean ± s.d. Even when downsampled to 100 genes, about 40% of the cell-to-cell correlation is retained, because the expression patterns of many genes are correlated. **f**, Scatter plots of seqFISH+ with scRNA-seq in different cell types. Each dot represents a gene and its mean expression z-score value in either seqFISH+ or scRNA-seq, in astrocytes, oligodendrocytes and excitatory neurons. In general, seqFISH+ and scRNA-seq 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. Coordinates are in units of one pixel (103 nm per pixel). Each camera FOV is 2,000 pixels.

a

seqFISH+

b
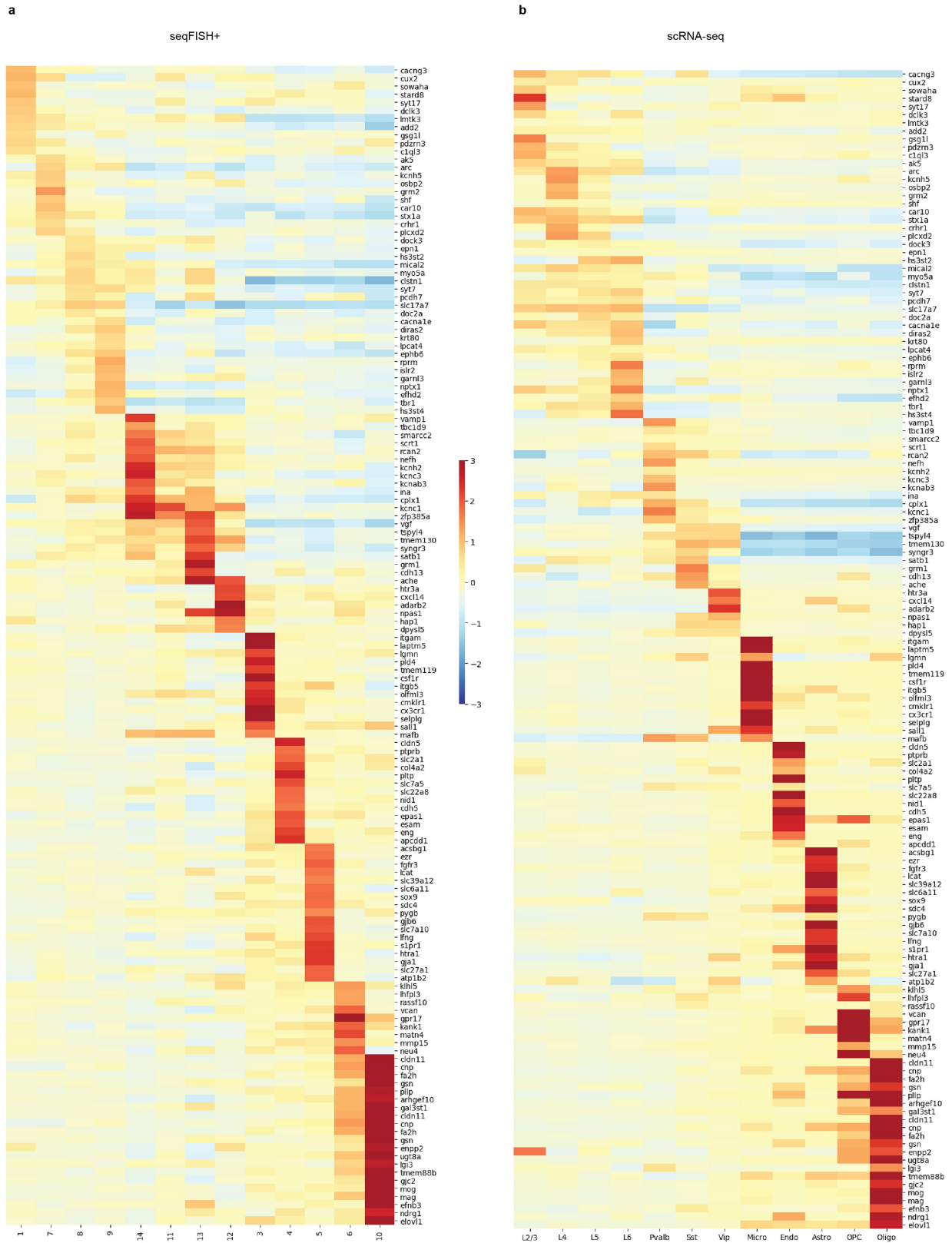
scRNA-seq

Extended Data Fig. 5 | Differential gene expressions between cell-type clusters. a, **b**, Expression measured by seqFISH+ (**a**) and scRNA-seq (**b**). The expression patterns of seqFISH+ clusters are similar to those shown by scRNA-seq clusters (*n* = 143 genes).
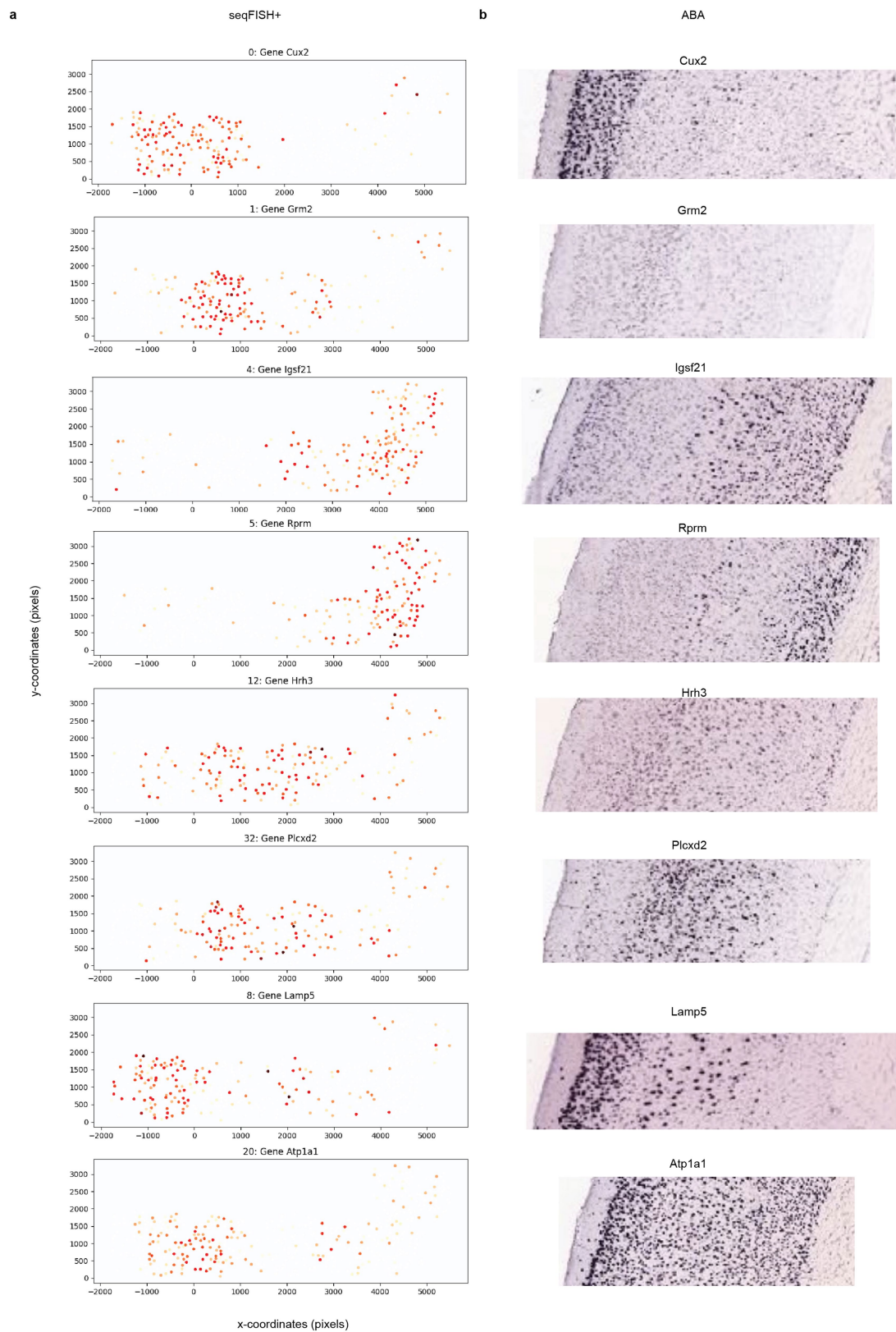
Extended Data Fig. 6 | Comparison of the spatial expression patterns across the cortex. a, **b**, seqFISH+ data (**a**) versus the Allen Brain Atlas (ABA)[41] (**b**). Coordinates are in units of one pixel (103 nm per pixel). Layers 1–6 are shown from left to right.

a

b Cortex & SVZ

Excitatory neuron: Clusters 3,4,5,6,9,11,13 & 18
Interneuron : Clusters 7 & 23
Neural Stem: Clusters 8& 14
Neural Progenitors: Clusters 15 & 24
Neuroblasts: Clusters 12,17 & 21
Endothelial: Cluster 2
Astrocytes: Clusters 10 & 19
Oligodendrocytes: Cluster 16
Microglia: Cluster 20
Ependymal : Cluster 22
Oligo Precursor: Cluster 25
unannotated: Cluster 26

c

Choroid Plexus: Cluster 1

d Contact Frequency

seqFISH+ Clusters Annotations

e

Layer 2 -> 6

f Htra1 Astrocytes

g Atp1b2 Astrocytes

h Nr4a1 Excitatory

i Kif5a Excitatory / Kif5a Interneuron

Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 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 classes in the cortex, normalized for the abundances of cells in each cluster. **e**, Each strip represents cells that cluster together, which breaks into layers in the cortex, consistent with expectations, as cells within a layer preferentially interact with each other ($n = 523$ cells). **f**, *Htra1* transcripts are preferentially localized to the periphery of the astrocytes in the cortex. Left, 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, 2 μm. Middle and right, a single *z*-slice of smFISH images of *Htra1* in cortical astrocytes (scale bars, 5 μm). **g**, *Atp1b2* localization in seqFISH+ (left; scale bar, 2 μm) and single *z*-slice smFISH images (middle and right; scale bars, 5 μm). 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, 40× 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, 2 μm). **i**, *Kif5a*, a kinesin, also exhibits periphery and process localizations in different cell types ($n = 60$ interneurons; scale bars, 2 μm).
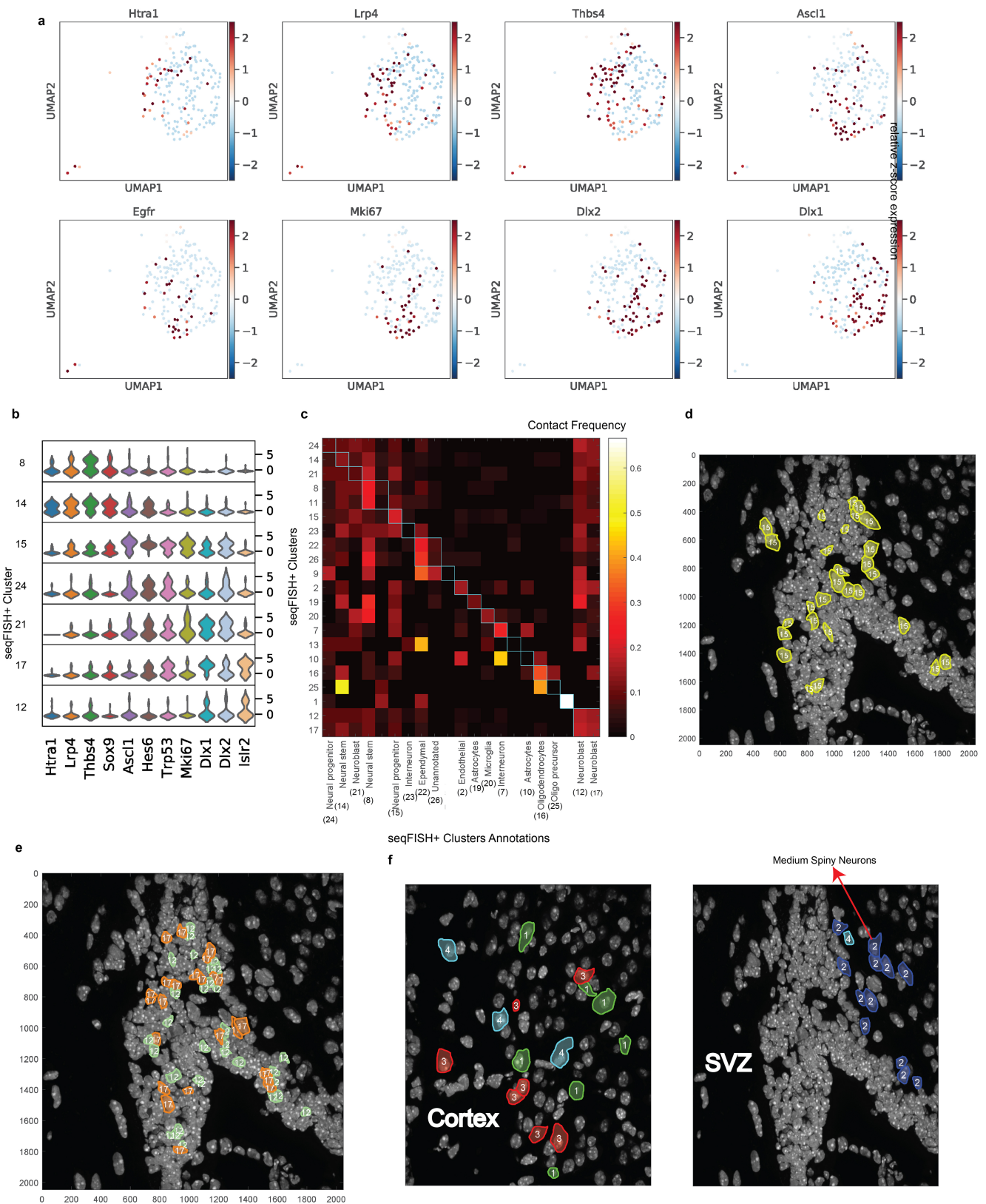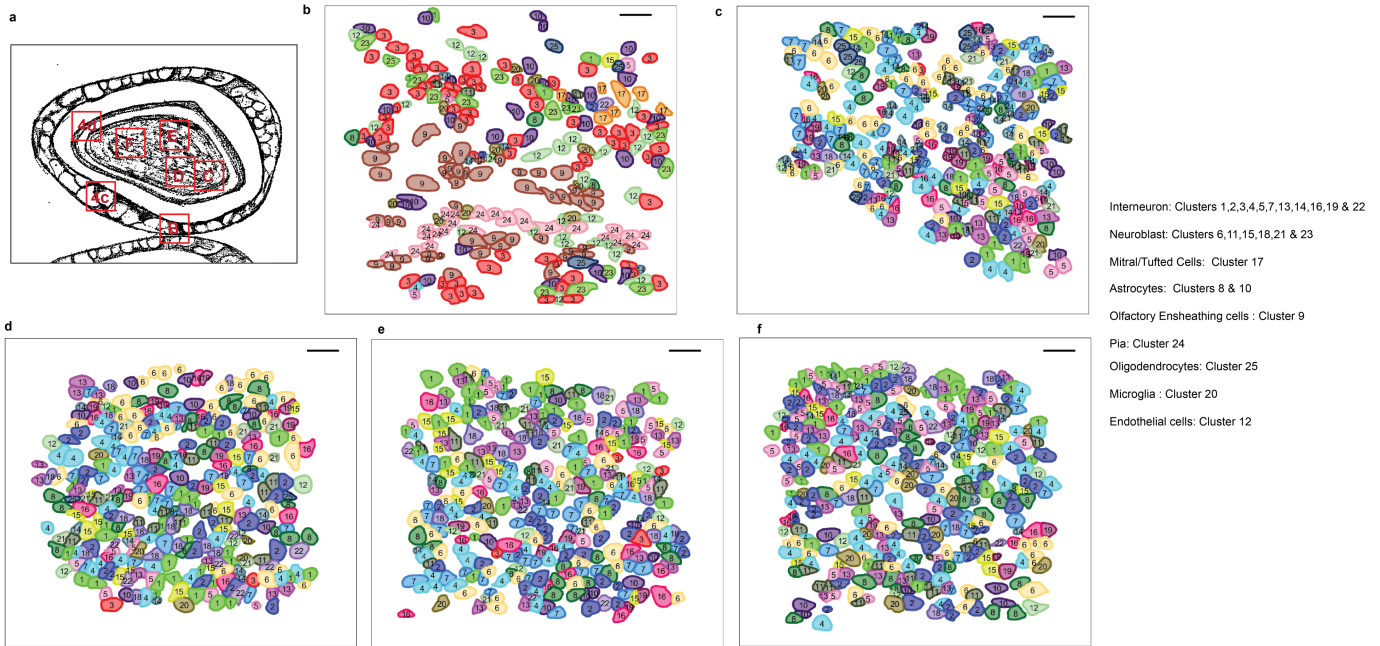
Extended Data Fig. 8 | Additional analysis of the SVZ. a, Expression of individual genes in the SVZ in UMAP representation ($n = 281$ cells). **b**, Violin plots showing z-scored gene expression patterns for Louvain clusters corresponding to NSCs for neuroblasts in the SVZ ($n = 281$ cells). **c**, Spatial proximity analysis of the cell clusters in the mouse SVZ. Frequency of contacts between the different cell classes in the SVZ, normalized for the abundances of cells in each cluster. **d**, Neural progenitors appear to be in spatial proximity with each other. $n = 281$ cells (**c**, **d**). **e**, Two neuroblast cell clusters are found to be in spatial proximity in the SVZ. **f**, Subclusters of cells from cluster 7 in the cortex (left). Medium spiny neurons that express *Adora2*, *Pde10a* and *Rasd2* marker genes form a separate cluster that is detected only in the striatum (right) ($n = 42$ cells in cluster 7).
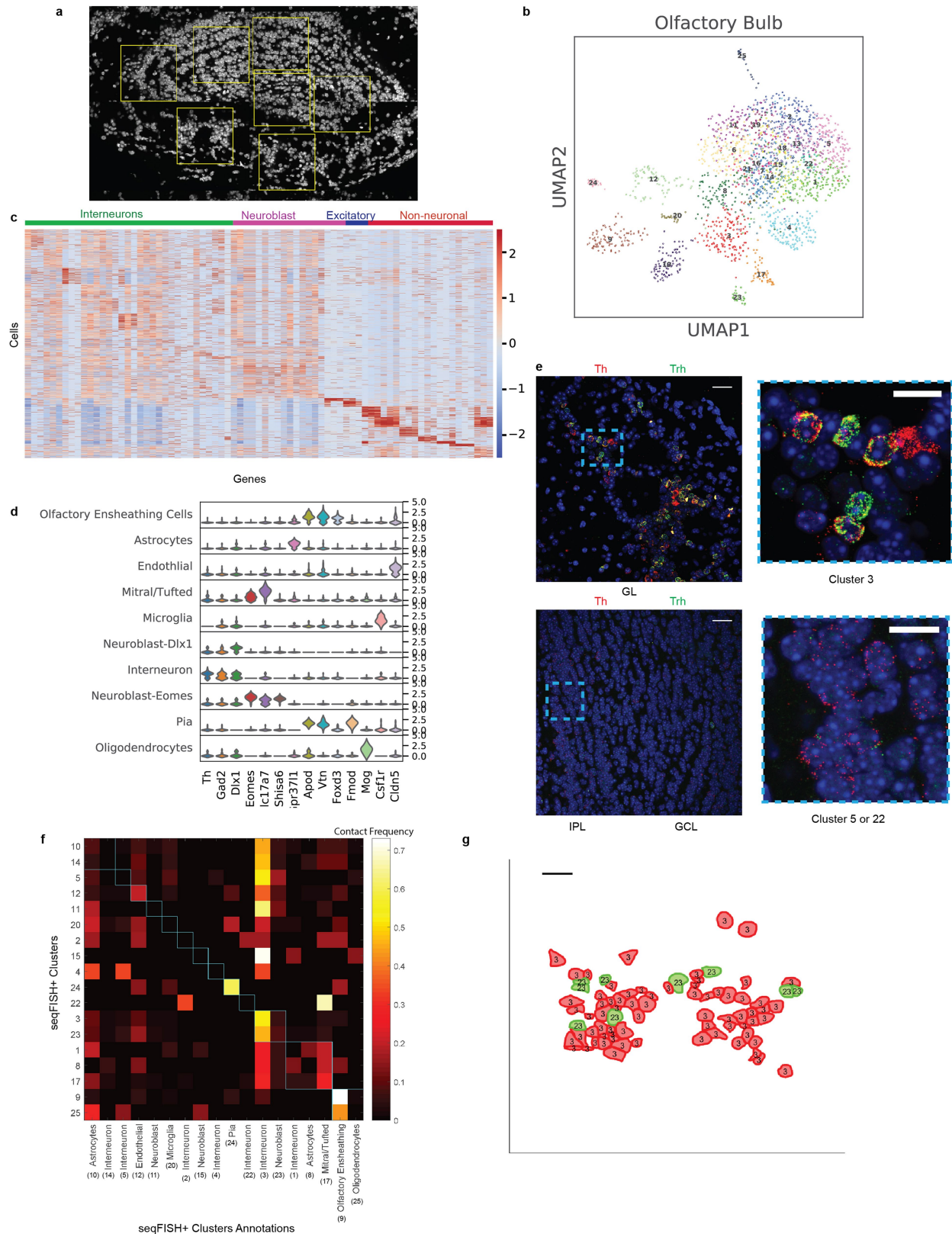
Extended Data Fig. 9 | Additional analysis of the olfactory bulb. a, Slide explorer image of the olfactory bulb FOVs imaged in the second brain slice. **b**, UMAP analysis of olfactory bulb cells. **c**, Heat map of *z*-scored gene expression patterns of cells in the olfactory bulb. **d**, Violin plots show *z*-scored marker gene expression patterns in the different classes of cells detected in the olfactory bulb. *n* = 2,050 cells (**a**–**d**). **e**, Representative smFISH images of *Th* and *Trh*. Images were maximum *z*-projected. In the glomeruli layer, cluster 3 cells express both *Th* and *Trh*, whereas in the GCL, cells express *Th* but not *Trh* (clusters 5 and 22). *n* = 10 FOVs, 40× objective. Scale bars, 13 μm (left images), 6.5 μm (right images). **f**, Frequency of contacts between the different cell classes 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. Scale bars, 20 μm (**f**, **g**).

Interneuron: Clusters 1,2,3,4,5,7,13,14,16,19 & 22

Neuroblast: Clusters 6,11,15,18,21 & 23

Mitral/Tufted Cells: Cluster 17

Astrocytes: Clusters 8 & 10

Olfactory Ensheathing cells : Cluster 9

Pia: Cluster 24

Oligodendrocytes: Cluster 25

Microglia : Cluster 20

Endothelial cells: Cluster 12

Extended Data Fig. 10 | Spatial organization of the olfactory bulb.
a, Schematics of the FOVs imaged in the olfactory bulb. **b**–**f**, Spatial mapping of the cell clusters in the glomeruli layer (**b**) and GCL (**c**–**f**) in the olfactory bulb. 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 GCL. Scale bars, 20 μm.

nature research

Corresponding author(s):	Long Cai
Last updated by author(s):	Jan 29, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Custom written scripts in Micro-Manager for automated fluids delivery and image acquisitions.
Data analysis	Custom written scripts in Matlab, Python 3, Scanpy, Mathematica, R, and Image J. See Methods for details on how each software is used.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Additional data from this study is available on https://github.com/CaiGroup/seqFISH-PLUS

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	No statistical methods were used to pre-determine sample size. Post hoc analysis showing high reproducibility and the agreement of cell type clusters with literature indicate that the sample size is sufficient.
Data exclusions	No raw data was excluded from the analyses. For downstream analysis, some single cells were excluded due to pre-established criteria. Details are included in the Methods section.
Replication	Pearson's r correlation was used to determine the reproducibility of the method. High correlation coefficient of 0.95 indicates that the method is robust and highly reproducible based on hundreds of cells from two independent experiments.
Randomization	The samples were not randomized in this study because only high quality samples/mice were used for the same experimental condition.
Blinding	There is no experimental group in this study and hence no blinding is needed.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	NIH/3T3 cell line from ATCC was used in this study.
Authentication	None of the cell lines have been authenticated.
Mycoplasma contamination	The cell line was not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	The cell line used was not listed in the ICLAC database.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	23days (Male), 40days(Male), and 10 weeks old(Female) wild type C57BL/6J mice were used in this study.
Wild animals	No wild animals were used in this study.
Field-collected samples	No field-collected samples were used in this study.
Ethics oversight	All animal care and experiments were carried out in accordance to Caltech Institutional Animal Care and Use Committee (IACUC) and NIH guidelines.

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