Three-dimensional intact-tissue sequencing of single-cell transcriptional states

Xiao Wang[1]*, William E. Allen[1,2]*, Matthew A. Wright[1,3], Emily L. Sylwestrak[1], Nikolay Samusik[4], Sam Vesuna[1], Kathryn Evans[1], Cindy Liu[1], Charu Ramakrishnan[1], Jia Liu[5], Garry P. Nolan[4]†, Felice-Alessio Bava[4]†, Karl Deisseroth[1,3,6]†

[1]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [2]Neuroscience Program, Stanford University, CA 94305, USA. [3]Department of Psychiatry and Behavioral Sciences, Stanford University, CA 94305, USA. [4]Baxter Laboratory, Department of Microbiology and Immunology, Stanford University, Stanford, CA 94305, USA. [5]Department of Chemical Engineering, Stanford University, CA 94305, USA. [6]Howard Hughes Medical Institute, Stanford University, CA 94305, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: deissero@stanford.edu (K.D.); alessio.bava@gmail.com (F.-A.B.); gnolan@stanford.edu (G.P.N.)

Retrieving high-content gene-expression information while retaining 3D positional anatomy at cellular resolution has been difficult, limiting integrative understanding of structure and function in complex biological tissues. Here we develop and apply a technology for 3D intact-tissue RNA sequencing, termed STARmap (Spatially-resolved Transcript Amplicon Readout Mapping), which integrates hydrogel-tissue chemistry, targeted signal amplification, and in situ sequencing. The capabilities of STARmap were tested by mapping 160 to 1,020 genes simultaneously in sections of mouse brain at single-cell resolution with high efficiency, accuracy and reproducibility. Moving to thick tissue blocks, we observed a molecularly-defined gradient distribution of excitatory-neuron subtypes across cubic millimeter-scale volumes (>30,000 cells), and discovered a short-range 3D self-clustering in many inhibitory-neuron subtypes that could be identified and described with 3D STARmap.

In biological tissues, diversity of function arises from diversity of form–in part via the complexity of cell-specific gene expression, which defines the unique three-dimensional molecular anatomy and cellular properties of each tissue. In situ transcriptomic tools for the spatial mapping of gene expression with subcellular resolution have emerged that may be applicable to probing these tissue structure-function relationships, including both multiplexed in situ RNA hybridization and in situ RNA sequencing (*1–10*). Current in situ sequencing approaches face the challenge of implementing enzymatic reactions in the dense, complex tissue environment and currently suffer from low efficiency (*2*), but the potential value of such intact-tissue sequencing could be enormous; in comparison to hybridization-based multiplexing/readout which utilizes multiple polynucleotide probes to encode gene identity (*3–5*), sequencing operates with single-nucleotide resolution, and thus inherently provides greater information. In addition, in situ sequencing methods typically utilize signal amplification, important for detection of short transcripts (e.g., neuropeptides) and for high-quality imaging in thick tissue blocks. However, current sequencing methods have not yet been successfully applicable to 3D volumes of intact tissue, due to fundamental limitations in requisite sensitivity, fidelity, and scalability for throughput in tissues such as the mammalian brain.

Hydrogels have been widely used for extracellular 3D scaffolding in applications across biology and medicine (*11–13*).

Recently-developed hydrogel-tissue chemistry (HTC) methodologies (*14*), beginning with CLARITY (*15*), physically link in situ-synthesized polymers with selected intracellular biomolecules. This process transforms the tissue, from within its constituent cells, into a new state suitable for high-resolution volumetric imaging and analysis compatible with many kinds of molecular phenotyping for proteins, nucleic acids, and other targets (*15*). HTC-based hydrogel-embedding strategies have been extended to nucleic acid analyses in the form of in situ hybridization for RNA (*16–19*), but these have not yet been extended to in situ RNA sequencing–which would have the potential to reveal the full molecular complexity of the transcriptome. In non-tissue environments, however, purely synthetic hydrogels have been used to accommodate enzymatic reactions that include DNA sequencing (*20*), and if biological tissue could be converted into a hydrogel-embedded form compatible with creation, retention, and functional presentation of RNA-derived or hybridized complementary DNA (cDNA), it might be possible to perform 3D in situ sequencing within such a tissue-hydrogel formulation–leveraging the crucial attendant properties of optical transparency, reduced background, elevated diffusion rate, and greater mechanical stability. Here we achieve this goal with the development and application of a sequencing-based method (Spatially-resolved Transcript Amplicon Readout Mapping, or STARmap) for targeted 3D in situ transcriptomics in intact

tissue (Fig. 1A); using STARmap we were able to identify organizational principles of a full spectrum of cell types, which would not have been otherwise accessible for identification, in the adult mammalian brain.

Results

Design and validation of STARmap principles

One component is an efficient approach for in situ-amplification of a library of cDNA probes hybridized with cellular RNAs (this approach is termed SNAIL, for Specific Amplification of Nucleic Acids via Intramolecular Ligation). Reverse transcription may be the major efficiency-limiting step for in situ sequencing (7, 21), and SNAIL bypasses this step with a pair of primer and padlock probes (fig. S1A) designed such that only when both probes hybridize to the same RNA molecule, the padlock probe can be circularized and rolling-circle-amplified to generate a DNA nanoball (amplicon) containing multiple copies of the cDNA probes (Fig. 1, A to D). This mechanism ensures target-specific signal amplification and excludes noise that invariably otherwise arises from non-specific hybridization of single probes. Indeed, the outcome includes much higher absolute intensity and signal-to-noise ratio (SNR) compared to commercial single-molecule fluorescent in situ hybridization (smFISH) probes (fig. S1, B to F), and substantial improvement of detection efficiency (comparable to single-cell RNA sequencing) with simplified experimental procedures compared to previous in situ RNA sequencing methods (fig. S1, G to I).

To enable cDNA amplicon embedding in the tissue-hydrogel setting, amine-modified nucleotides were spiked into the rolling circle amplification reaction, functionalized with an acrylamide moiety using acrylic acid N-hydroxysuccinimide esters, and copolymerized with acrylamide monomers to form a distinct kind of hydrogel-DNA amplicon network (Fig. 1A and fig. S2A). The resulting tissue-hydrogel was then subjected to protein digestion and lipid removal to enhance transparency (fig. S2, B to E). This design chemistry dictates that amplicons are covalently linked with the hydrogel network, and such crosslinking is essential to maintain the position and integrity of the amplicons through many cycles of detection (fig. S2, F to H).

A 5-base barcode (library size of 1,024) was designed and built into each padlock probe as a gene-unique identifier to be sequenced, thus enabling multiplexed gene detection (Fig. 1A). Sequencing-by-synthesis paradigms were avoided as these require elevated reaction temperatures, which in turn are problematic for high-resolution imaging and sample stability (16) in comparison to sequencing-by-ligation methods that can be implemented at room temperature. However, none of the reported or commercially available sequencing-by-ligation methods approach exhibit the necessary SNR or accuracy for this challenging intact-tissue application: SOLiD

sequencing causes strong background fluorescence in biological samples (10) while cPAL sequencing (22) lacks an error-rejection mechanism (fig. S3). For this reason, an approach we term Sequencing with Error-reduction by Dynamic Annealing and Ligation (SEDAL) was devised specifically for STARmap (fig. S3).

SEDAL employs two kinds of short, degenerate probes: reading-probes to decode bases, and fluorescence-probes to transduce decoded sequence information into fluorescence signals. The two short probes only transiently bind to the target DNA and ligate to form a stable product for imaging only when a perfect match occurs; after each cycle corresponding to a base readout, the fluorescent products are stripped by formamide, which eliminates error accumulation as sequencing proceeds (Fig. 1E and fig. S3B). In contrast to SOLiD, SEDAL exhibits minimal background (fig. S3, C to F). A 2-base encoding scheme was designed and implemented to mitigate any residual errors related to imaging high densities of spots (fig. S3, G and H). Based upon a panel of four very-highly expressed test genes in mouse brain (to mimic amplicon crowdedness as would be encountered in highly-multiplexed gene-detection), we found that the error rate of STARmap was more than an order of magnitude lower than prior cPAL methods (~1.8% versus 29.4%) (fig. S3, I to L) (17).

Spatial cell typing in primary visual cortex with 160-gene STARmapping

To test if STARmap could deliver on the initial goal of high-content 3D intact-tissue sequencing of single-cell transcriptional states with the necessary sensitivity and accuracy, we applied STARmap to a pressing current challenge in neuroscience: detecting and classifying cell types and corresponding tissue-organization principles in neocortex of the adult mouse brain. The anatomy and function of mouse primary visual neocortex have been extensively studied (23), a setting which here allows validation of our results by comparison with prior findings that span multiple papers, methodologies, and data sources (but the full diversity of deeply molecularly-defined cell types within visual cortex has not yet been spatially resolved in a single experiment, precluding identification of potentially fundamental joint statistics and organizational principles across 3D volumes). Among many examples of the experimental leverage such information could provide, joint 3D cell-typology mapping might be employed to help decode the spatiotemporal logic of neural-activity-triggered gene expression as a function of cell type and spatial location.

We therefore used 5-base barcoded SNAIL probes over six rounds of in situ SEDAL sequencing in coronal mouse brain slices (Figs. 1A and 2, A and B) to survey a large but focused and curated gene set (160 genes including 112 putative cell-

type markers collated from mouse cortical single-cell RNA sequencing (*24*, *25*) and 48 activity-regulated genes (ARGs) (*26*, *27*). In one arm of the experiment, visually-evoked neural activity was provided to a cohort of mice via 1 hour light-exposure after four days of housing in the dark; other mice were kept continuously in the dark (*27*, *28*). 8 μm-thick volumes containing up to 1,000 cells covering all cortical layers were imaged. After six rounds of sequencing, fluorescent Nissl staining was used to segment cell bodies, allowing attribution of amplicons to individual cells (fig. S4, A and B). The values corresponding to amplicons-per-cell and genes-per-cell varied substantially (Fig. 2C), while the 160-gene expression pattern was consistent between biological replicates (R = 0.94-0.95) (Fig. 2D), revealing reliable detection of transcript diversity at the single-cell level. Because only 160 genes were encoded out of the 1024 possible barcodes from five bases, we were able to quantify sequencing errors that resulted in sequences being corrupted from the 160 true barcodes to the 864 invalid barcodes, which was remarkably low at 1-4%. We found that this 160-gene pilot faithfully reproduced the spatial distribution of known cortical layer markers and interneurons, illustrated here via comparison of in situ images from paired public atlases (*29*) and STARmap results (Fig. 2E).

We next performed cell classification using expression data of the 112 cell-type markers. First, >3,000 cells pooled from four biological replicates were clustered into three major cell types (excitatory neurons, inhibitory neurons and non-neuronal cells) using graph-based clustering following principal-component decomposition (*30*), and then further sub-clustered under each category (Fig. 2, F to H, and fig. S4C). Intriguingly, the richly-defined excitatory neurons segregated into four major types, here denoted (eL2/3, eL4, eL5 and eL6) (Fig. 2, I to K, and fig. S5, A and B) by spatial correspondence with anatomic cortical layers and expression profiles of known layer-specific gene markers. Although spatial organization of the four excitatory types exhibited a layered pattern, there was extensive intermixing among different cell types within each layer. Inhibitory neurons were also clustered into four major types, here denoted by the dominant interneuron marker of each subtype [VIP, SST, NPY and PV (Pvalb)] (Fig. 2, L to N, and fig. S5, C and D); the VIP and NPY type were observed to distribute more to the upper layers (L1-3) while SST and PV types were found more commonly in the lower layers (L4-6). We also detected non-neuronal cell types, including astrocytes, oligodendrocytes, endothelial cells, and smooth muscle cells (fig. S6). The number of major cell-types illustrated here (12 in total) can be further broken down (single-cell RNA sequencing can lead to classification into 40 or more subtypes, consistent with the readily-apparent heterogeneity of gene expression within each type) (figs. S5 and S6). Notably, with our targeted 112-gene set and at the size of 600-

800 cells per sample, all 12 major cell types could be reliably detected without batch effects with highly similar spatial patterning among four biological replicates (defined as samples prepared from different animals) (fig. S7), and matched with published single-cell RNA sequencing results (fig. S8)

We next sought to take advantage of STARmap's quantitative capabilities at the single-cell level, to test differential gene expression analyses across experimental conditions, in molecularly-defined cell types. To this end, we assessed visual-stimulus-dependent gene expression patterns (via 48 defined ARGs with single-cell resolution in situ). Further developing the single-cell RNA sequencing procedure, mouse brains here were flash-frozen with minimal handling time after sacrifice (<5 min), for maximal preservation of native transcriptional signatures. Consistent with prior reports (*26–28*), we observed global induction of known immediate-early genes (*Fos*, *Egr1*, and *Egr2*) (Fig. 3A) in primary visual cortex upon 1 hour of light exposure. At single-cell resolution, the quantitative extent (fold change in expression) of ARG changes exhibited striking diversity across neuronal cell types (Fig. 3, B and C, and fig. S9) (*28*). In general, ARG expression programs in excitatory neurons across different layers were highly similar, whereas ARG expression programs in inhibitory cells exhibited much more distinct cell-type specific characteristics (fig. S9C); for example, *Egr2* exhibited light-induction across excitatory neurons (Fig. 3D) but not in inhibitory neurons, while in contrast, *Prok2* was upregulated in *Vip* inhibitory neurons (Fig. 3C) (*22*). Finally, since neural activity can trigger co-transcription of noncoding RNAs from within enhancers of ARGs (*26*, *31*), we also studied exemplars of these enhancer RNAs (here, eRNAs 1-5 of the *Fos* gene); these transcripts, not polyadenylated, would be very difficult to measure with current single-cell RNA sequencing. Intriguingly, eRNA3 was identified as the most significant and consistent ARG marker (fig. S9B).

Comparing spatial cell-type distributions in frontal and sensory cortices

We then investigated to what extent the cell types of higher cognitive cortex resemble those of sensory cortex, as exemplified by primary visual cortex. We applied the same 160-gene set to STARmapping the cell populations of medial prefrontal cortex (mPFC) (Fig. 4A) which is involved in high-level cognitive functions such as attention and memory, and is thought to be dysregulated in major psychiatric disorders (*32*). We identified 15 distinct molecular cell types including six excitatory neuron subtypes (eL2/3, eL5-1, eL5-2, eL5-3, eL6-1 and eL6-2, annotated by anatomic cortical layers), five inhibitory neuron subtypes (VIP, Reln, SST, Lhx6 and NPY, annotated by dominant gene markers) and four non-neuronal types (astrocytes, oligodendrocytes, endothelial cells and smooth-muscle cells) (Fig. 4B and fig. S10).

The spatial organization of broad cell types in mPFC resembled that of V1 with intermixed excitatory neuronal layers and sparsely distributed inhibitory neurons (Fig. 4C), however, the nature and composition of neuronal subtypes in mPFC and V1 strikingly differed (Fig. 4, D and E). For excitatory subtypes, mPFC lacks eL4 (consistent with previous reports) (*33*), and exhibits reduced eL2/3 and vast expansion of eL5 and eL6 compared to V1 (Fig. 4E). Many new types of cell were discovered, including three eL5 subtypes and two eL6 subtypes, as characterized by gene markers *Sema3e*, *Plcxd2*, *Tpbg*, *Syt6*, and *Ctgf*, respectively (Fig. 4D).

Substantially different tissue organization by cell type was also observed for inhibitory subclusters. *Sst-*, *Vip-* and *Npy-*positive subtypes in mPFC were represented similarly among all inhibitory neurons compared with those in V1, while *Pvalb*-positive cells were comparatively much sparser. In V1 *Reln*-positive neurons coexist with *Sst* and *Npy*, while in mPFC these segregate as a single cluster with ~50% co-marked by *Ndnf*; we also discovered a new inhibitory subtype labeled by *Lhx6* which in fact constitutes the most abundant inhibitory subtype in mPFC (Fig. 4E). While the 5-HT(3A) receptor (*Htr3a*) expression has been reported in cortical inhibitory neurons (*34*), *Htr3a* has not been ranked as a critical genetic marker of inhibitory subtypes in V1. In mPFC, however, we find that *Htr3a* distinguishably marks a large fraction of *Vip*$^+$ neurons and a subset of *Reln*$^+$ neurons (fig. S10D).

Superficial layers (L1-3) were found to contain *Vip*, *Reln* and *Npy* subtypes while deeper layers (L5-6) were found to contain all of the inhibitory subtypes. Indeed, all of the 15 cell types with tissue-level spatial organization could be reliably detected by STARmap across four biological replicates (fig. S11). The capability of STARmap for multidimensional cell typing in mPFC was further demonstrated in the setting of activity-dependence, supporting the possibility of defining cell types in part by communication properties including activity during behavior (*35*, *36*). 1 hour after cocaine injection (*37*), a specific subpopulation of deep-layer excitatory neurons (e.g., *Tpbg* labeled eL5-2) in mPFC was activated compared to saline-injected control mice (Fig. 4F), revealing STARmap capability for identifying functional segregation of neuronal subtypes in mPFC.

Scaling STARmap to >1,000 genes

To further test the scalability of STARmap, we extended our gene list from 160 to 1,020 genes, leveraging previously published single-cell RNA sequencing data (*24*). The 1,020-gene set was first validated in mouse hippocampal neuron culture, with successful resolution of neuronal and glial cells (fig. S12). We then probed mouse V1 neocortex with the 1,020-gene set to evaluate performance in spatial cell typing in comparison with the 160-gene set. Amplicons obtained in 1,020-gene experiment were much denser in cells compared to

those in 160-gene experiments but were optically-resolvable in 3D with high-resolution imaging and post-imaging deconvolution (Fig. 5A).

We observed that a higher percentage (40%) of amplicons were filtered out in the 1020-gene experiments by our error-rejection mechanism (fig. S3H) in comparison to the 4-gene experiments (20%) (fig. S3L), indicating a more frequent initial color-misassignment potentially resulted from amplicon merging or optical resolution, and further demonstrating the importance of our designed error-rejection mechanism. Crucially though despite the read loss, we successfully clustered single cells of the imaging area into 16 annotated cell types using 1,020 genes and the same data analysis pipeline from the focused 160 gene probe set (Fig. 5, B and C, and fig. S13). Three new cell types were identified in addition to the 13 cell types detected by 160 genes (Fig. 5B): eL6 was resolved into two subtypes (eL6-1 and eL6-2); a novel hippocampal excitatory subtype (HPC) was identified; and microglial cells were cleanly identified with an expansion of non-neuronal cell type markers in the 1,020-gene set.

Beyond those advances, the 1,020-gene findings also successfully reproduced the cell types (and their spatial patterning) from the 160-gene findings, and further allowed discovery of multiple new gene markers for each cell type (e.g., *3110035E14Rik* for deep layers, *Cnot6l* for *Sst* neurons, and *Cplx1* for *Pvalb* neurons) (Fig. 5D and fig. S13). These molecularly-defined cell types were highly reproducible between biological replicates for 1,020-gene detection and were concordant with published single-cell RNA sequencing results (fig. S14). We further assessed the possibility of scaling up STARmap to accommodate higher gene numbers; while the STARmap scheme can encode and decode more than one million codes and the physical volume of mammalian cells is not limiting for amplification of more than 1,000 genes (fig. S15), the 1,020-gene experiments approached the upper limit of the optical volume of cells (fig. S15E); for those cases where more genes are needed, STARmap may cover the whole transcriptome with optical resolution enhanced by super-resolution microscopy (*38*, *39*) or the physical swelling typical of the hydrogel-tissue chemistries (*14*, *19*).

Adapting STARmap to thick tissue blocks for 3D analyses

In neuroscience, addressing the 3D complexity of both neurons and neural circuits has generally required development and use of thick tissue blocks or fully-intact brains for functional and structural readouts, including electrophysiology, imaging of activity, and analysis of morphology and connectivity. Therefore, for linking these readout measures from intact or semi-intact tissue preparations with cellular-resolution gene expression readouts from the very same preparations, methods of 3D spatial transcriptomic analysis in thick

tissues have been long-sought, to achieve datastream registration as well as preserve 3D morphology and to obtain readouts from very much larger cell numbers (2). The initial experiments were carried out in brain slices no more than 1 cell body thick; we therefore next developed and tested STARmap to overcome limitations in diffusional access and imaging throughput for large tissue volumes, with a modified strategy for linearly reading out gene expression at cellular resolution to enable high-throughput molecular analysis in tissue volumes (Fig. 6A and fig. S16). Specificity and penetration depth of large-volume STARmap were tested initially using Thy1::YFP mouse brains, wherein STARmap successfully detected YFP mRNA across 150 μm of tissue thickness, and specifically co-localized YFP protein and mRNA at single-cell resolution (Fig. 6B) without labeling the tens of thousands of interspersed neighboring cells.

We then extended the spatial cell-typing of mouse primary visual cortex to more than 30,000 cells across volumes spanning all six layers and the corpus callosum. Using a curated gene set including 23 cell type markers and 5 ARGs read out over 7 cycles of linear SEDAL sequencing (Fig. 6, C and D, and fig. S17), we applied K-means clustering of marker genes (Methods) for each cell-type (recovering 11 cell types corresponding to the majority of those extracted by the 160-gene experiment). We found that 3D patterning of the 11 cell types (Fig. 6, E and F) was consistent with the 160-gene thin-section tissue findings, but provided a novel accurate and quantitative profiling of cellular distribution across space, with much larger cell numbers. As reflected by both spatial-histogram (Fig. 6E) and correlational analyses (fig. S17B), excitatory subtypes exhibited an unanticipated layered gradient distribution, with the spatial density of each subtype decaying across space into adjacent layers. In contrast, inhibitory subtypes were dispersed, albeit with layer preferences exhibited by the *Vip* subtype (largely located in layer 2/3), and the *Sst* and *Pvalb* subtypes (in layers 4 and 5). Non-neuronal cells were largely seen in layer 1 and white matter.

To discover yet-finer volumetric patterns, we further analyzed the distribution of distances from each individual cell from each sequencing-defined subtype to its nearest-neighbors, finding unexpectedly that the nearest neighbor of any inhibitory neuron tended to be its own subtype, rather than excitatory neurons or others inhibitory subtypes (Fig. 6G). If inhibitory neurons were randomly dispersed among the more abundant excitatory neurons in a purely salt-and-pepper distribution, the distance between inhibitory neurons would be larger than that from inhibitory to excitatory neurons (Fig. 6H). Remarkably instead, the actual intra-subtype distance of inhibitory neurons was much shorter (~15 μm, equivalent to the size of a single neuron, indicating direct somatic juxtaposition) (Fig. 6I), revealing a short-range self-clustering organization of inhibitory subtypes across volumes that could only

be accurately measured in 3D, but not in 2D (fig. S18A). When guided by this initial STARmap observation, evidence for such patterning could be also obtained in transgenic mouse lines (fig. S18, B and C). This discovery bears considerable relevance to previous functional work; for example, electrophysiological studies have suggested that inhibitory neurons in spatial proximity tend to be connected by electric (gap) junctions important for setting up synchronized firing patterns (40, 41), and in vivo imaging has suggested that inhibitory-neuron groupings in visual cortex could sharpen visual responses (42).

Discussion

STARmap defines a platform for 3D in situ transcriptomics, enabled by state-of-the-art DNA library preparation/sequencing and novel hydrogel-tissue chemistry. Here, STARmap was shown to be applicable to the study of molecularly-defined cell types and activity-regulated gene expression in mouse cortex, and to be scalable to larger 3D tissue blocks to visualize short- and long- range spatial organization of cortical neurons on a volumetric scale not previously accessible. In future work, STARmap may also be adapted to longer sequencing lengths or higher gene numbers; there is no intrinsic limit to the number of genes or RNA species that can be simultaneously and quantitatively accessed by STARmap (fig. S15); STARmap may also be capable of integrating cell type information with single-neuron morphology and projection anatomy (e.g., via Brainbow and MAPseq) (43, 44) as well as with in vivo neural activity imaging and electrophysiology. This platform can also be generalized to study other heterogeneous cell populations in diverse tissues across the body, though the brain poses special challenges well suited to STARmap analysis. For example, the polymorphic ARG expression observed across different cell types is likely to depend on both intrinsic cell-biological properties (such as signal transduction pathway-component expression), and on extrinsic properties such as neural circuit anatomy that routes external sensory information to different cells (here in visual cortex). In general, it may not be possible to fully define brain cell typology independent of such 3D anatomy as well as activity patterns exhibited and experienced by cells during behavior; the nature of input and output communication pathways for the cells in question in fact can form the foundation for defining cell types (35, 36). Toward this end, in situ transcriptomics exemplified by STARmap can effectively link this imaging-based molecular information with complementary cellular-resolution datastreams describing anatomy, natural activity, and causal significance, thus promising to fundamentally deepen our understanding of brain function and dysfunction (2).

Methods summary

All animal procedures followed animal care guidelines approved by Stanford University's Administrative Panel on Laboratory Animal Care (APLAC) and guidelines of the National Institutes of Health. For thin sections, animals were anesthetized and rapidly decapitated; the brain tissues were sliced using a cryostat. For thick sections, animals were anesthetized and transcardially perfused with PFA; the brain tissues were sliced using a vibratome. In STARmap experiments, tissues were hybridized with SNAIL probes, enzymatically amplified, hydrogel embedded, and sequentially imaged using SEDAL process and a confocal microscope. The resulting image datasets were registered across multiple cycles using the positions of all amplicons in each cycle and decoded. For cell-typing and single-cell gene expression analyses, the amplicons were attributed to individual cells based on segmentation images of fluorescent Nissl staining. All the detailed procedures for the experiments and data analyses are described in the supplementary materials.

REFERENCES AND NOTES

1. N. Crosetto, M. Bienko, A. van Oudenaarden, Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015). doi:10.1038/nrg3832 Medline
2. E. Lein, L. E. Borm, S. Linnarsson, The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358**, 64–69 (2017). doi:10.1126/science.aan6827 Medline
3. E. Lubeck, L. Cai, Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nat. Methods* **9**, 743–748 (2012). doi:10.1038/nmeth.2069 Medline
4. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015). doi:10.1126/science.aaa6090 Medline
5. S. Shah, E. Lubeck, W. Zhou, L. Cai, *In situ* transcription profiling of single cells reveals spatial organization of cells in the mouse hippocampus. *Neuron* **92**, 342–357 (2016). doi:10.1016/j.neuron.2016.10.001 Medline
6. R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby, M. Nilsson, *In situ* sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013). doi:10.1038/nmeth.2563 Medline
7. J. H. Lee, E. R. Daugharthy, J. Scheiman, R. Kalhor, J. L. Yang, T. C. Ferrante, R. Terry, S. S. F. Jeanty, C. Li, R. Amamoto, D. T. Peters, B. M. Turczyk, A. H. Marblestone, S. A. Inverso, A. Bernard, P. Mali, X. Rios, J. Aach, G. M. Church, Highly multiplexed subcellular RNA sequencing *in situ*. *Science* **343**, 1360–1363 (2014). doi:10.1126/science.1250212 Medline
8. A. P. Frei, F.-A. Bava, E. R. Zunder, E. W. Y. Hsieh, S.-Y. Chen, G. P. Nolan, P. F. Gherardini, Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat. Methods* **13**, 269–275 (2016). doi:10.1038/nmeth.3742 Medline
9. X. Chen, Y.-C. Sun, G. M. Church, J. H. Lee, A. M. Zador, Efficient *in situ* barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res.* **46**, e22 (2018). doi:10.1093/nar/gkx1206 Medline
10. M. Nagendran, D. P. Riordan, P. B. Harbury, T. J. Desai, Automated cell-type classification in intact tissues by single-cell molecular profiling. *eLife* **7**, e30510 (2018). doi:10.7554/eLife.30510 Medline
11. N. A. Peppas, J. Z. Hilt, A. Khademhosseini, R. Langer, Hydrogels in biology and medicine: From molecular principles to bionanotechnology. *Adv. Mater.* **18**, 1345–1360 (2006). doi:10.1002/adma.200501612
12. A. M. Rosales, K. S. Anseth, The design of reversible hydrogels to capture extracellular matrix dynamics. *Nat. Rev. Mater.* **1**, 1–15 (2016). Medline
13. R. Y. Tam, L. J. Smith, M. S. Shoichet, Engineering cellular microenvironments with photo- and enzymatically responsive hydrogels: Toward biomimetic 3D cell culture models. *Acc. Chem. Res.* **50**, 703–713 (2017). doi:10.1021/acs.accounts.6b00543 Medline
14. V. Gradinaru, J. Treweek, K. Overton, K. Deisseroth, Hydrogel-tissue chemistry: Principles and applications. *Annu. Rev. Biophys.* **47**, 355–376 (2018). doi:10.1146/annurev-biophys-070317-032905 Medline
15. K. Chung, J. Wallace, S.-Y. Kim, S. Kalyanasundaram, A. S. Andalman, T. J. Davidson, J. J. Mirzabekov, K. A. Zalocusky, J. Mattis, A. K. Denisin, S. Pak, H. Bernstein, C. Ramakrishnan, L. Grosenick, V. Gradinaru, K. Deisseroth, Structural and molecular interrogation of intact biological systems. *Nature* **497**, 332–337 (2013). doi:10.1038/nature12107 Medline
16. E. L. Sylwestrak, P. Rajasethupathy, M. A. Wright, A. Jaffe, K. Deisseroth, Multiplexed intact-tissue transcriptional analysis at cellular resolution. *Cell* **164**, 792–804 (2016). doi:10.1016/j.cell.2016.01.038 Medline
17. S. Shah, E. Lubeck, M. Schwarzkopf, T.-F. He, A. Greenbaum, C. H. Sohn, A. Lignell, H. M. T. Choi, V. Gradinaru, N. A. Pierce, L. Cai, Single-molecule RNA detection at depth by hybridization chain reaction and tissue hydrogel embedding and clearing. *Development* **143**, 2862–2867 (2016). doi:10.1242/dev.138560 Medline
18. J. R. Moffitt, J. Hao, D. Bambah-Mukku, T. Lu, C. Dulac, X. Zhuang, 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). doi:10.1073/pnas.1617699113 Medline
19. F. Chen, A. T. Wassie, A. J. Cote, A. Sinha, S. Alon, S. Asano, E. R. Daugharthy, J.-B. Chang, A. Marblestone, G. M. Church, A. Raj, E. S. Boyden, Nanoscale imaging of RNA with expansion microscopy. *Nat. Methods* **13**, 679–684 (2016). doi:10.1038/nmeth.3899 Medline
20. J. Shendure, G. J. Porreca, N. B. Reppas, X. Lin, J. P. McCutcheon, A. M. Rosenbaum, M. D. Wang, K. Zhang, R. D. Mitra, G. M. Church, Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005). doi:10.1126/science.1117389 Medline
21. J. H. Lee, E. R. Daugharthy, J. Scheiman, R. Kalhor, T. C. Ferrante, R. Terry, B. M. Turczyk, J. L. Yang, H. S. Lee, J. Aach, K. Zhang, G. M. Church, Fluorescent *in situ* sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015). doi:10.1038/nprot.2014.191 Medline
22. R. Drmanac, A. B. Sparks, M. J. Callow, A. L. Halpern, N. L. Burns, B. G. Kermani, P. Carnevali, I. Nazarenko, G. B. Nilsen, G. Yeung, F. Dahl, A. Fernandez, B. Staker, K. P. Pant, J. Baccash, A. P. Borcherding, A. Brownley, R. Cedeno, L. Chen, D. Chernikoff, A. Cheung, R. Chirita, B. Curson, J. C. Ebert, C. R. Hacker, R. Hartlage, B. Hauser, S. Huang, Y. Jiang, V. Karpinchyk, M. Koenig, C. Kong, T. Landers, C. Le, J. Liu, C. E. McBride, M. Morenzoni, R. E. Morey, K. Mutch, H. Perazich, K. Perry, B. A. Peters, J. Peterson, C. L. Pethiyagoda, K. Pothuraju, C. Richter, A. M. Rosenbaum, S. Roy, J. Shafto, U. Sharanhovich, K. W. Shannon, C. G. Sheppy, M. Sun, J. V. Thakuria, A. Tran, D. Vu, A. W. Zaranek, X. Wu, S. Drmanac, A. R. Oliphant, W. C. Banyai, B. Martin, D. G. Ballinger, G. M. Church, C. A. Reid, Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010). doi:10.1126/science.1181498 Medline
23. L. L. Glickfeld, R. C. Reid, M. L. Andermann, A mouse model of higher visual cortical function. *Curr. Opin. Neurobiol.* **24**, 28–33 (2014). doi:10.1016/j.conb.2013.08.009 Medline
24. B. Tasic, V. Menon, T. N. Nguyen, T. K. Kim, T. Jarsky, Z. Yao, B. Levi, L. T. Gray, S. A. Sorensen, T. Dolbeare, D. Bertagnolli, J. Goldy, N. Shapovalova, S. Parry, C. Lee, K. Smith, A. Bernard, L. Madisen, S. M. Sunkin, M. Hawrylycz, C. Koch, H. Zeng, Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016). doi:10.1038/nn.4216 Medline
25. A. Zeisel, A. B. Muñoz-Manchado, S. Codeluppi, P. Lönnerberg, G. La Manno, A. Juréus, S. Marques, H. Munguba, L. He, C. Betsholtz, C. Rolny, G. Castelo-Branco, J. Hjerling-Leffler, S. Linnarsson, Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015). doi:10.1126/science.aaa1934 Medline
26. T. K. Kim, M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman, M. E. Greenberg, Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010). doi:10.1038/nature09033 Medline
27. A. R. Mardinly, I. Spiegel, A. Patrizi, E. Centofante, J. E. Bazinet, C. P. Tzeng, C. Mandel-Brehm, D. A. Harmin, H. Adesnik, M. Fagiolini, M. E. Greenberg, Sensory experience regulates cortical inhibition by inducing IGF1 in VIP neurons. *Nature* **531**, 371–375 (2016). doi:10.1038/nature17187 Medline

28. S. Hrvatin, D. R. Hochbaum, M. A. Nagy, M. Cicconet, K. Robertson, L. Cheadle, R. Zilionis, A. Ratner, R. Borges-Monroy, A. M. Klein, B. L. Sabatini, M. E. Greenberg, Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat. Neurosci.* **21**, 120–129 (2018). doi:10.1038/s41593-017-0029-5 Medline

29. E. S. Lein, M. J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, A. F. Boe, M. S. Boguski, K. S. Brockway, E. J. Byrnes, L. Chen, L. Chen, T.-M. Chen, M. C. Chin, J. Chong, B. E. Crook, A. Czaplinska, C. N. Dang, S. Datta, N. R. Dee, A. L. Desaki, T. Desta, E. Diep, T. A. Dolbeare, M. J. Donelan, H.-W. Dong, J. G. Dougherty, B. J. Duncan, A. J. Ebbert, G. Eichele, L. K. Estin, C. Faber, B. A. Facer, R. Fields, S. R. Fischer, T. P. Fliss, C. Frensley, S. N. Gates, K. J. Glattfelder, K. R. Halverson, M. R. Hart, J. G. Hohmann, M. P. Howell, D. P. Jeung, R. A. Johnson, P. T. Karr, R. Kawal, J. M. Kidney, R. H. Knapik, C. L. Kuan, J. H. Lake, A. R. Laramee, K. D. Larsen, C. Lau, T. A. Lemon, A. J. Liang, Y. Liu, L. T. Luong, J. Michaels, J. J. Morgan, R. J. Morgan, M. T. Mortrud, N. F. Mosqueda, L. L. Ng, R. Ng, G. J. Orta, C. C. Overly, T. H. Pak, S. E. Parry, S. D. Pathak, O. C. Pearson, R. B. Puchalski, Z. L. Riley, H. R. Rockett, S. A. Rowland, J. J. Royall, M. J. Ruiz, N. R. Sarno, K. Schaffnit, N. V. Shapovalova, T. Sivisay, C. R. Slaughterbeck, S. C. Smith, K. A. Smith, B. I. Smith, A. J. Sodt, N. N. Stewart, K.-R. Stumpf, S. M. Sunkin, M. Sutram, A. Tam, C. D. Teemer, C. Thaller, C. L. Thompson, L. R. Varnam, A. Visel, R. M. Whitlock, P. E. Wohnoutka, C. K. Wolkey, V. Y. Wong, M. Wood, M. B. Yaylaoglu, R. C. Young, B. L. Youngstrom, X. F. Yuan, B. Zhang, T. A. Zwingman, A. R. Jones, Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007). doi:10.1038/nature05453 Medline

30. K. Shekhar, S. W. Lapan, I. E. Whitney, N. M. Tran, E. Z. Macosko, M. Kowalczyk, X. Adiconis, J. Z. Levin, J. Nemesh, M. Goldman, S. A. McCarroll, C. L. Cepko, A. Regev, J. R. Sanes, Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* **166**, 1308–1323.e30 (2016). doi:10.1016/j.cell.2016.07.054 Medline

31. J. Y. Joo, K. Schaukowitch, L. Farbiak, G. Kilaru, T.-K. Kim, Stimulus-specific combinatorial functionality of neuronal *c-fos* enhancers. *Nat. Neurosci.* **19**, 75–83 (2016). doi:10.1038/nn.4170 Medline

32. M. Carlén, What constitutes the prefrontal cortex? *Science* **358**, 478–482 (2017). doi:10.1126/science.aan8868 Medline

33. H. B. Uylings, H. J. Groenewegen, B. Kolb, Do rats have a prefrontal cortex? *Behav. Brain Res.* **146**, 3–17 (2003). doi:10.1016/j.bbr.2003.09.028 Medline

34. S. Lee, J. Hjerling-Leffler, E. Zagha, G. Fishell, B. Rudy, The largest group of superficial neocortical GABAergic interneurons expresses ionotropic serotonin receptors. *J. Neurosci.* **30**, 16796–16808 (2010). doi:10.1523/JNEUROSCI.1869-10.2010 Medline

35. A. Paul, M. Crow, R. Raudales, M. He, J. Gillis, Z. J. Huang, Transcriptional architecture of synaptic communication delineates GABAergic neuron identity. *Cell* **171**, 522–539.e20 (2017). doi:10.1016/j.cell.2017.08.032 Medline

36. T. N. Lerner, L. Ye, K. Deisseroth, Communication in neural circuits: Tools, opportunities, and challenges. *Cell* **164**, 1136–1150 (2016). doi:10.1016/j.cell.2016.02.027 Medline

37. L. Ye, W. E. Allen, K. R. Thompson, Q. Tian, B. Hsueh, C. Ramakrishnan, A.-C. Wang, J. H. Jennings, A. Adhikari, C. H. Halpern, I. B. Witten, A. L. Barth, L. Luo, J. A. McNab, K. Deisseroth, Wiring and molecular features of prefrontal ensembles representing distinct experience. *Cell* **165**, 1776–1788 (2016). doi:10.1016/j.cell.2016.05.010 Medline

38. M. J. Rust, M. Bates, X. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006). doi:10.1038/nmeth929 Medline

39. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, H. F. Hess, Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006). doi:10.1126/science.1127344 Medline

40. J. R. Gibson, M. Beierlein, B. W. Connors, Two networks of electrically coupled inhibitory neurons in neocortex. *Nature* **402**, 75–79 (1999). doi:10.1038/47035 Medline

41. Y. Amitai, J. R. Gibson, M. Beierlein, S. L. Patrick, A. M. Ho, B. W. Connors, D. Golomb, The spatial dimensions of electrically coupled networks of interneurons in the neocortex. *J. Neurosci.* **22**, 4142–4152 (2002). doi:10.1523/JNEUROSCI.22-10-04142.2002 Medline

42. T. Ebina, K. Sohya, I. Imayoshi, S.-T. Yin, R. Kimura, Y. Yanagawa, H. Kameda, H. Hioki, T. Kaneko, T. Tsumoto, 3D clustering of GABAergic neurons enhances inhibitory actions on excitatory neurons in the mouse visual cortex. *Cell Reports* **9**, 1896–1907 (2014). doi:10.1016/j.celrep.2014.10.057 Medline

43. J. Livet, T. A. Weissman, H. Kang, R. W. Draft, J. Lu, R. A. Bennis, J. R. Sanes, J. W. Lichtman, Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007). doi:10.1038/nature06293 Medline

44. J. M. Kebschull, P. Garcia da Silva, A. P. Reid, I. D. Peikon, D. F. Albeanu, A. M. Zador, High-throughput mapping of single-neuron projections by sequencing of barcoded RNA. *Neuron* **91**, 975–987 (2016). doi:10.1016/j.neuron.2016.07.036 Medline

45. A. McDavid, G. Finak, P. K. Chattopadyay, M. Dominguez, L. Lamoreaux, S. S. Ma, M. Roederer, R. Gottardo, Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* **29**, 461–467 (2013). doi:10.1093/bioinformatics/bts714 Medline

46. J. L. Bentley, Multidimensional binary search trees used for associative searching. *Commun. ACM* **18**, 509–517 (1975). doi:10.1145/361002.361007

47. C. Larsson, I. Grundberg, O. Söderberg, M. Nilsson, In situ detection and genotyping of individual mRNA molecules. *Nat. Methods* **7**, 395–397 (2010). doi:10.1038/nmeth.1448 Medline

48. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015). doi:10.1016/j.cell.2015.05.002 Medline

49. P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, J. Frisén, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016). doi:10.1126/science.aaf2403 Medline

ACKNOWLEDGMENTS

staining with other segmentation methods. W.E.A. and E.L.S. designed the visual-stimulus and cocaine-stimulus procedures. S.V. generated the CLARITY data with PV transgenicmice. K. E. and C. R. contributed to preparation of cell cultures. C. L. assisted with experiments. J.L. assisted with experiments and graphic design. K.D. supervised all aspects of the work. X.W., W.E.A., and K.D. interpreted the STARmap data, designed and prepared the figures, and wrote the manuscript with edits from all authors. **Competing interests:** The design, steps and applications of STARmap are covered in pending patent application material from Stanford University; all methods, protocols, and sequences are freely available to nonprofit institutions and investigators. **Data and materials availability:** All data are available in the main text or the supplementary materials. Computational tools/code and other materials are available at http://clarityresourcecenter.org.

SUPPLEMENTARY MATERIALS
www.sciencemag.org/cgi/content/full/science.aat5691/DC1
Materials and Methods
Figs. S1 to S18
References (*45–49*)
Tables S1 and S2

A Solid heterogeneous tissue — cell — SNAIL probe — gene-unique identifier — DNA amplicon

Sequencing hydrogel — 3D cDNA library — DNA-gel

Spatial cell typing — 3D imaging — Readout

TTGCA Gene 1
ACGTT Gene 2
....
....
GTACT Gene n

B primer · padlock → AND → amplification

C primer_A · padlock_A · mRNA_A · mRNA_B

D padlock_A · mRNA_A · primer_B · mRNA_B

E Cycle 1 · Cycle 2 · Cycle 3 · Cycle 4 · Cycle 5 · Cycle 6

Ch1 Ch2 Ch3 Ch4
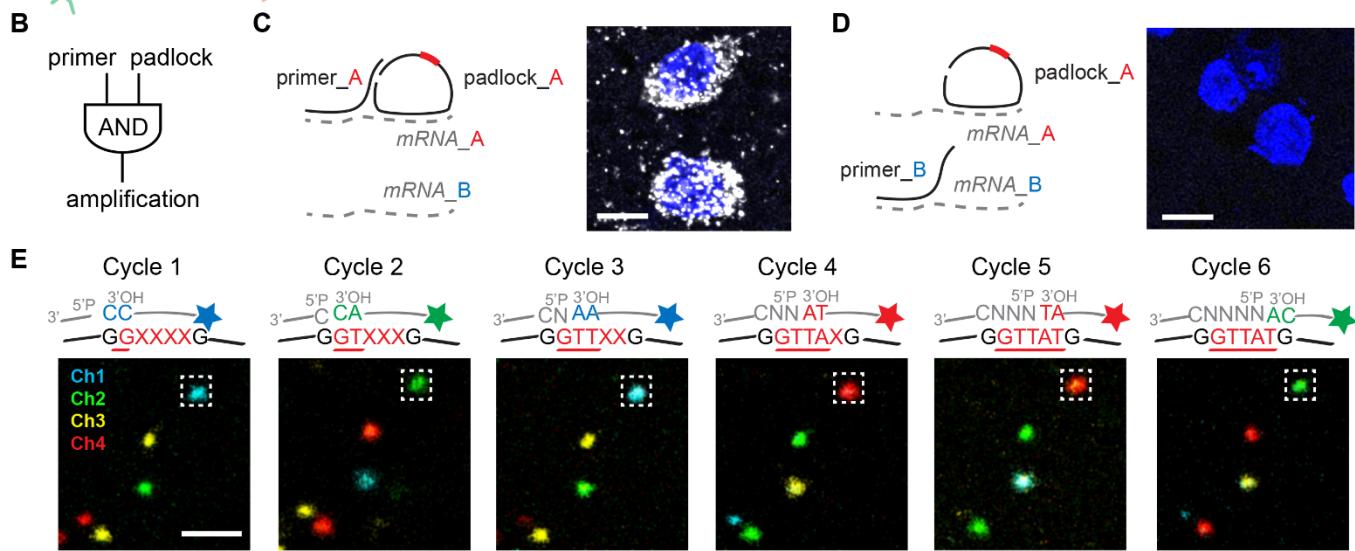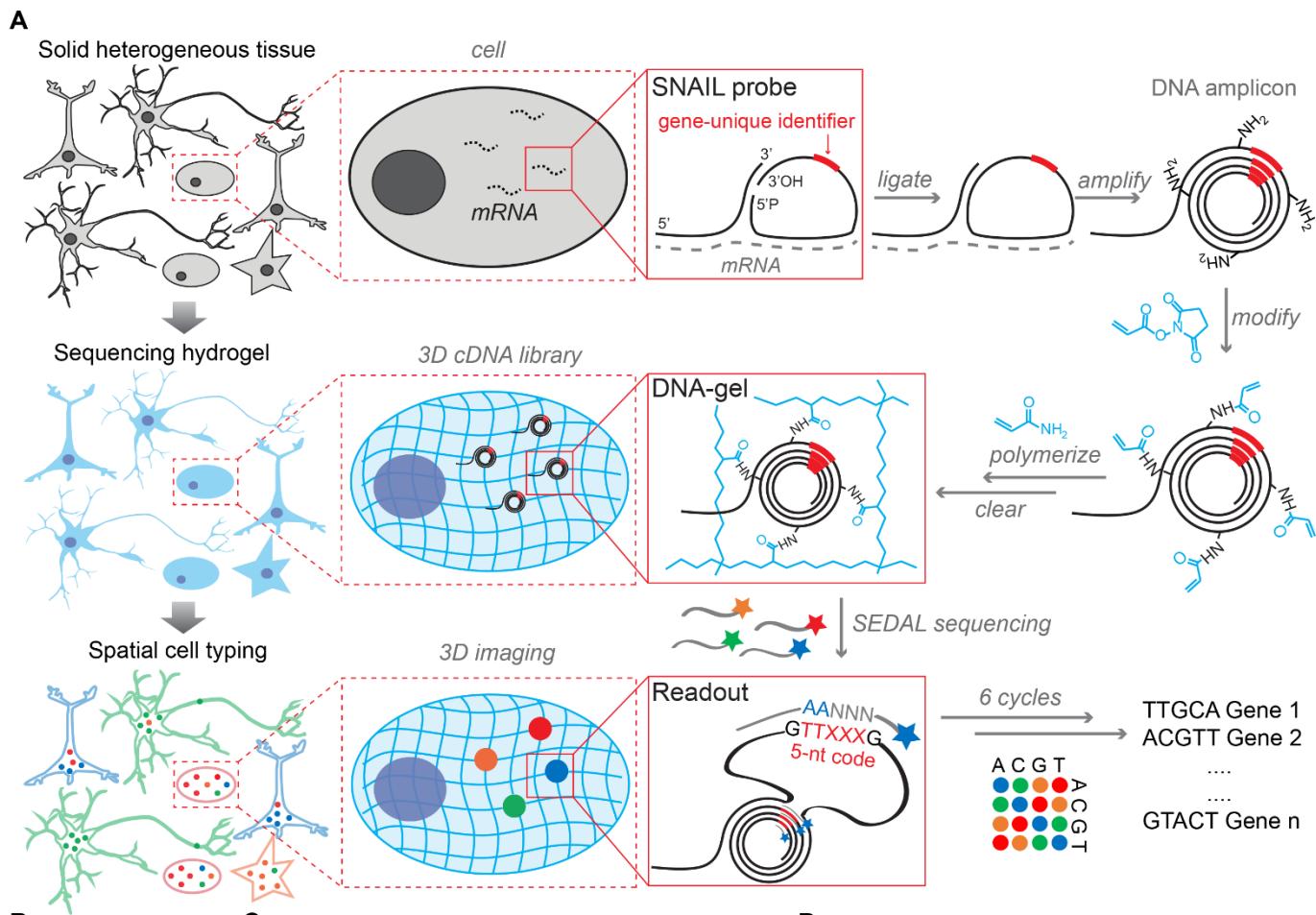
Fig. 1. STARmap principles: in situ RNA sequencing for spatial transcriptomics within the 3D tissue environment. (**A**) STARmap overview schematic. After brain tissue is prepared (see Methods for mouse brain protocols), the custom SNAIL probes that encounter and hybridize to intracellular mRNAs (dashed lines) within the intact tissue are enzymatically replicated as cDNA amplicons. The amplicons are constructed in situ with an acrylic acid N-hydroxysuccinimide moiety modification (blue) and then copolymerized with acrylamide to embed within a hydrogel network (blue wavy lines), following by clearance of unbound lipids and proteins (fig. S2). Each SNAIL probe contains a gene-unique identifier segment (red) which is read-out through in situ sequencing with 2-base encoding for error correction (SEDAL) (fig. S3). Finally, highly multiplexed RNA quantification in 3D reveals gene expression and cell types in space. (**B**) SNAIL logic: a pair of primer and padlock probes amplifies target-specific signals and excludes noise known to commonly arise from non-specific hybridization of a single probe. (**C** and **D**) Only adjacent binding of primer and padlock probes leads to signal amplification. mRNA A represents *Gapdh* and mRNA B represents *Actb*. Both fluorescent images showing *Gapdh* (gray) mRNA and cell nuclei (blue) labeling in mouse brain slice; note the absence of labeling with mismatched primer and padlock (right). Scale bar, 10 μm. (**E**) In situ sequencing of DNA amplicons in the tissue-hydrogel complex via SEDAL, the novel sequencing-by-ligation method devised for STARmap: for each cycle, the reading probes (gray line without star-symbol label) contain an incrementally increasing-length run of degenerate bases (N representing an equal mixture of A, T, C and G) with phosphate at the 5′ end (5′P) to set the reading position; the decoding probes (gray line with star-symbol label) are labeled by fluorophores with color coding for the dinucleotide at the 3′ end. Only if both probes are perfectly complementary to the DNA template (black lower sequence), the two kinds of probes can then be ligated) to form a stable product with high melting-temperature, allowing later imaging after unligated probes are washed away. After each imaging cycle, probes are stripped away from the robust tissue-hydrogel using 60% formamide so that the next cycle can begin. X: unknown base to be read; red underline: decoded sequence; Ch1-4: fluorescence channels. Scale bar, 2 μm.
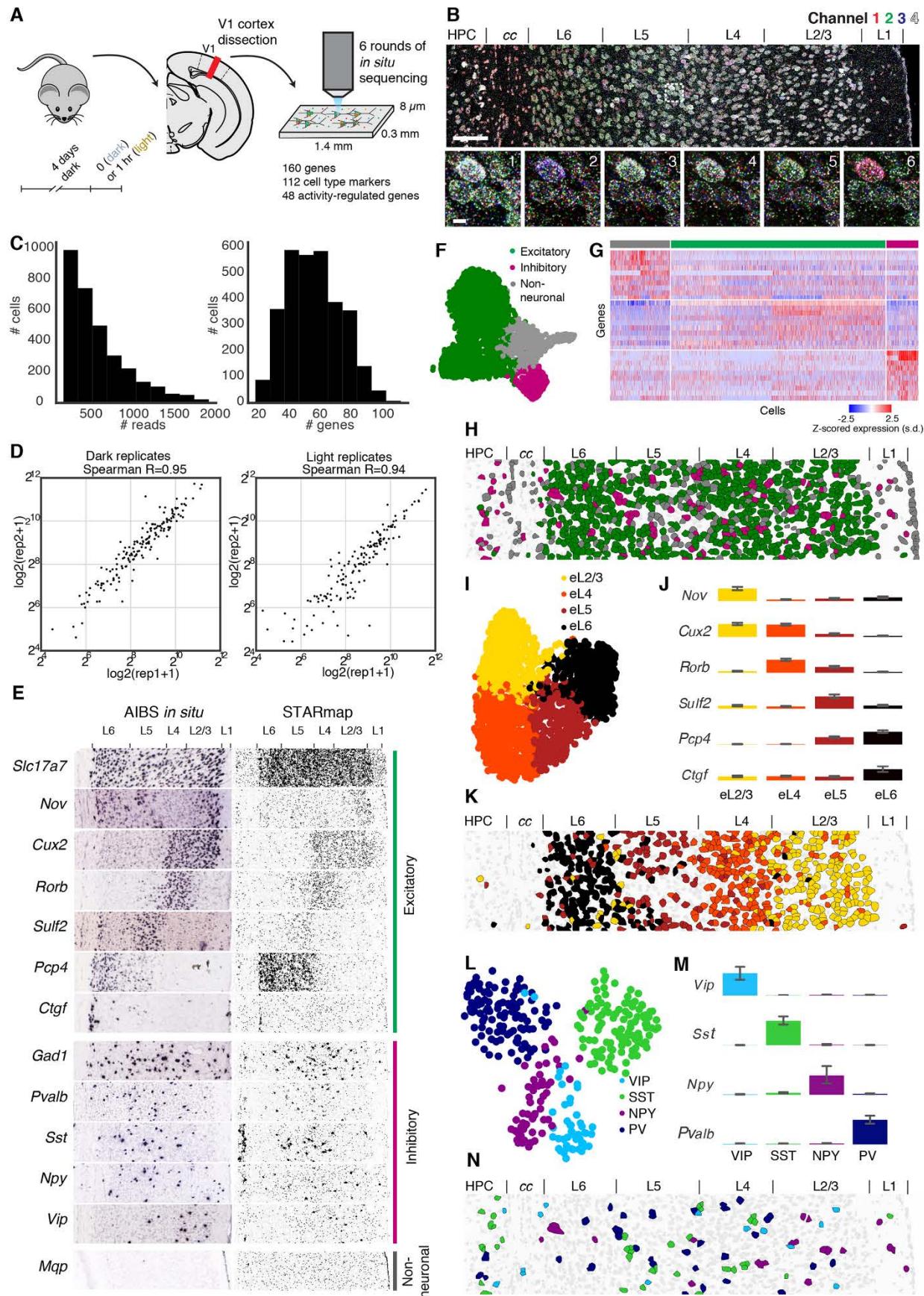
Fig. 2. STARmapping cell types in primary visual cortex. (**A**) Experimental design; mice were dark housed, prior to sacrifice, for 4 days and then either kept in the dark or exposed to light for 1 hour. Primary visual cortex (V1) was coronally sectioned and RNAs of 112 cell type markers and 48 activity-regulated genes were quantified by STARmap. (**B**) Raw fluorescence images of in-process STARmap with the full view of Cycle 1 (top) and zoomed views across all six cycles (bottom). Full field: 1.4 mm × 0.3 mm, scale bar, 100 μm; zoomed region: 11.78 μm × 11.78 μm, scale bar, 2 μm; Channel: color code for the four fluorescence channels; L1-6: the six neocortical layers; *cc*: corpus callosum; HPC: hippocampus. (**C**) Histograms: detected reads (DNA amplicons) per cell (left), and genes per cell (right). (**D**) Quantitative reproducibility of biological replicates, whether in the light or dark condition: $\log_2$(amplicon quantity) for 160 genes across the whole imaging region plotted. Rep1: expression value in first replicate, rep2: expression value in second replicate. (**E**) Validation of STARmap: left column, in situ images from Allen Institute of Brain Science (AIBS); right column, RNA pattern of individual genes extracted from 160-gene STARmap, which reliably reproduced the spatial gene expression pattern from AIBS. (**F**) Uniform Manifold Approximation (UMAP) plot, a non-linear dimensionality reduction technique used to visualize the similarity of cell transcriptomes in two dimensions, showing consistent clustering of major cell types across 3,142 cells pooled from four biological replicates: 2,199 excitatory neurons, 324 inhibitory neurons and 619 non-neuronal cells. (**G**) Gene expression heatmap for 112 cell-type markers aligned with each cell cluster, showing clustering by inhibitory, excitatory, or non-neuronal cell types. Expression for each gene is z-scored across all genes in each cell. (**H**) Representative cell-resolved spatial map in neocortex and beyond: cell-types color-coded as in panel (F). (**I** to **N**) Clustering of excitatory and inhibitory subtypes: UMAP plots [(I) and (L)], bar plots of representative genes [(J) and (M)] (mean ± 95% confidence interval expression across all cells in that cluster, with each bar scaled to the maximum mean expression across all clusters), and in situ spatial distribution [(K) and (N)] of excitatory [(I) to (K)] and inhibitory [(L) to (N)] neurons. The number of cells in each cluster was as follows: eL2/3: 589; eL4: 649; eL5: 393; eL6: 368; PV neurons: 111; VIP neurons: 46; SST neurons: 46; NPY neurons: 56. Inclusion of cells in clusters was guided entirely by amplicon representation in each cell without using spatial information; excitatory cell clusters were then named according to the spatial layering observed for that cluster, while inhibitory cell clusters were named according to the dominant cell-type amplicon based on the strong segregation of amplicon markers.
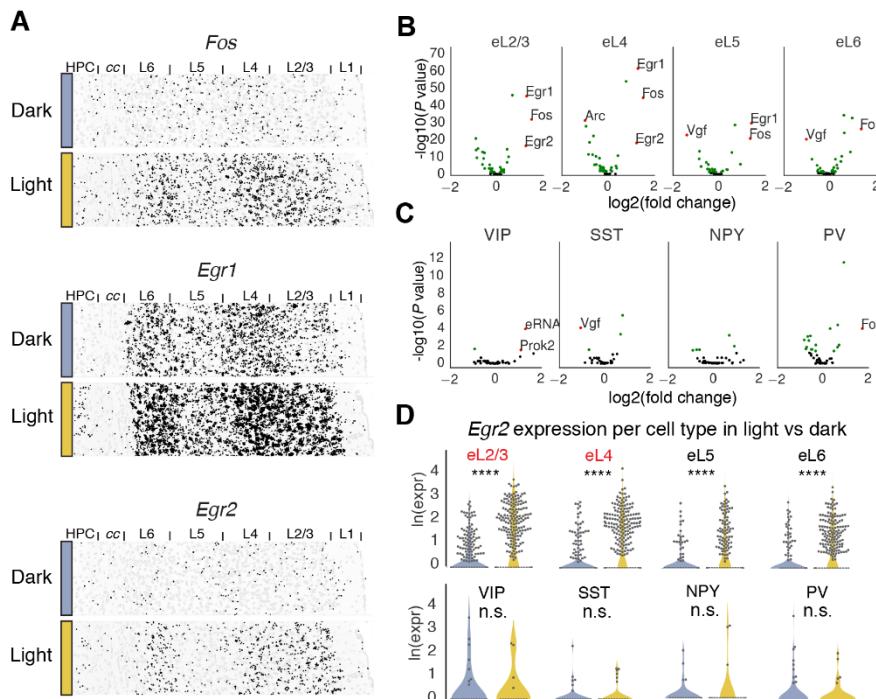
Fig. 3. STARmapping behavioral experience: detecting and quantifying cell type-specific-regulation of activity-regulated genes (ARGs). (A) Validation: spatial expression pattern in visual cortex of prototypical ARGs known as immediate early genes (IEGs): sacrifice was in darkness or after 1 hour light exposure. (B and C) Volcano plots of log fold-change in gene expression between light and dark conditions in inhibitory and excitatory cell types. Genes with significantly increased or decreased expression (false discovery rate adjusted P value < 0.05, Wilcoxon rank-sum test) are labeled in green and the most significantly changed genes (P value <0.05 and fold change > 2) are labeled in red. Many ARGs showed cell-type specificity pointing to discovery of unanticipated cell type-specific logic of excitation-transcription coupling. (D) Violin plot of *Egr2* expression by cell type. ****P < 0.0001, n.s. not significant, Wilcoxon rank-sum test; red-labeled cell types, fold change >2.
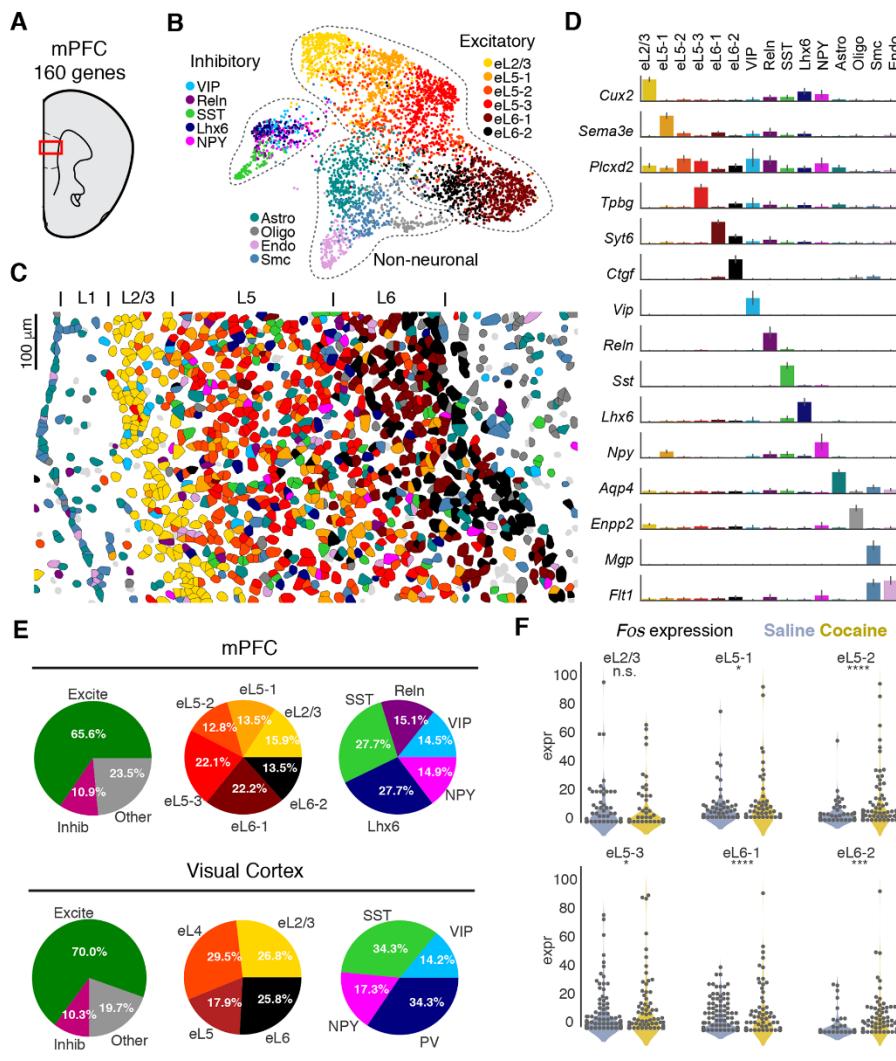
Fig. 4. STARmapping cell types and neural activity in medial prefrontal cortex. (**A**) Diagram of targeted region (red box) containing prelimbic cortex (PrL) within medial prefrontal cortex (mPFC). (**B**) UMAP visualization of all inhibitory (VIP, Reln, SST, Lhx6, NPY), excitatory (eL2/3, eL5-1, eL5-2, eL5-3, eL6-1, eL6-2), and non-neuronal (Astro, Oligo, Smc, Endo) cell types. (**C**) Spatial visualization of cell type layout in mPFC, using the same color scheme as in (B). (**D**) Barplot of representative genes per cluster (mean ± 95% confidence interval), with each bar scaled to the maximum mean expression for that gene across clusters. (**E**) Piecharts showing the relative proportion of each major and minor cell type in both mPFC and visual cortex. (**F**) Violin plots of *Fos* gene induction in different excitatory cell types in mPFC in response to cocaine. The mice were sacrificed after one hour of cocaine or saline injection. Expr = normalized expression. n.s. not significant, *$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$, likelihood ratio test. Astro, astrocytes; Oligo, oligodendrocytes; Smc, smooth muscle cells; Endo, endothelial cells.
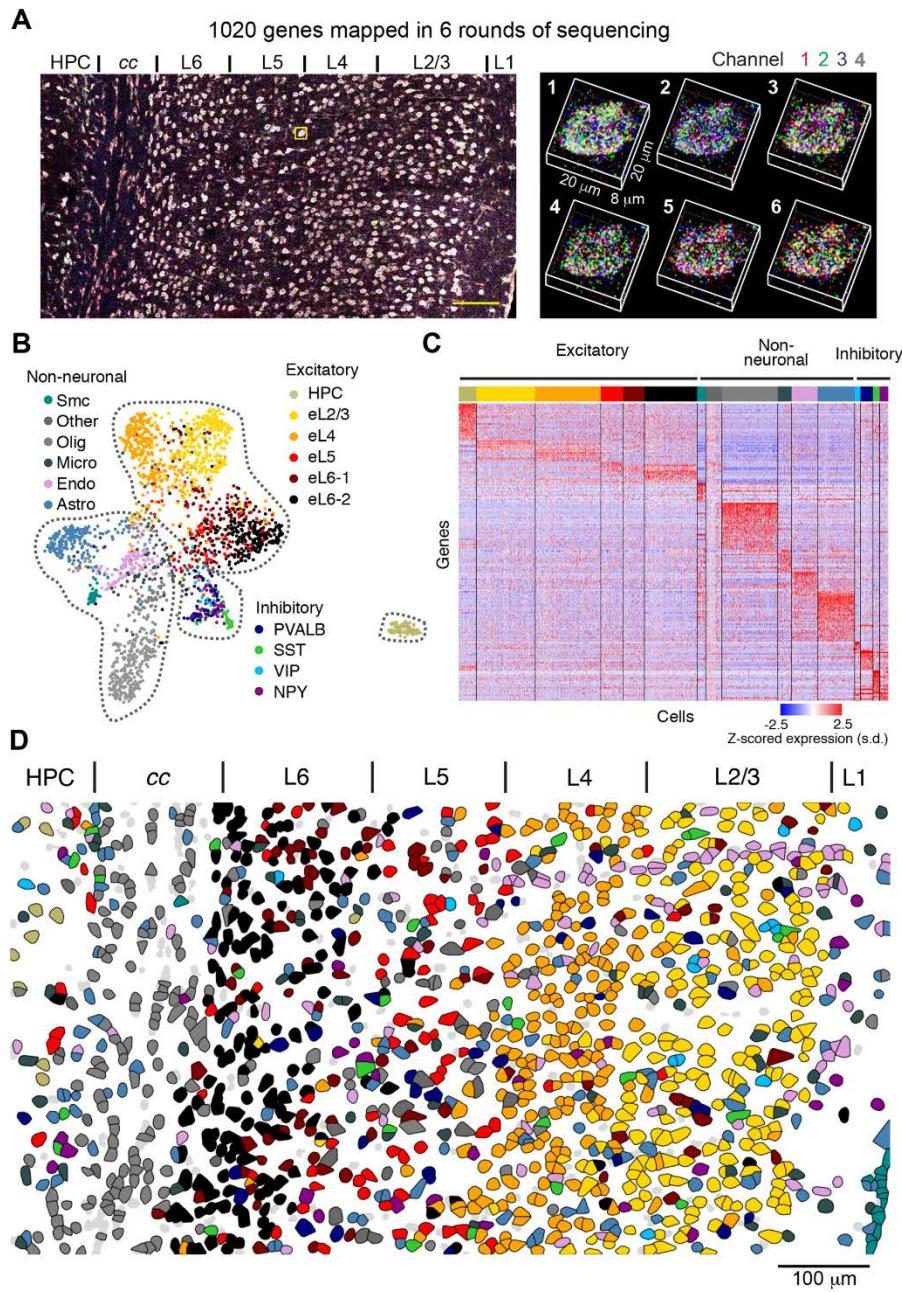
Fig. 5. Simultaneous mapping of 1,020 genes in primary visual cortex by STARmap. (A) Input fluorescence data. Left: maximum-intensity projection of the first sequencing round for 1020 gene experiment, showing all four channels simultaneously; yellow square, zoom region; scale bar, 100 µm. Right: zoom into a single cell showing spatial arrangement of amplicons in 3D across six sequencing rounds. (B) Joint UMAP plot showing all excitatory (HPC, eL2/3, eL4, eL5, eL6-1, eL6-2), non-neuronal (Smc, Other, Olig, Micro, Endo, Astro), and inhibitory (PVALB, SST, VIP, NPY) cell types. (C) Plot of all differentially expressed genes across every cluster, with $P < 10^{-12}$ and log fold change > 1.5. (D) Spatial map of all excitatory, non-neuronal, and inhibitory cell types in visual cortex using the same color code of (B). HPC, hippocampus; Smc: smooth muscle cells; Other, other unclassified cells; Oligo, oligodendrocytes; Micro, microglia; Endo, endothelia cells; Astro, astrocytes.
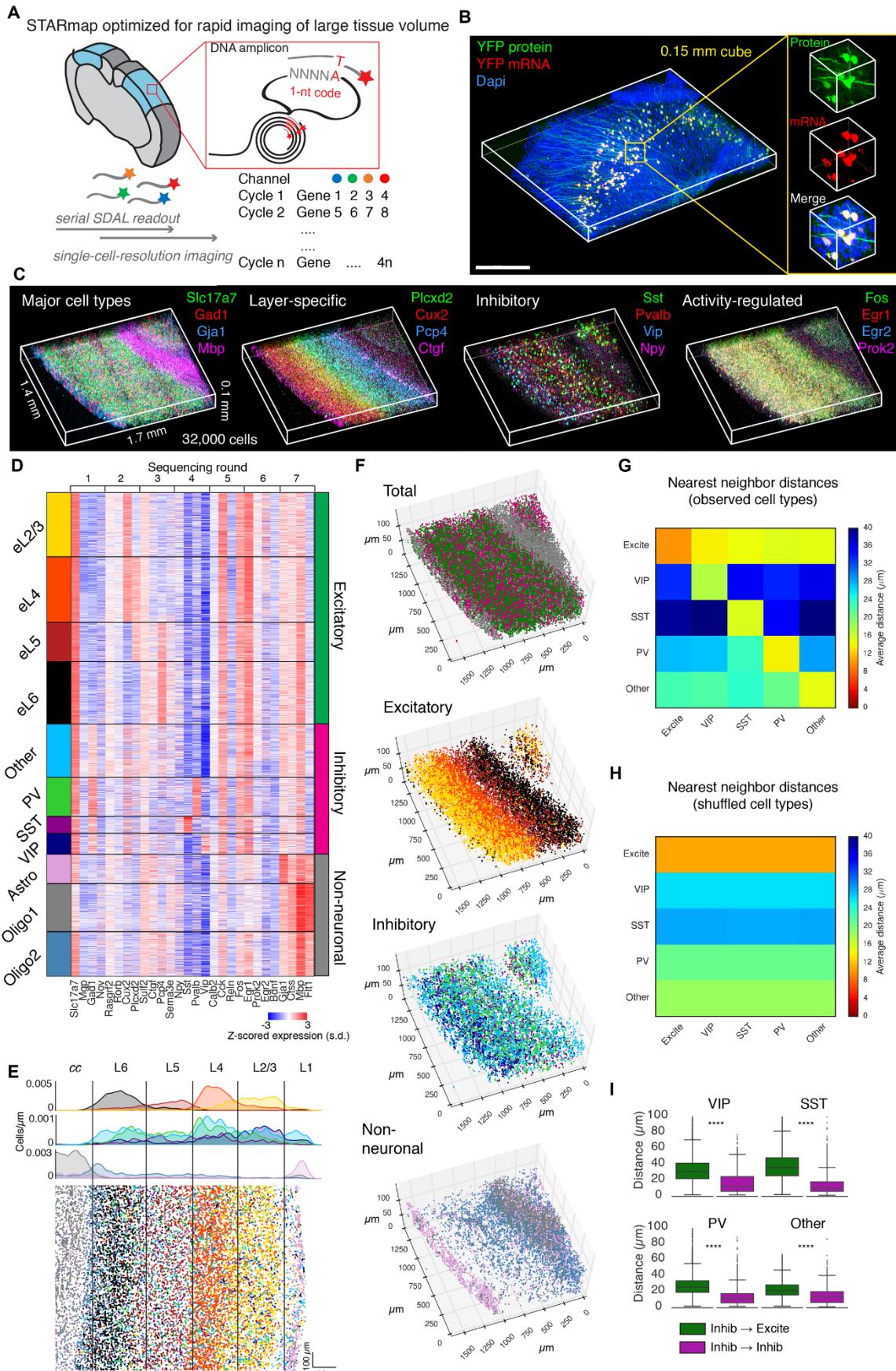
Fig. 6. Three-dimensional architecture of cell types in visual cortex volumes. (A) Volumetric STARmapping via sequential SEDAL gene readout. Using a modified STARmap procedure (fig. S16) and cyclic gene readout (4 genes in each cycle), large tissue volumes can be rapidly mapped at single-cell resolution without oversampling each amplicon. (B) Validation showing specific STARMAP labeling of YFP-expressing neurons (from transgenic Thy1::YFP mouse line) in 3D cortical volume. Scale bar, 0.5 mm. (C) Representative labeling of major cell types (left), layer-specific markers (left center), inhibitory markers (right center), and activity-regulated genes (right) acquired over multiple rounds in visual cortex STARmap volumes. (D) Per-cell expression matrix of 28 genes from 32,845 single cells from one volume clustered into multiple excitatory, inhibitory, and non-neuronal cell types, z-scored across genes for each cell in order to normalize for mean differences in total signal between cells. Columns are sorted by order of sequencing rounds as conducted, in groups of 4. (E) Top: Spatial histograms of excitatory, inhibitory, and non-neuronal cell types using same color labels as (D). Cells were counted in 5 μm bins in a 2D max-projection, and plotted in cell count/μm units as a function of distance from the corpus callosum (cc) to pia, averaged across the bins perpendicular to the cortical layers. Bottom: plot of max-projected cell locations color coded by cluster as in (D). (F) Spatial distribution of each cell type (excitatory, inhibitory, non-neuronal) and subtypes in three dimensions. Each dot represents a single cell; spatial dimensions are in μm. (H) Average nearest-neighbor distances computed in 3D between all excitatory cells (Excite) and each inhibitory cell type. For self-comparisons, the nearest neighbor was defined as the closest non-identical cell; note the persistent self-correlation revealing self-clustering of inhibitory subtypes. (I) Same distances as (H) but using shuffled (randomized) cell type labels. (J) Nearest-neighbor distances computed in 3D between each inhibitory cell of a certain type and any member of the same type (Inhib → Inhib, eg VIP→VIP) or any excitatory neuron (Inhib → Excite). ****$P < 0.0001$, Wilcoxon rank-sum test.

Science

Three-dimensional intact-tissue sequencing of single-cell transcriptional states

Xiao Wang, William E. Allen, Matthew A. Wright, Emily L. Sylwestrak, Nikolay Samusik, Sam Vesuna, Kathryn Evans, Cindy Liu,
Charu Ramakrishnan, Jia Liu, Garry P. Nolan, Felice-Alessio Bava and Karl Deisseroth

ARTICLE TOOLS	http://science.sciencemag.org/content/early/2018/06/20/science.aat5691
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2018/06/20/science.aat5691.DC1
REFERENCES	This article cites 49 articles, 13 of which you can access for free http://science.sciencemag.org/content/early/2018/06/20/science.aat5691#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science

AAAS

Supplementary Materials for

Title: Three-dimensional intact-tissue sequencing of single-cell transcriptional states

Authors: Xiao Wang[1]*, William E. Allen[1,2]*, Matthew A. Wright[1,3], Emily L. Sylwestrak[1], Nikolay Samusik[5], Sam Vesuna[1], Kathryn Evans[1], Cindy Liu[1], Charu Ramakrishnan[1], Jia Liu[4], Garry P. Nolan[5],†, Felice-Alessio Bava[5],†, and Karl Deisseroth[1,3,6],†

Affiliations:
[1] Department of Bioengineering, Stanford University, Stanford, CA94305, USA.
[2] Neuroscience Program, Stanford University, CA 94305, USA.
[3] Department of Psychiatry and Behavioral Sciences, Stanford University, CA94305, USA.
[4] Department of Chemical Engineering, Stanford University, CA94305, USA.
[5] Baxter Laboratory in Stem Cell Biology, Department of Microbiology and Immunology, Stanford University, Stanford, CA 94305, USA.
[6] Howard Hughes Medical Institute, Stanford University, CA94305, USA.
*These authors contributed equally to this work.
†Corresponding authors. Email: deissero@stanford.edu (K.D.); alessio.bava@gmail.com (F.-A. B.); gnolan@stanford.edu (G.P.N.)

This PDF file includes:

Materials and Methods

Mice

All animal procedures followed animal care guidelines approved by Stanford University's Administrative Panel on Laboratory Animal Care (APLAC) and guidelines of the National Institutes of Health. Male C57/BL6 mice (6-10 weeks) were used for experiments. For dark/light experiments, mice were housed on a standard light cycle followed by placement in constant darkness for 5 days. After dark housing, mice were either sacrificed or light-exposed for 1 hour before sacrifice. For cocaine experiments, mice were injected with either saline or 15 mg/kg cocaine 1 hour before sacrifice. For thin sections, animals were anesthetized with isofluorane and rapidly decapitated. Brain tissue was removed, placed in O.C.T, frozen in liquid nitrogen, and sliced using a cryostat (Leica CM1900; detailed information below in the thin-tissue slice section). For large volume samples, animals were anesthetized with Buprenex (100 mg/ml, i.p.), transcardially perfused with cold PBS, then perfused with 4% PFA (detailed information below in the large-volume sample section). Thy1-YFP mice were B6.Cg-Tg(Thy1-YFP)HJrs/J. Transgenic parvalbumin mice were generated by crossing Parv-IRES-Cre (JAX#8069) and Ai14 (JAX # 7908) mice.

Chemicals and enzymes

Chemicals and enzymes listed as name (supplier, catalog number): Gel Slick Solution (Lonza, 50640). PlusOne Bind-Silane (GE Healthcare, 17-1330-01). Poly-L-lysine solution, 0.1% w/v (Sigma, P8920). Ultrapure distilled water (Invitrogen, 10977-015). Glass bottom 12-well and 24-well plates (MatTek, P12G-1.5-14-F and P24G-1.5-13-F). #2 Micro coverglass, 12 mm diameter (Electron Microscope Sciences, 72226-01). O.C.T. Compound (Fisher, 23-730-571). 16% PFA, EM grade (Electron Microscope Sciences, 15710-S). Methanol for HPLC (Sigma-Aldrich, 34860-1L-R). PBS, 7.4 (Gibco, 10010-023 for 1× and 70011-044 for 10×). Tween-20, 10% solution (Calbiochem, 655206). Triton-X-100, 10% solution (Sigma-Aldrich, 93443). OminiPur Formamide (Calbiochem, 75-12-7). 20×SSC buffer (Sigma-Aldrich, S6639). Ribonucleoside vanadyl complex (New England Biolabs, S1402S). Sheared salmon sperm DNA (Invitrogen, AM9680). SUPERase·In (Invitrogen, AM2696). T4 DNA ligase, 5 Weiss U/μL (Thermo Scientific, EL0011). Phi29 DNA polymerase (Thermo Scientific, EP0094). 10 mM dNTP mix (Invitrogen, 100004893). BSA, molecular biology grade (New England Biolabs, B9000S). 5-(3-aminoallyl)-dUTP (Invitrogen, AM8439). BSPEG9 (Thermo Scientific, 21582). Acrylic acid NHS ester, 90% (Sigma-Aldrich, A8060). Methacrylic acid NHS ester, 98% (Sigma-Aldrich, 730300). DMSO, anhydrous (Molecular Probes, D12345). Acrylamide solution, 40% (Bio-Rad, 161-0140). Bis Solution, 2% (Bio-Rad, 161-0142). Ammonium persulfate (Sigma-Aldrich, A3678). N,N,N′,N′-Tetramethylethylenediamine (Sigma-Aldrich, T9281). OminiPur SDS, 20% (Calbiochem, 7991). Protease K, RNA grade (Invitrogen, 25530049). Shrimp Alkaline Phosphatase (New England Biolabs, M0371L). DAPI (Molecular Probes, D1306). NeuroTrace Fluorescent Nissl Stains, yellow (Molecular Probes, N-21480). PMSF (Sigma, 93482). Papain (Worthington, LS003127). Matrigel (Corning Life Sciences, 356234). Neurobasal-A medium (Invitrogen, 21103-049). FBS (HyClone, SH3007103). B-27 supplement (Gibco, 17504044). 2 mM Glutamax (Gibco, 35050-061). Fluorodeoxyuridine (Sigma, F-0503). Anti-NeuN antibody (Abcam, 190565).

Primary mouse cortical neuron culture

Neoortices or hippocampi from mouse pups were removed at postnatal day 0 (P0), digested with 0.4 mg/mL papain and plated onto 24-well glass-bottomed plates pre-coated with 1:30 Matrigel at a density of 65,000 cells per well. Cultured neurons were maintained in Neurobasal-A medium containing 1.25% FBS, 4% B-27 supplement, 2 mM Glutamax and 2 mg/ml fluorodeoxyuridine, and kept in a humid culture incubator with 5% CO_2 at 37°C.

<u>smFISH</u>
Stellaris ShipReady smFISH probes of mouse *Gapdh* with Quasar 570 were purchased from LGC Bioresearch Technologies (SMF-3002-1). smFISH experiments were performed according to the manufacturer's protocols for adhesive cell and frozen mouse brain tissue.

<u>STARmap probe design</u>
SNAIL probes were designed as follows: (1) For genes with multiple transcript isoforms, only the shortest isoforms were considered and the coding regions were used except for non-coding RNAs; (2) Picky 2.2 was used to design the hybridization sequence of each probe pair with length restriction of 40-46 nucleotide; 4 sequences for each gene were designed; (3) the resulting complementary DNA (cDNA) sequences (40-46 nt) were split into halves of 20-25 nt, with an 0-2 nt gap in between, and with the best match of melting temperature (T_m) between the two halves. All the probes were under 60 nt and manufactured as by Integrated DNA Technologies (IDT). For the 160- and 1020-transcript experiments, the homemade sequencing reagents included six reading probes (R1 to R6) and sixteen 2-base encoding fluorescent probes (2base_F1 to 2base_F16) labeled with Alexa 488, 546, 594 and 647. YFP probes were ordered with Acrydite modification, the SNAIL primer probes for the 28-transcript large-volume experiment were ordered with 5'amino modification and further modified by AA-NHS; and the sequencing was carried out using 11-nt orthogonal reading probes (OR1 to OR7) and four 1-base fluorescent probes (1base_F1 to 1base_F4). All sequences were included in Table S1 (SNAIL probes) and Table S2 (SEDAL probes).

<u>STARmap procedure for cell culture and thin tissue sections</u>

Sample preparation: Glass-bottom 12- or 24-well plates were treated by methacryloxypropyltrimethoxysilane (Bind-Silane). For brain-tissue slices, 12-well plates were further treated with poly-L-lysine solution. #2 Micro coverglasses (12 mm) were pretreated with Gel Slick for later polymerization following manufacturer's instructions. Primary neuron cell cultures were fixed with 1.6% PFA in PBS for 10 min then transferred to pre-chilled (-20°C) methanol and kept at -80°C for at least 15 min (and up to 1 wk). For brain-tissues, freshly harvested mouse brains were immediately embedded in O.C.T. and snap-frozen. Tissues were either stored at -80°C or transferred to the cryostat and cut as 16-µm slices. Slices containing primary visual cortex were mounted in the pretreated glass-bottom plates. Brain slices were fixed with 4% PFA in PBS at r.t. for 10 min, permeabilized with -20°C methanol and then placed at -80°C for 15 min before hybridization.

Library construction: SNAIL probes were dissolved at 100 or 200 µM in ultrapure RNase-free water and pooled. The probe mixture was heated at 90°C for 2 to 5 min and then cooled down at r.t. The samples were taken from -80°C and equilibrated to r.t. for 5 min, washed by PBSTR (0.1% Tween-20, 0.1 U/µL SUPERase·In in PBS) for 2-5 min and incubated in 1× hybridization buffer (2X SSC, 10% formamide, 1% Tween-20, 20 mM RVC, 0.1 mg/ml salmon sperm DNA and pooled SNAIL probes at 100 nM per oligo) in 40°C humidified oven with gentle shaking overnight. The samples were then washed for 20 min, twice, with PBSTR, followed by one 20 min wash in 4X SSC dissolved in PBSTR at 37°C. Finally, the sample was briefly rinsed with PBSTR once at r.t. The samples were then incubated for two hours with T4 DNA ligation mixture (1: 50 dilution of T4 DNA ligase supplemented with 1X BSA and 0.2 U/µL of SUPERase-In) at room temperature with gentle agitation. Then samples were washed twice with PBSTR, incubated with RCA mixture (1: 50 dilution of Phi29 DNA polymerase, 250 µM dNTP, 1X BSA and 0.2

U/µL of SUPERase-In and 20 µM 5-(3-aminoallyl)-dUTP) at 30 °C for two hours under agitation. The samples were next washed twice in PBST (PBSTR omitting SUPERase·In) and treated with 20 mM Acrylic acid NHS ester in PBST for two hours at r.t. The samples were briefly washed with PBST once, then incubated with monomer buffer (4% acrylamide, 0.2% bis-acrylamide, 2X SSC) for 30 min at RT. The buffer was aspirated and 10 µL of polymerization mixture (0.2% ammonium persulfate, 0.2% tetramethylethylenediamine dissolved in monomer buffer) was added to the center of the sample, which were immediately covered by Gel Slick coated coverslip and incubated for 1 hour at r.t., then washed by PBST twice for 5 min each. The tissue-gel hybrids were digested with Proteinase K (0.2 mg/ml) at 37ºC for one hour, then washed with PBST three times (5 min each).

Imaging and sequencing: For single-gene detection, the 19-nt fluorescent oligo complementary to DNA amplicon was diluted at 500 nM in 1X SSC dissolved in PBST and samples incubated at r.t. for 30 min, then washed by PBST three times for 5 min each before imaging. For sequencing, each sequencing cycle began with treating the sample with stripping buffer (60% formamide,0.1% Triton-X-100) at r.t. for 10 min twice, followed by three PBST washes, 5 min each. The sample was incubated with the sequencing mixture (1XT4 DNA ligase buffer, 1: 25 dilution of T4 DNA ligase, 1XBSA, 10 µM reading probe and 5 µM fluorescent oligos) at r.t. for 3 hours. Then the sample was rinsed with washing and imaging buffer (2XSSC and 10% formamide) for three times (10 min) each before proceeding to imaging. DAPI staining was performed following manufacturer's instruction before Cycle 1 and after Cycle 6 for the purpose of registering sequencing images with Nissl staining. Nissl staining was performed after Cycle 6 following manufacturer's instruction for the purpose of cell segmentation. Images were acquired using Leica TCS SP8 confocal microscopy with a 405 diode, white light laser, 40× oil-immersed objective (NA 1.3), with voxel size of 78 nm × 78 nm × 315 nm.

Thin-section STARmap Data Processing
All image processing steps were implemented using MATLAB R2017A. For the full list of software packages used, see "Software References Note" below. The STARmap pipeline and analysis tools can be found at: https://github.com/weallen/STARmap

Image Registration: Image registration was accomplished using a three-dimensional fast Fourier transform (FFT) to compute the cross-correlation between two image volumes at all translational offsets. The position of the maximal correlation coefficient was identified and used to translate image volumes to compensate for the offset. All images were registered to the first round of sequencing, first through a global transform across the entire field of view, and then separately for each tile (corresponding to the individual tiled fields of view used by the microscope during image acquisition).

Spot Calling: After registration, individual dots were identified separately in each color channel on the first round of sequencing. For the 160 gene experiment, dots of approximately 7 pixels in diameter were identified by first filtering the volume with a three-dimensional Laplacian of Gaussian filter, and then finding local maxima in 3D. For the 1020 gene experiment, dots were found using a similar approach but in which the Laplacian of Gaussian was computed at multiple scales and the maxima was found across these scales. After identifying each dot, the dominant color for that dot across all four channels was determined on each round in a 3x3x1 voxel volume surrounding the dot location.

Barcode Filtering: Dots were first filtered based on quality score. The quality score quantified the extent to which each dot on each sequencing round came from one color rather than a mixture of color. The

barcode codebook was converted into colorspace, based on the expected color sequence following 2-base encoding of the barcode DNA sequence. Dot color sequences that passed the quality threshold and matched sequences in the codebook were kept and identified with the specifc gene that that barcode represented; all other dots were rejected. The high-quality dots and associated gene identities in the codebook were then saved out for downstream analysis.

2D Cell Segmentation: Nuclei were manually identified from a maximum intensity projection of the DAPI channel following the final round of sequencing. Cell bodies were first identified using a Random Forests classifier implemented in Ilastik, based on Nissl staining. The classifier was trained on a randomly selected subset of cropped regions from all samples, and then applied to the full image. The thresholded probability map was then pixel-dilated to fill in remaining holes. Finally, a marker-based watershed transform was then applied to segment the thresholded cell bodies based on the combined thresholded cell body map and identified locations of nuclei. A convex hull was computed around each cell body. Points overlapping each convex hull in 2D were then assigned to that cell, to compute a per-cell gene expression matrix.

Single-cell data preprocessing: All single-cell analyses were implemented using a custom software package in Python for the analysis of STARmap experiments. The per-cell expression matrix was first normalized for the expression value E_{ij} across all genes j for each cell i with the formula:

$$N_{ij} = \ln(1+\text{median}(E_{i\cdot})*(E_{ij}/\Sigma\ E_{i\cdot}))$$

For clustering, effects relating to the number of transcripts per cell, the identity of the sample, and the experimental condition (light vs dark) were regressed out using the linear model:

$$E_{ij} = nSpots_i + exptID_i + exptCond_i$$

with the assumption that E_{ij} is Poisson distributed.

Top-level single-cell clustering: After normalization and scaling, principal-components analysis was applied to reduce dimensionality of the cellular expression matrix. Based on the explained variance ratio, the top PCs were then used for top-level clustering, based on manual analysis of the explained variance ratio per PC. The top PCs were then clustered using the shared nearest neighbor (SNN) algorithm with Louvain distances (*23*). Clusters enriched for the excitatory neuron marker *Slc17a7* (vesicular glutatamate transporter), inhibitory neuron marker *Gad1*, and non-neuronal marker *Mqp* were manually merged to form three major clusters representing these cell types. The cells were displayed using the Uniform Manifold Approximation and Projection (UMAP) (https://github.com/lmcinnes/umap). The cells for each cluster were then subclustered using PCA decomposition followed by Louvain SNN (*25*) clustering to determine specific cell types.

Single-cell subclustering: The inhibitory, excitatory, and non-neuronal clusters were then subclustered using the same approach as applied to the major clusters.

Differential expression analysis: Genes specifically variable in this subcluster were selected by computing the *P* value of differential expression of each gene between each cluster and all other clusters, using a bimod test (likelihood-ratio test, *45)*. The *P* values were FDR corrected based on the number of clusters.

STARmap procedure for thick tissue sections

Sample preparation: Glass-bottom 12- well plates and micro coverglass (12 mm) were pretreated the same as for thin tissue sections. The mice were perfused with PFA, post-fixed on ice for 2-3 hours, transferred to PBS on ice for 30 min, and cut as 150-μm slices. The slices containing primary visual cortex were transferred into the pretreated glass-bottom plates and washed once with ice-cold PBS.

Library construction: The samples were precleared with -20°C-chilled methanol at 4°C for 1 hour, then incubated in 1× permeabilization and hybridization buffer (2X SSC, 10% formamide, 1% Triton-X-100, 20 mM RVC, 0.1 mg/ml salmon sperm DNA and pooled SNAIL probes at 100 nM per oligo, and 0.2% SDS) in 40°C humidified oven with gentle shaking for two days. The methanol treatment protects samples from expansion and deformation; for co-detection of protein and RNA (Thy1-YFP), methanol preclearance was skipped. The samples were then washed with PBSTV (0.1% Triton and 2 mM RVC) at 37°C twice for 1 hour each, then PBS once for another hour. The samples were next incubated with polymerization mixture I (4% Acrylamide, 0.2% bis-acrylamide, 2X SSC, 0.1% VA-044) for 1 hour at 4°C. The buffer was aspirated and 40 µL of polymerization mixture II (0.1% ammonium persulfate, 0.1 % tetramethylethylenediamine dissolved in polymerization mixture I) was added to the center of the samples, which were immediately covered by Gel Slick coated coverslip and incubated for 1 hour at 40°C. The samples were then washed by PBSTV twice for 20 min each. The tissue-gel hybrids were digested with Proteinase K (0.2 mg/ml in 2XSSC, 1% SDS) at 37°C overnight, then washed with PBSF (2 mM PMSF in PBS) and PBSTR twice times (30 min each). The samples were next incubated for 12 hours with T4 DNA ligation mixture (1: 50 dilution of T4 DNA ligase supplemented with 1X BSA and 0.2 U/µL of SUPERase-In) at room temperature with gentle agitation. Then the samples were washed twice with PBSTR (1 hour each), then incubated with RCA mixture (1: 50 dilution of Phi29 DNA polymerase, 250 µM dNTP, 1X BSA and 0.2 U/µL of SUPERase-In and 20 µM 5-(3-aminoallyl)-dUTP) at 30 °C for 12 to 24 hours. Finally, the samples were crosslinked by BSPEG9 (*16*).

Imaging and sequencing: The samples were first stained with DAPI overnight then rinsed by PBS for 1 hour. Each cycle began with treating the sample with stripping buffer (60% formamide,0.1% Triton-X-100) at r.t. for 20 min twice, followed by three PBST washes, 20 min each. The sample was incubated with large volume sequencing mixture (1XT4 DNA ligase buffer, 1: 50 dilution of T4 DNA ligase, 1XBSA, 5 µM orthogonal reading probe and 400 nM 1-base fluorescent oligos) at r.t. for 4 hours. Then the sample was rinsed with PBS twice (20 min each) before proceeding to imaging. Images were acquired using Leica TCS SP8 confocal microscopy with a 405 nm diode, white light laser, and 25× water-immersion objective (NA 0.95), with voxel size of 0.9 μm × 0.9 μm × 1 μm.

Large-volume STARmap Data Processing
Image Registration 3D FFT registration was again applied, except using the DAPI channel for registration. Specifically, the DAPI channel on each round was registered to the first round globally, and then piece-wise in a 4x5 grid corresponding to the field-of-view tiles used to acquire the image.

Cell Finding and Quantification: After registration, cells were identified using minima of a Laplacian-of-Gaussian filter applied to the DAPI channel. To quantify the expression of each gene, the average intensity in each color channel was averaged in a 10×10×3 voxel volume around each nucleus.

3D Cellular Analysis: Cells were first clustered into inhibitory, excitatory, and non-neuronal using *Gad1*, *Slc17a7*, and several non-neuronal genes using K-means clustering. Each cluster was then subclustered using K-means clustering. The initial values of the K-means clustering were set to the average expression of each marker genes. To compute distances between cell types, the nearest-neighbor distance was computed between cells in all excitatory neurons and each inhibitory neuron subtype, using a *k*D-tree (*46*) for fast nearest neighbor computations with Euclidean distance.

<u>*Actb* spike-in to evaluate the physical limits of STARmap</u>
The 100 and 1000 genes all shared a common 18 nt sequence (sequence A) in the padlock which would be amplified in the final nanoballs. We then designed another set of SNAIL probes for *Actb* with a different 18 nt sequence (sequence B). The SNAIL probes of *Actb* and those of 0, 100, or 1000 genes were mixed together and used in the hybridization. Both the *Actb* spike-in and 100 & 1,000 genes had gone through the same ligation and amplification step to ensure equal efficiency. For readout, two fluorescent detection oligos (Alexa488-probes complementary to sequence B and Alexa 647-probes targeting sequence A) were added to the samples; the amplicons of *Actb* and amplicons of the rest of the genes were be imaged in two separate channels. We then tested if the number of amplicons of *Actb* RNA would be diluted by increased numbers of other genes, as an indication of molecular crowding (fig. S15).

<u>Immunostaining of CLARITY tissue</u>
PFA fixed tissue was processed as described previously (*15*). Briefly, 200 μm thick PFA fixed brain sections were placed in a 1% acrylamide embedding solution at 4°C for 23 hours, embedded at 37°C for 4 hours, and then passively cleared for 5 days. Cleared sections were washed in PBST for 2 days, stained with anti-NeuN (1:100) for 24 hours at RT, and then washed for 24 hours in PBST. Sectioned were imaged using confocal microscopy.

Open-source Software References Note
The following open-source software was used in the STARmap image processing and analysis pipeline:

- SciPy: Jones E., Oliphant E., Peterson P., *et al.* SciPy: Open Source Scientific Tools for Python, (2001-). http://www.scipy.org

- Numpy: Oliphant E. A., Guide to NumPy, *Trelgol Publishing* (2006). http://www.numpy.org

- Matplotlib: Hunter J. D. Matplotlib: A 2D graphics environment, *Computing In Science & Engineering* (2007). http://www.matplotlib.org

- Scikit-learn: Pedregosa *et al.*, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* **12**, 2825-2830 (2011). http://scikit-learn.org/stable/index.html

- Ilastik: Sommer C., Strähle C., Köthe U., Hamprecht F. A. ilastik: Interactive Learning and Segmentation Toolkit, *Eighth IEEE International Symposium on Biomedical Imaging (ISBI) Proceedings* (2011). http://www.ilastik.org

- UMAP: McInnes L. and Healy J.. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction https://arxiv.org/abs/1802.03426 (2018). https://github.com/lmcinnes/umap

- Pandas: McKinney W. Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 51-56 (2010). https://pandas.pydata.org/

- Scikit-image: van der Walt S. *et al.,* scikit-image: Image processing in Python. *PeerJ* **2**:e453 (2014). http://scikit-image.org/

- IPython: Perez F., Granger B.E. IPython: A System for Interactive Scientific Computing. *Computing in Science & Engineering*, **9**, 21-29 (2007). https://ipython.org/

- Seaborn: http://seaborn.pydata.org

- Statsmodels: Skipper S., and Perktold J. Statsmodels: Econometric and statistical modeling with Python. *Proceedings of the 9th Python in Science Conference (*2010). https://www.statsmodels.org/stable/index.html

- OpenCV: Bradski G., The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000). https://opencv.org/

- ImageJ: Ruden C.T., *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18:529** (2017). https://imagej.nih.gov/ij/

References

45. A. McDavid *et al.*, Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* **15**, 461-467 (2013).

46. J. L. Bentley, Multidimensional binary search trees used for associative searching. *Commun.* ACM **18**, 509-517 (1975).

47. C. Larsson, I. Grundberg, O. Soderberg and M. Nilsson, *Nat. Methods* **7**, 395‑397, (2010).

48. E. Z. Macosko *et al*., Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202-1214 (2015).

49. P. L. Ståhl et al., Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82 (2016).

A

DNA-DNA complementary length: 6 nt

primer   padlock

mRNA

DNA-RNA complementary length: 19-25 nt

T_m (DNA-DNA) $\ll T_{hybridazation} < T_m$ (DNA-RNA)

B

■ commercial smFISH ■ SNAIL probe

SNR

cortical cell culture   mouse visual cortex

C commercial smFISH

Gapdh

50% laser

D SNAIL probe

0.5% laser

cortical cell culture

E

50% laser

F

1% laser

mouse visual cortex

G

FISSEQ   **Padlock**   **PLAYR**   **STARmap: SNAIL probe**

RT   RT   probe1 probe2   primer padlock

mRNA   mRNA   mRNA   mRNA

Reverse transcription *Reverse transcription* *1st Hybridization* *Hybridization*
(efficiency-limiting) *(efficiency-limiting)*

cDNA   padlock   probe 3   Lig
                              probe 4   mRNA
             cDNA

mRNA removal *mRNA removal & hybridization* *2nd Hybridization* *Ligation*

Cir   Lig   Lig Lig
cDNA   cDNA

Circularization *Gap-filling & ligation* *Ligation* *Amplification*

cDNA   cDNA   Phi29   mRNA   Phi29

Annealing primer *Annealing primer* *Amplification*

Phi29   Phi29

Amplification *Amplification*

enzymes

H

gene markers

mean of RNAs/cell

0.26   0.67

n = 151
p = 6.6 x 10⁻⁵

scRNA-seq STARmap

I

| Methods	Efficiency
scRNA-seq	6%~40% (*2*)
FISSEQ	0.2% (*21*); 0.01% (*2*)
Padlock	30% (with LNA probe, *6*); 5% (*2*)
PLAYR	N/A
STARmap	no less than scRNA-seq (panel H)

9

Fig. S1. SNAIL probes for high-quality RNA imaging: mouse brain instantiation. (**A**) Design of SNAIL probes (one component of STARmap): each primer or padlock probe has 19-25 nt (labeled by blue double-headed arrows) to hybridize with target RNA with a designed T_m of 60°C, while the complementary sequence between the primer and padlock is only 6 nt on each arm (labeled by red double-headed arrow) with T_m below room temperature, so that primer-padlock DNA-DNA hybridization is negligible during DNA-RNA hybridization at 40°C, but allows DNA ligation by T4 DNA ligase in the following step. T_m, melting temperature of nucleic acids. (**B**) Comparison of signal-to-noise ratios (SNR; mean intensity of signal spots/mean intensity of background) of commercial smFISH probes and SNAIL probes targeting *Gapdh* mRNAs in mouse cortical cell cultures and mouse visual cortex sections. Error bars: standard deviation (s.d.) of spot intensity. Error bars are s.d. of 39,398 pixels; 30,297 pixels; 97,555 pixels; and 19,392 pixels corresponding to RNA signals out of 640,000 pixels in acquired images; *** $P<0.0001$, Student's *t*-test. (**C-F**) Fluorescence images of *Gapdh* smFISH (**C** and **E**) and SNAIL probes (**D** and **F**) in cortical cell cultures (**C** and **D**) and visual cortex sections (**E** and **F**); scale bars, 10 μm. (**G**) Comparison of multiplexed RNA imaging methods using rolling circle amplification (RCA): in comparison with FISSEQ and padlock probes, SNAIL probes have overcome the efficiency-limiting step of reverse transcription and greatly simplified the experimental procedure; while PLAYR requires four probes, one additional step and two ligation sites, SNAIL only requires a pair of probes and one ligation site. (**H**) Boxplots of RNAs per cell of 151 cell type gene markers measured by single-cell RNA sequencing (scRNA-seq; *25*) and STARmap (extracting from 160-gene mapping of visual cortex). Box: the first and third quartiles; middle line: median; whisker: 5% and 95% data points. *P* value, rank-sum test. (**I**) Summary of single-cell RNA sequencing and RCA-based multiplexed RNA detection methods, numbers were extracted from references (*2, 6, 21, 47*).

A DNA amplicon — 5-(3-aminoallyl)-dU — acrylic acid NHS (AA-NHS) — acrylamide moiety

B visual cortex: without hydrogel & clearing

C visual cortex: with hydrogel & clearing

D habenula: without hydrogel & clearing

E habenula: with hydrogel & clearing

F + AA-NHS / − AA-NHS; Fluorescence intensity; fresh, 24h at r.t.

G + AA-NHS fresh | + AA-NHS 24h at r.t. | − AA-NHS fresh | − AA-NHS 24h at r.t.

**H** Cycle1 | Strip | Cycle2 | Merge Cycle1 and Cycle2

11

Fig. S2. Distinct hydrogel-tissue chemistry (HTC) for background reduction and amplicon immobilization in STARmap. (**A**) Schematic of hydrogel-tissue chemistry for STARmap in thin tissue slices. DNA amplicons are synthesized in the presence of minor levels of 5-(3-aminoallyl)-dUTP, which replaces T at a low rate and allows further functionalization with the polymerizable acrylamide moiety using acrylic acid N-hydroxysuccinimide ester (AA-NHS), so that the DNA amplicons are covalently anchored within the polyacrylamide network at multiple sites. (**B-D**) Fluorescence images (summed intensity from all four fluorescent channels) showing 160-gene detection in mouse visual cortex (**B** and **C**) and 16-gene detection in mouse medial habenula (**D** and **E**). Compared to untreated samples (**B** and **D**), samples treated with 5-(3-aminoallyl)-dUTP HTC and clearing of lipids and proteins (**C** and **E**) showed reduced opacity and autofluorescence. Scale bars: 50 μm. (**F-G**) DNA-gel crosslinking is indispensable to maintain DNA amplicons in the gel. 160-gene samples were prepared with or without AA-NHS and imaged within medial prefrontal cortex. Fresh samples prepared without AA-NHS had 36% signal loss compared to AA-NHS treated samples and had suffered from further 40% signal loss after stored at room temperature for 24 hours, while AA-NHS treated sample only had 9 % signal change (**F**). (**F**) Fluorescence intensities are the mean of four technical replicates with the imaging dimension of 120 μm ×120 μm × 3μm. Error bar, mean ± s.d.; *** $p < 0.001$, two-sided t-test. (**G**) Fluorescence images. Scale bars, 3.5 μm. (**H**) Zoomed-in fluorescence images of one neuron in visual cortex detecting *Gapdh* RNA by STARmap for Cycle 1, stripping, Cycle 2 and merged images of Cycle 1 and Cycle 2, demonstrating stable spatial position of DNA amplicons over sequencing cycles. Scale bars, 2 μm.

A

Sequencing with Error-reduction by Dynamic Annealing and Ligation (SEDAL)

reading probe mixture (11 nt):
T_m ~ r.t.

8-nt fluorescent decoding probe

19-nt product for imaging:
T_m >> r.t.

illustrating SEDAL by Fig.1E Cycle 4
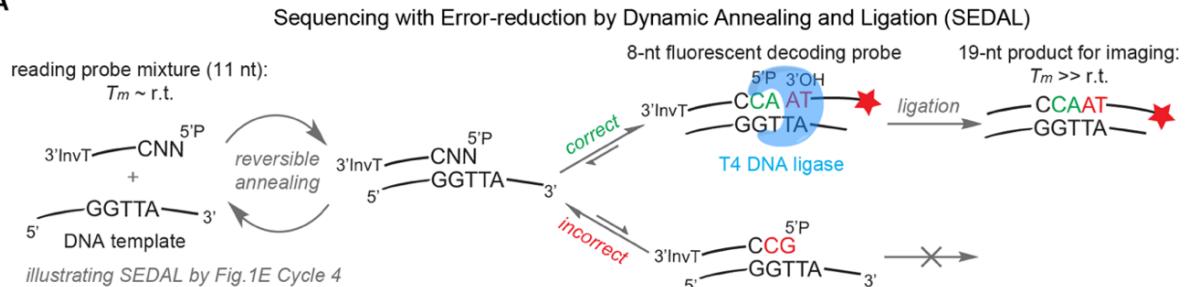
B

| Method	Reagents	Background issue	Cycles to decode 5 continuous bases	Encoding scheme	Cumulative error	Error rejection
SOLiD	Commercial	Yes	9	2-base	Yes	Yes
cPAL	Homemade	No	5	1-base	No	No
SEDAL	Homemade	No	6	2-base	No	Yes

C — SOLiD: signal
D — SOLiD: background
E — SEDAL: signal
F — SEDAL: background

G

1-base encoding

ACGT
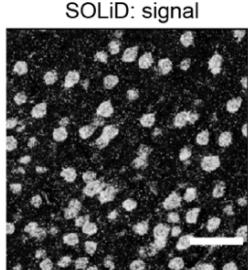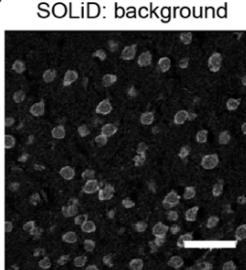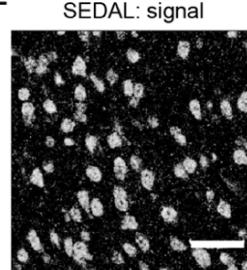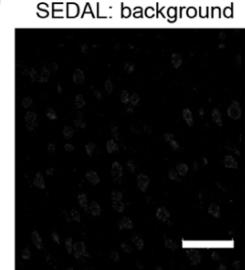
Example

Cycle	1	2	3	4	5
Color	🟢	🔵	🔵	🟠	🔵
Letter	G	T	T	A	T
Results	**GTTAT**				

cycle 3 error
🔵 → 🟢

one error: wrong code

Cycle	1	2	3	4	5
Color	🟢	🔵	🟢	🟠	🔵
Letter	G	T	T	A	T
Results	**GTGAT**				

H

2-base encoding

ACGT

Example

Cycle	1	2	3	4	5	6
Color	🔵	🟢	🔵	🔴	🔴	🟢
Letter	GG	GT	TT	TA	AT	TG
Results	**GGTTATG**					

correct reads:
GXXXXXG
one error:
GXXXXXH,H ≠ G
rejected

cycle 3 error
🔵 → 🟢

Cycle	1	2	3	4	5	6
Color	🔵	🟢	🟢	🔴	🔴	🟢
Letter	GG	GT	TG	GC	CG	GT
Results	**GGTGCGT**					

I

cPAL: 1-base encoding

Malat1   *Actb*   *Calm1*   *Snap25*

J

SEDAL: 2-base encoding

Malat1   *Actb*   *Calm1*   *Snap25*

K

cPAL: 1-base encoding

TTATC: *Calm1*
AGGAA: *Snap25*
CACCG: *Actb*
GCTGT: *Malat1*

Code: XXXXX
All codes: 1024
Correct codes: 4
Correct spots: 70.6%
Error rate: 29.4%

L

SEDAL: 2-base encoding

TTATC: *Calm1*
GCTGT: *Malat1*
CACCG: *Actb*
AGGAA: *Snap25*

Code: GXXXXXX
All codes: 4096
Correct codes: 4
Spots in GXXXXXG: 79.8%
Correct spots: 98.2%
Error rate: 1.8%

13
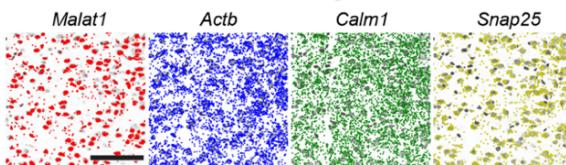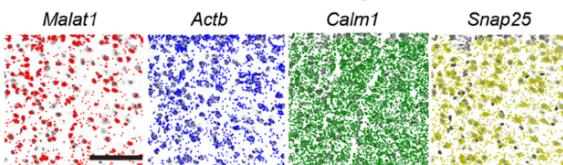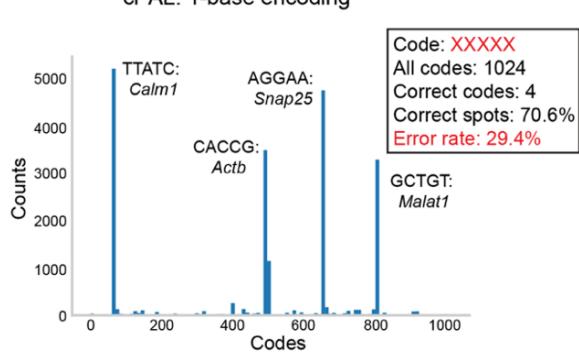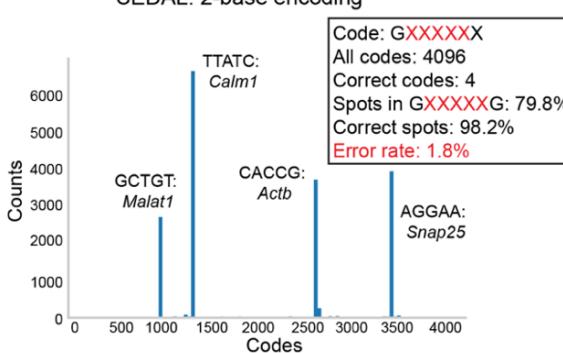
Fig. S3. The low-background error-correcting *in situ* sequencing component of STARmap: sequencing by dynamic annealing and ligation (SEDAL). (A) Schematic of SEDAL. SEDAL involves a T4 DNA ligase with activity strongly hindered by base mismatches, and two kinds of sequencing probes: reading probes that set the base position to be interrogated, and fluorescent decoding probes that transduce base information into colors for imaging. Unlike other sequencing-by-ligation methods which use pre-annealed reading probes (or equivalent), the reading probe in SEDAL is short (11 nt, with T_m near room temperature), partially degenerate (as shown here, for cycle 4, the first two base at 5' end are N, equal amount mixture of A, T, C and G), and mixed with decoding probes and T4 DNA ligase for a one-step reaction. At room temperature, the reading probe remains in a dynamic state of annealing with and detaching from the DNA template. Only when the reading probe perfectly matches the DNA template, T4 DNA ligase (blue) ligates it to the fluorescent 8-nt decoding probe. The short reading and decoding probes are then washed away, leaving fluorescent 19-nt products stably hybridized to the DNA amplicon for imaging. For the next cycle, previous fluorescent products are stripped and the reading probe includes one more degenerate base to shift the reading frame by one base (Fig. 1E). 5'P: 5' phosphate. 3'InvT: 3' inverted dT base that prevents self-ligation of the reading probe. 3'OH: 3' hydroxyl group. **(B)** Comparison of key properties of all sequencing-by-ligation methods. **(C-F)**. Background problem associated with the commercial SOLiD sequencing kit when applied to mouse brain tissue (**C** and **D**) while custom SEDAL reagents exhibit minimal background (**E** and **F**). Signal images (**C** and **E**) represent the first cycle of sequencing for the *Malat1*, *Actb*, *Calm1* and *Snap25* genes. Background images (**D** and **F**) were acquired after cleavage/stripping of the first cycle. Scale bars, 50 μm. **(G** and **H)** Schematic of the 1-base and 2-base encoding paradigms, along with example results with or without a single sequencing error (wrong color) during cycle 3. For 1-base encoding, a single sequencing error leads to one base mutation and thus the wrong 5-base code (**G**). For 2-base encoding, the six-cycle paradigm plays an error-reduction role: since a single error during any sequencing cycle will propagate and cause the flanking known base G to mutate into other bases; thus erroneous reads can be rejected (**H**). **(I-L)** Actual data from cPAL (representing a 1-base encoding scheme) and SEDAL (representing a 2-base encoding scheme) applied to 4-gene detection in mouse visual cortex. The SNAIL probes for *Malat1*, *Actb*, *Calm1* and *Snap25* were identical for the two conditions and the Hamming distance for each pair of the four 5-base codes was 5 (i.e. complete non-homology); with such sparse coding, the sequencing error rate was estimated by the percentage of wrong spots (not the four 5-base codes used) out of all detected spots. **(I-J)** Spatial map of the four genes detected by cPAL (**I**) and SEDAL (**J**). **(K-L)** The error rate of cPAL in the 4-gene experiment was 29.4% (**K**) while the error rate of SEDAL was 1.8% after the built-in error reduction (**L**), related to **I** and **J**, respectively.
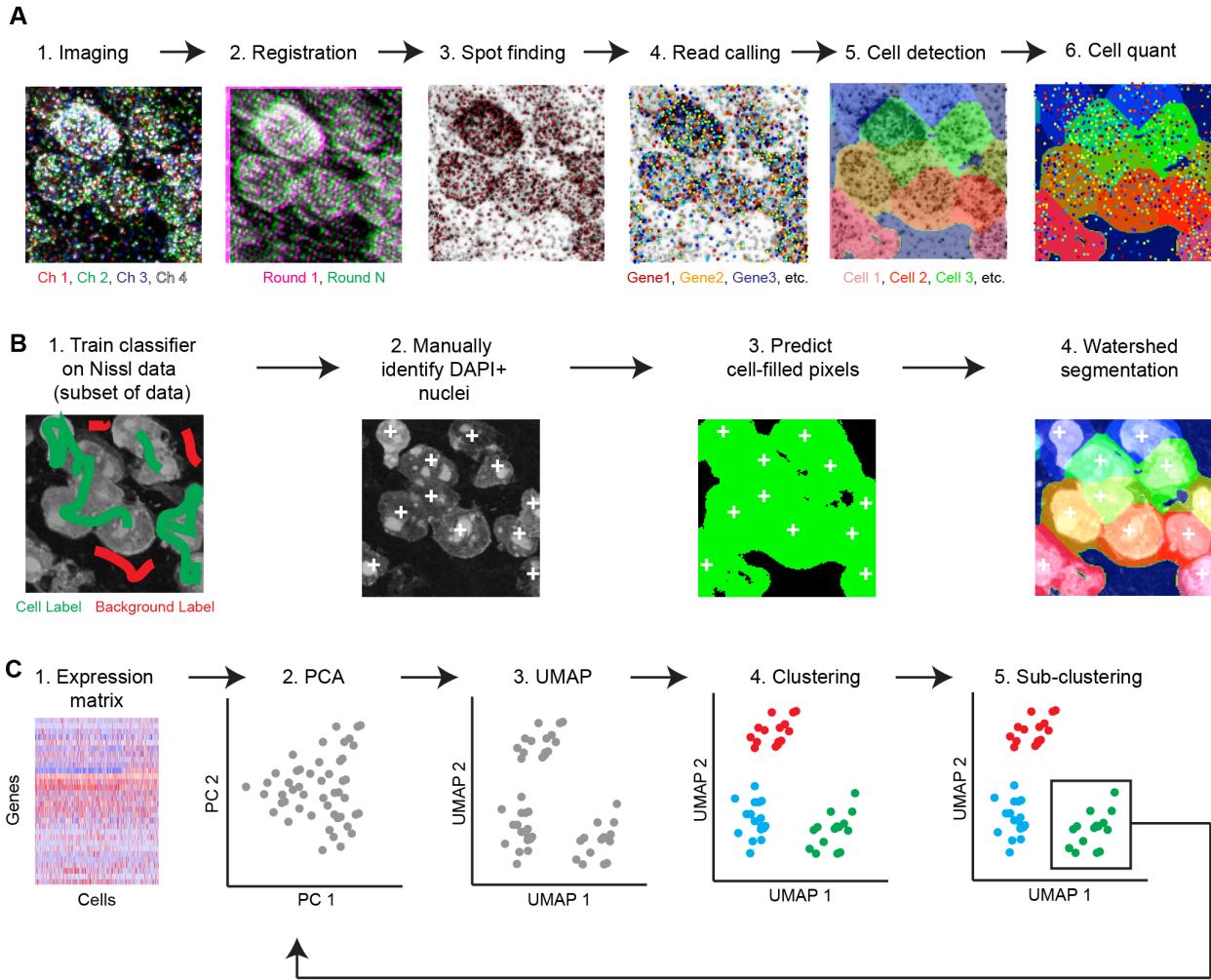
Fig. S4. Data processing pipeline for STARmap. (**A**) Diagram showing processing pipeline to extract decoded reads from raw imaging data (see also Methods for detail corresponding to each step). (1) Samples are imaged over multiple rounds. (2) Samples are registered across rounds, showing two rounds (green and purple) with a misalignment that must be corrected by registration. (3) Spots are automatically identified in each color channel (independently in the first round) as putative amplicons that will be decoded based on the color values at the point in each round. (4) Reads are called based on comparing the maximum intensity of each spot in each round across channels with the predicted color sequences for each barcoded DNA sequence (colorspace-encoded barcodes). (5) Cells are detected using machine learning-based segmentation that takes into account various intensity and texture features in order to segment Nissl containing cells from background (described in (B)). (6) Reads are assigned to cells by computing the overlap between each valid read's position and a convex hull of the segmented area for each cell. The gene encoded by a read that overlap with a cell's convex hull are assigned to that cell. (**B**) Method for determining cell extents. (1) A random forest classifier (a non-parametric machine learning algorithm for label prediction) is trained on a subsampled set of Nissl-stained data to discriminate cell-containing areas vs. background. (2) Cell locations are manually selected using the DAPI (cell nucleus) channel. (3) The classifier is applied to the whole image to predict the location of cells. (4) Cells are segmented from this

15

prediction using marker-based watershed, which segments the cell-labeled areas of the image into discrete cell bodies based on the known locations of nuclei. (**C**) Method for clustering and subclustering per-cell expression data. (1) Data are represented in a matrix of cells-by-genes, as z-scored log-transformed counts. (2) Principal components analysis (PCA) is applied to the matrix to reduce to a cells-by-factors matrix. (3) The location of cells is plotted using uniform manifold approximation for visualization (UMAP) (a nonlinear dimensionality reduction technique for the 2D visualization of high-dimensional data). (4) Cells are clustered by PCA values using shared-nearest-neighbor-based graph clustering. (5) The expression values of cells corresponding to individual clusters are then taken and used again for sub-clustering.
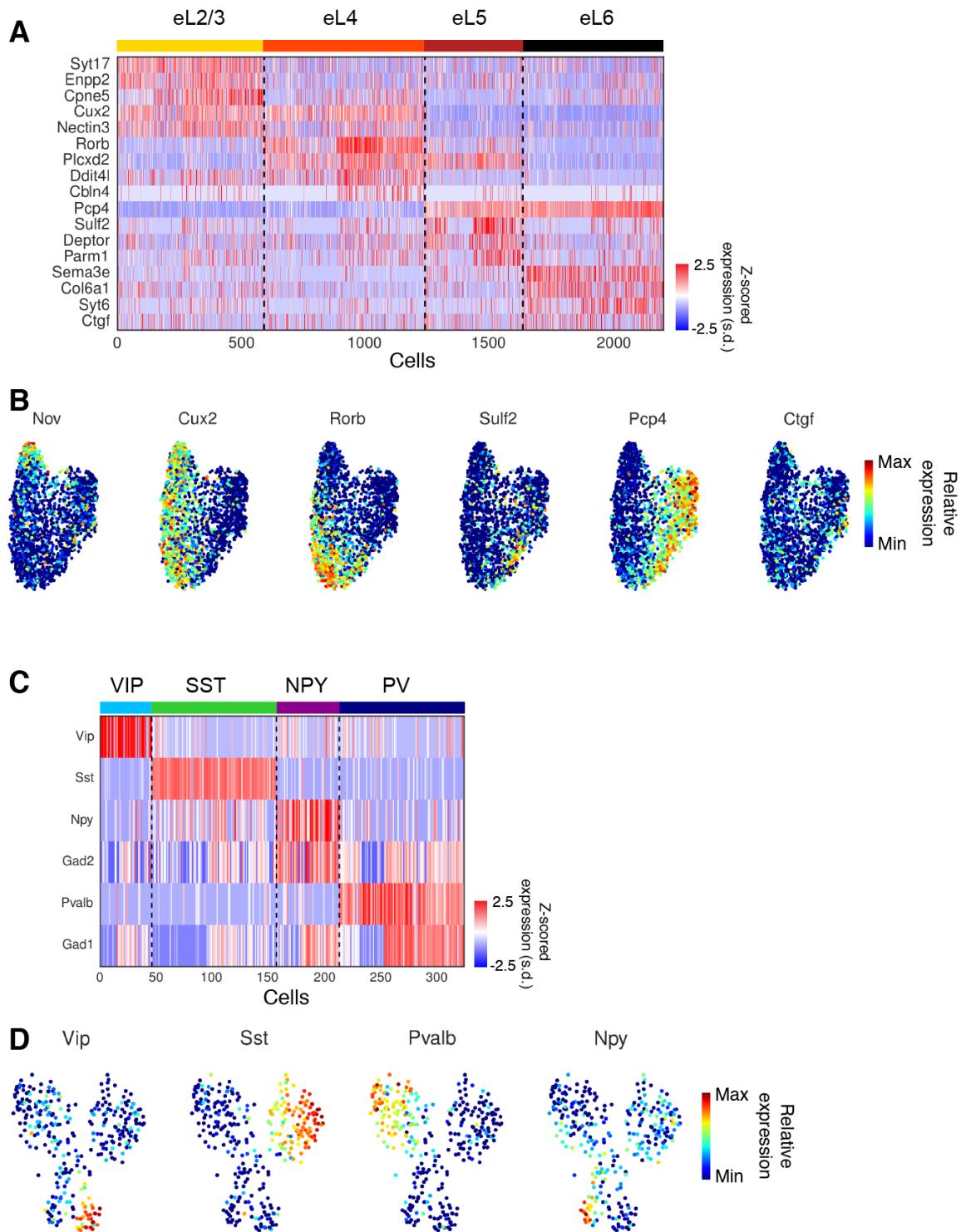
Fig. S5. Additional gene expression information for STARmapping of inhibitory and excitatory subclusters. (**A**) Z-scored expression matrix of excitatory cell types, showing clustering of multiple

differentially-expressed genes per cell type. Genes shown are selected based on a false discovery rate (FDR)-adjusted *P* value threshold of 10^{-10} and a minimum log10 fold change of 0.1, using a likelihood-ratio test, for genes that are expressed in cells within each cluster versus cells in any other cluster. (**B**) UMAP visualization of relative expression (normalized to min and max across all excitatory cells) of multiple known layer-specific genes enriched in each cluster across cells, showing that most are enriched in a specific excitatory subtype. (**C**) Expression matrix of inhibitory cell types, selected as in (**A**). (**D**) UMAP visualization showing relative expression of known interneuron marker genes, showing each is enriched specifically in an inhibitory neuronal subtype.
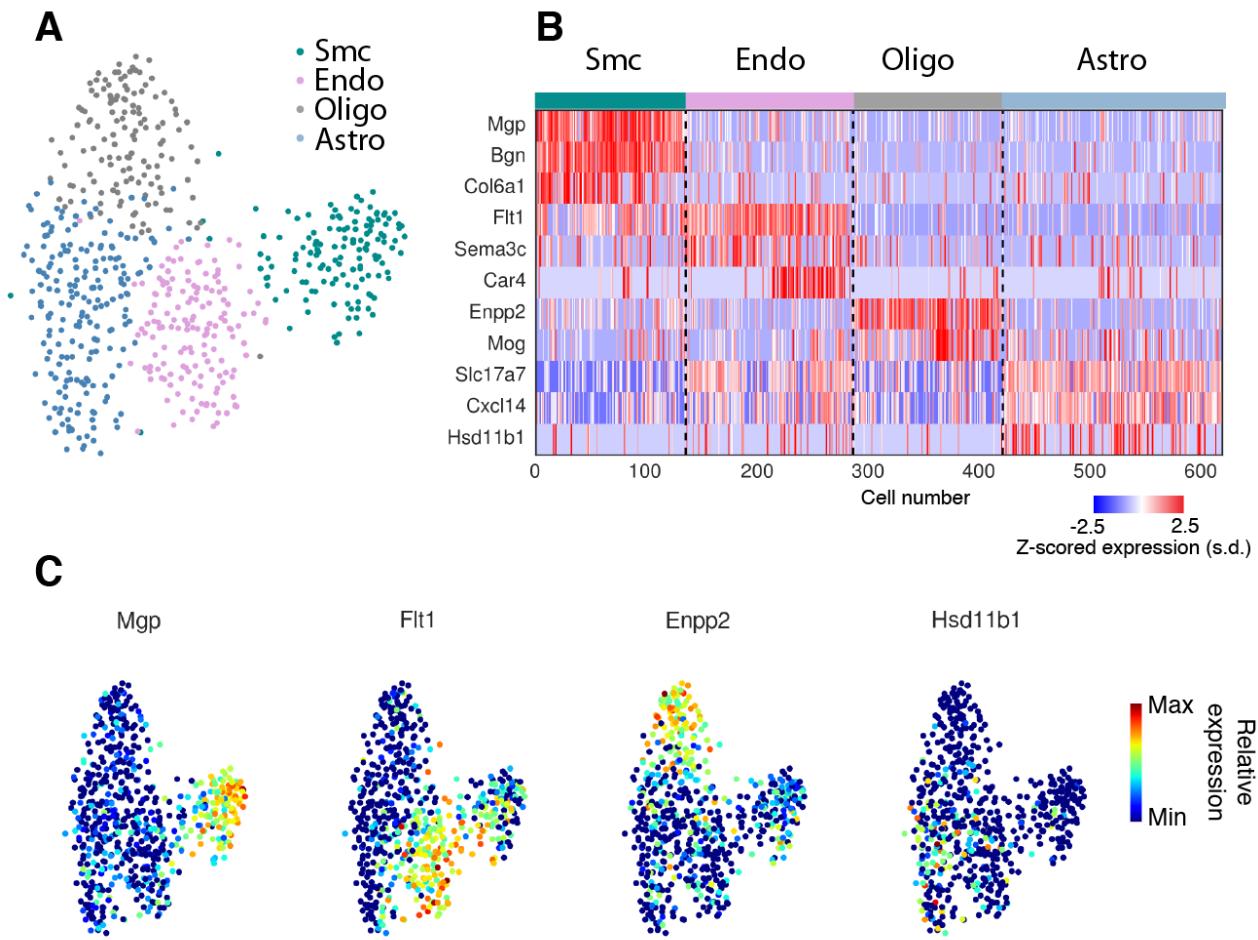
Fig. S6. Subclustering of non-neuronal cell types. (**A**) UMAP visualization of 4 non-neuronal cell types. (**B**) Z-scored expression matrix of non-neuronal cell types. Genes shown are selected based on a false discovery rate (FDR)-adjusted *P* value threshold of 10^{-10} and a minimum log10 fold change of 0.1, using a likelihood-ratio test, for genes that are expressed in cells within each cluster versus cells in any other cluster. (**C**) UMAP visualization of per-cell expression of marker genes (top differentially expressed genes per cluster) for non-neuronal cell types, showing specificity for that cluster. Color indicates relative expression (normalized to min and max) of each gene across all non-neuronal cells.
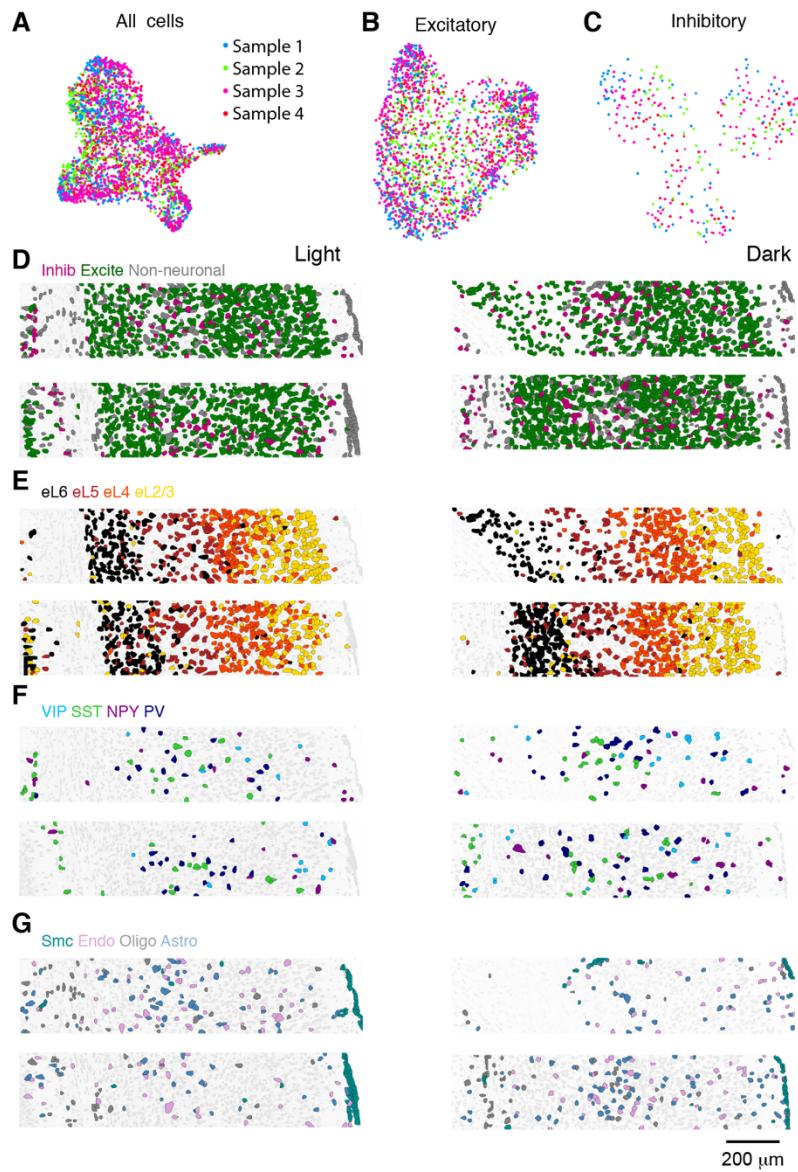
Fig. S7. Lack of batch effects with reproducibility of cell clustering in STARmap. (**A-C**) UMAP visualization of cells color coded by sample replicates, and then grouped by major clusters (**A**), excitatory subclusters (**B**), and inhibitory subclusters (**C**) (**D-G**) Spatial maps of light and dark replicate pairs (top and bottom) of all cell types (**D**), excitatory cell types (**E**), inhibitory cell types (**F**), and non-neuronal cell types (**G**).
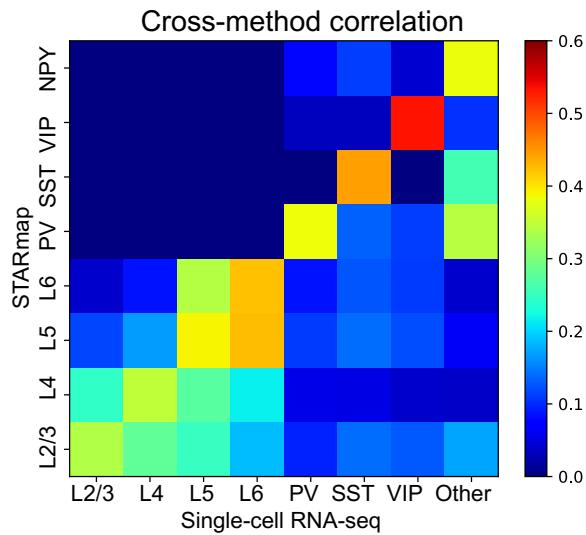
Fig. S8. Correlation of neuron types identified in STARmap 160-gene experiments and published single-cell RNA sequencing results. Pearson correlation of average gene expression across all genes within identified STARmap excitatory and inhibitory clusters, and corresponding clusters identified by single-cell RNA-seq from the Allen Brain Institute (*24*). For the single-cell RNA-seq data, the expression was averaged across all subtypes within a major type (e.g. L2/3), and only genes that were common between the single-cell RNA-seq and the 160-gene V1 experiment were used to compute the correlation. Scale, 0-0.6, Pearson correlation coefficients.
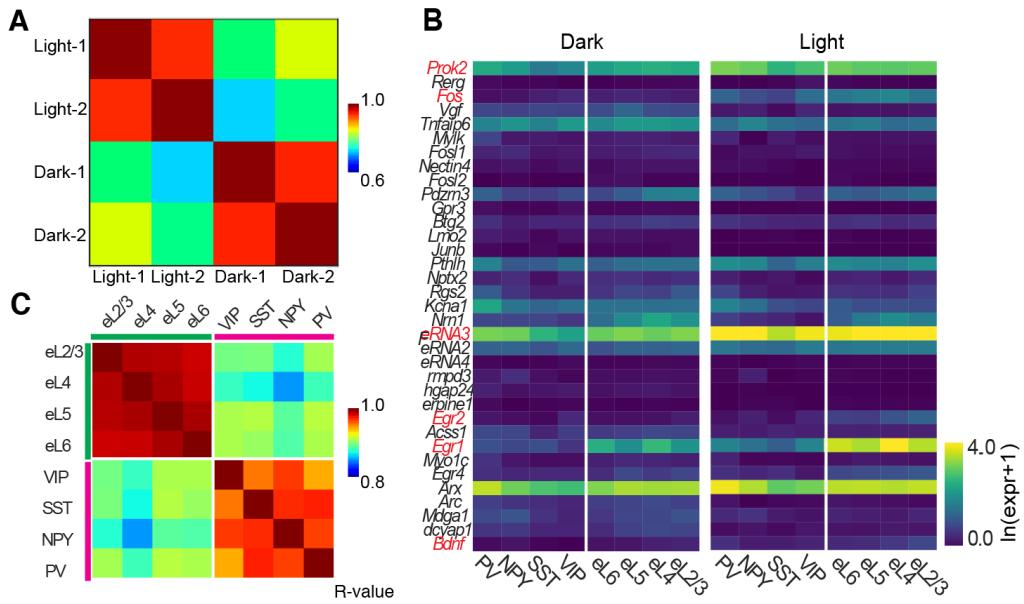
Fig. S9. Additional STARmap expression information of dark/light samples and activity-regulated genes. (**A**) Correlation of 160 genes of dark/light biological replicates showing that samples of the same condition are highly correlated than samples under different conditions; scale, Spearman R-value. (**B**) Log-scaled expression data (counts per cell) of ARGs in inhibitory and excitatory neuron subtypes. Genes with significantly increased expression in any cell type are highlighted in red. (**C**) Heatmap of the correlation of neuronal subtypes based on the correlation of mean expression of all ARGs in that cluster in response to light, showing that cells from inhibitory cell types are more correlated with other inhibitory cell types than with excitatory cell types, and vice versa. Note that scale ranges from R=0.8 to R=1.
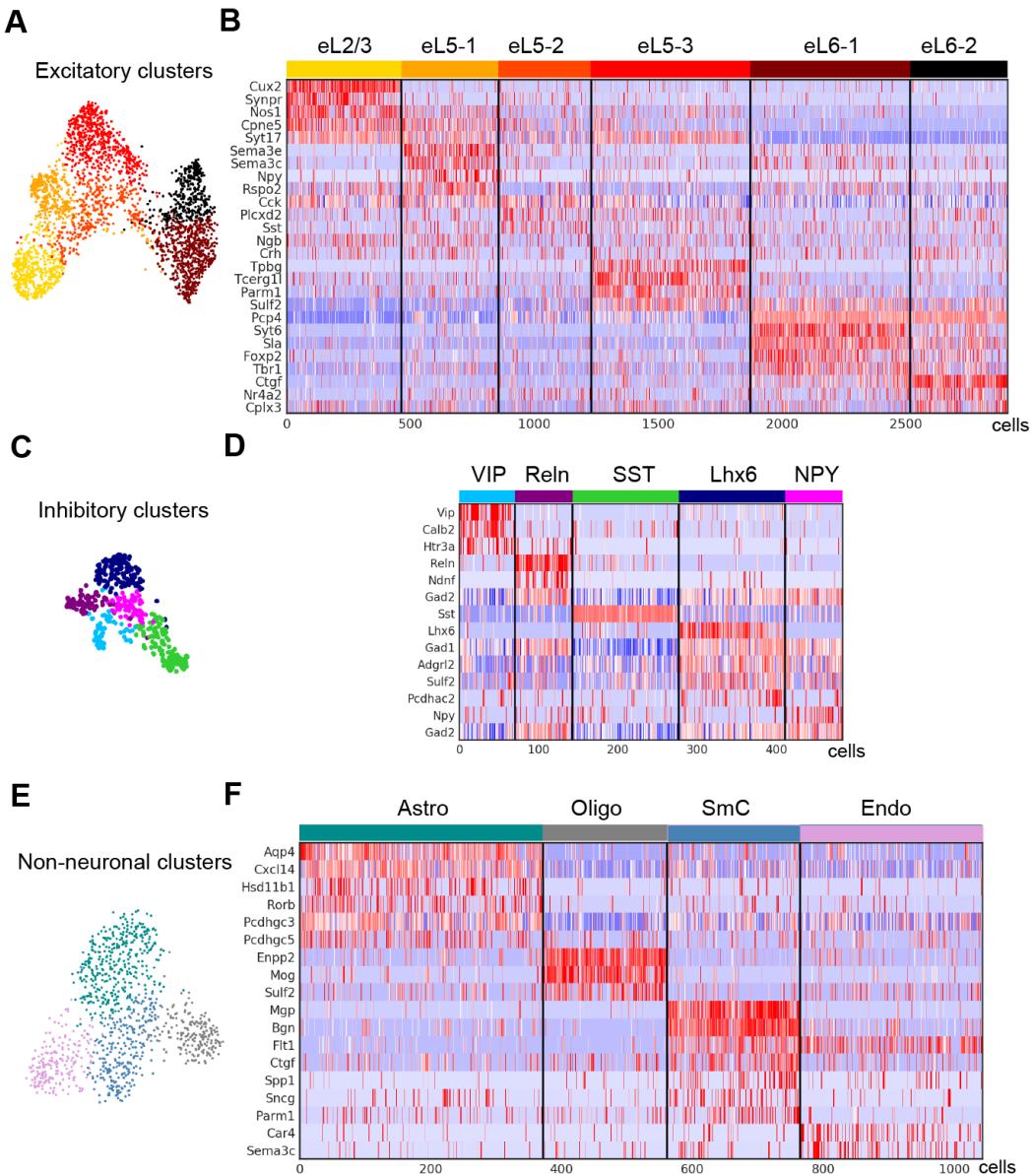
Fig. S10. Additional gene expression information for STARmapping of cell-type subclusters of medial prefrontal cortex (mPFC). (**A**) UMAP visualization of excitatory subclusters. (**B**) Differentially expressed genes per excitatory subcluster. (**C**) UMAP visualization of inhibitory subclusters. (**D**) Differentially expressed genes per inhibitory subcluster. (**E**) UMAP visualization of non-neuronal subclusters. (**F**) Differentially expressed genes per non-neuronal subcluster.
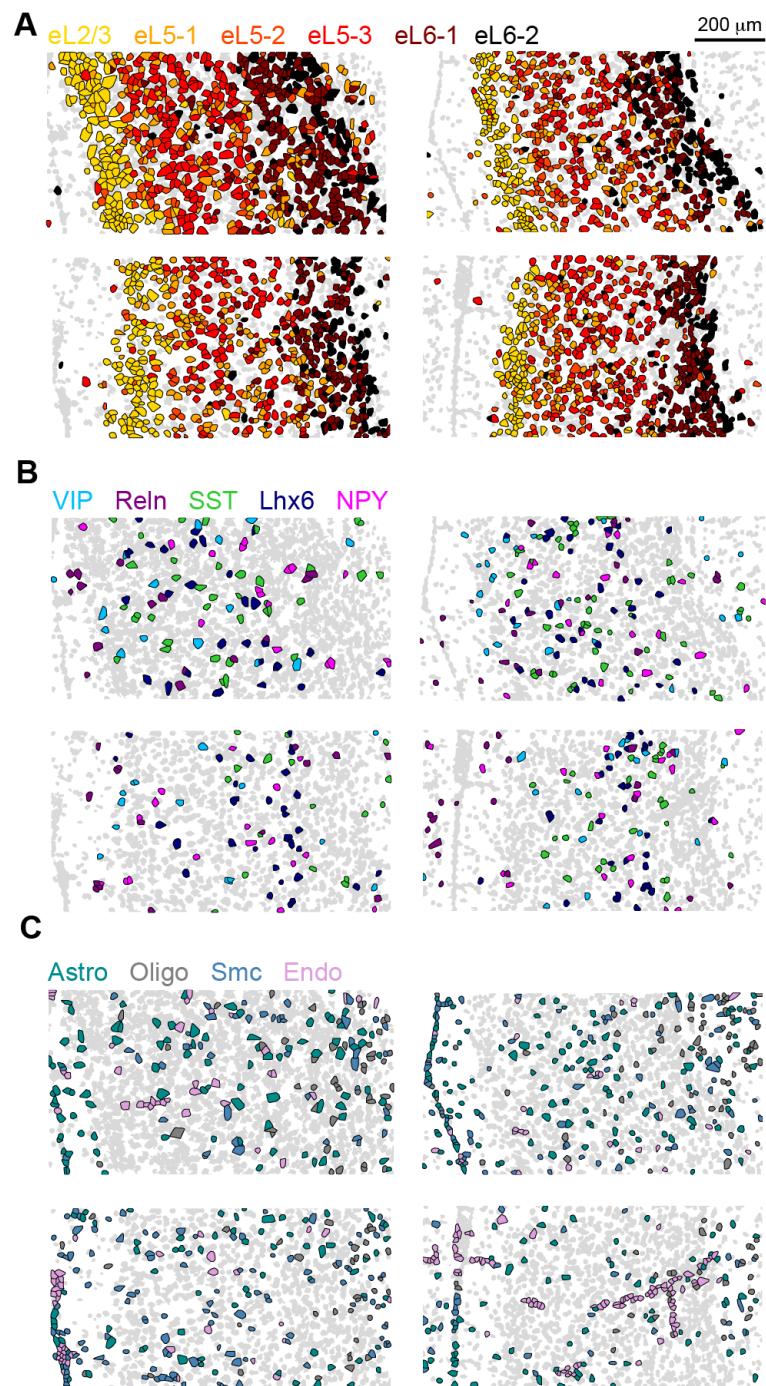
Fig. S11. Reproducibility of cell clusters and spatial organization of mPFC by STARmap. (A-C) Spatial maps across four biological replicates of excitatory subclusters (**A**), inhibitory subclusters (**B**), and non-neuronal subclusters (**C**).
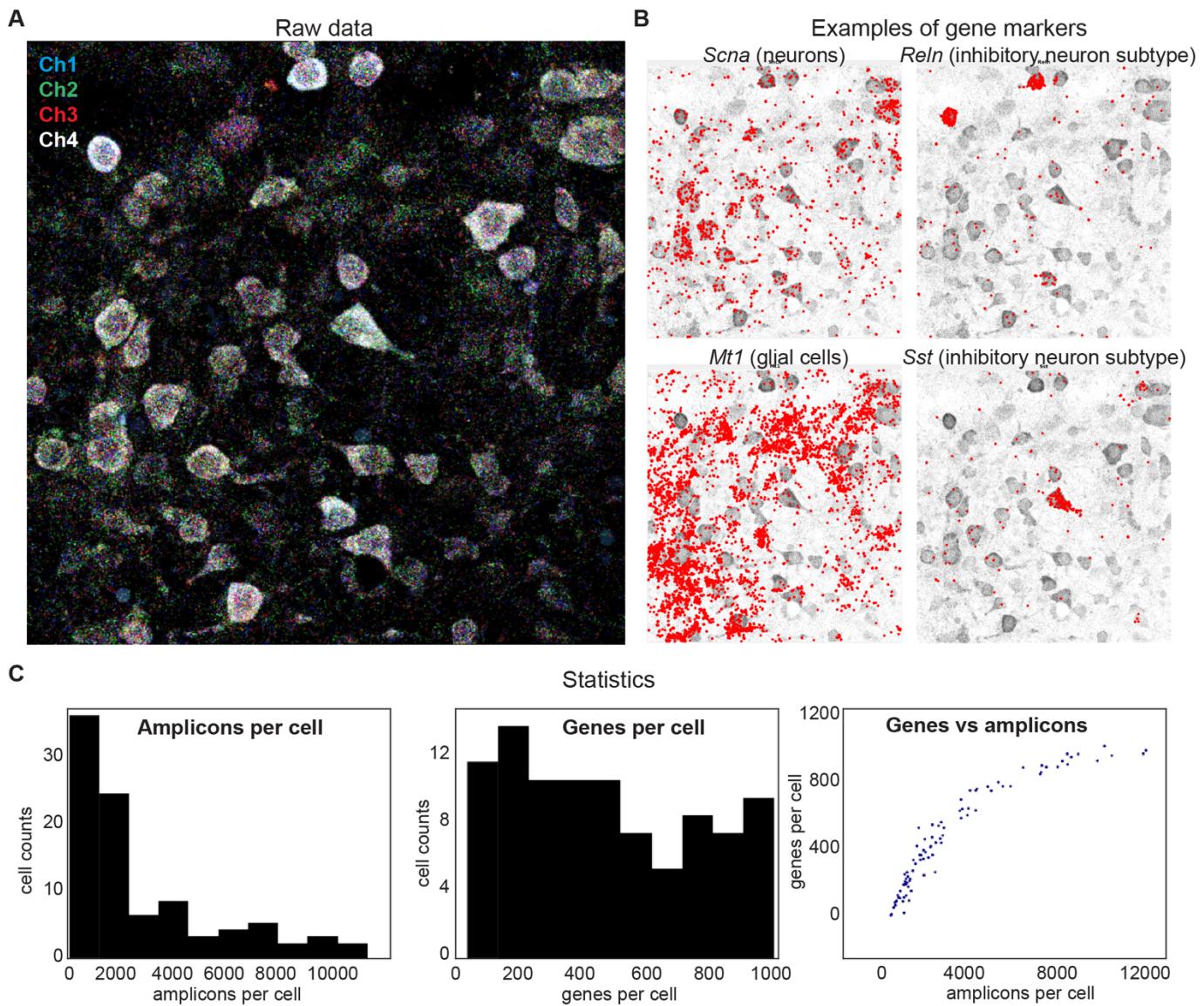
Fig. S12. STARmapping 1,020 genes in mouse hippocampal cell culture in 6-round sequencing. (A) Raw fluorescence image merging four fluorescent channels of the first round. **(B)** Examples of cell type markers. Neuronal gene marker (*Scna*) is well separated from non-neuronal gene marker (*Mt1)* and the distribution of neuronal subtype markers (*Reln*, *Sst*) are distinct. **(C)** Statistics of amplicons and genes per cell. Imaging area: 270×270 μm.
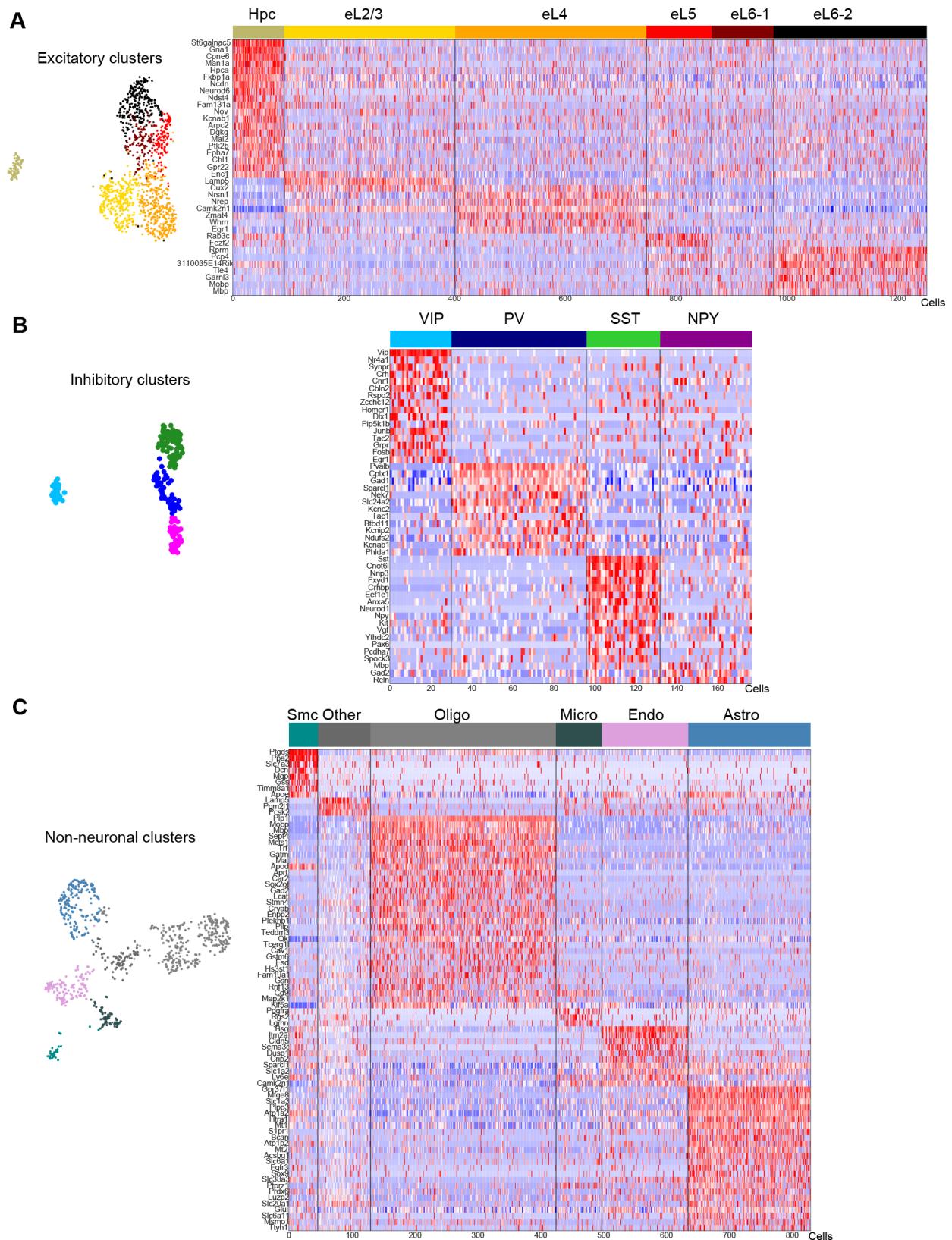
Fig. S13. Additional gene expression information for STARmapping of 1020 genes in mouse primary visual cortex. (**A**) Left: UMAP visualization of excitatory subclusters. Right: Differentially

expressed genes per excitatory subcluster. (**B**) Left: UMAP visualization of inhibitory subclusters. Right: Differentially expressed genes per inhibitory subcluster. (**C**) Left: UMAP visualization of non-neuronal subclusters. Right: Differentially expressed genes per non-neuronal subcluster.
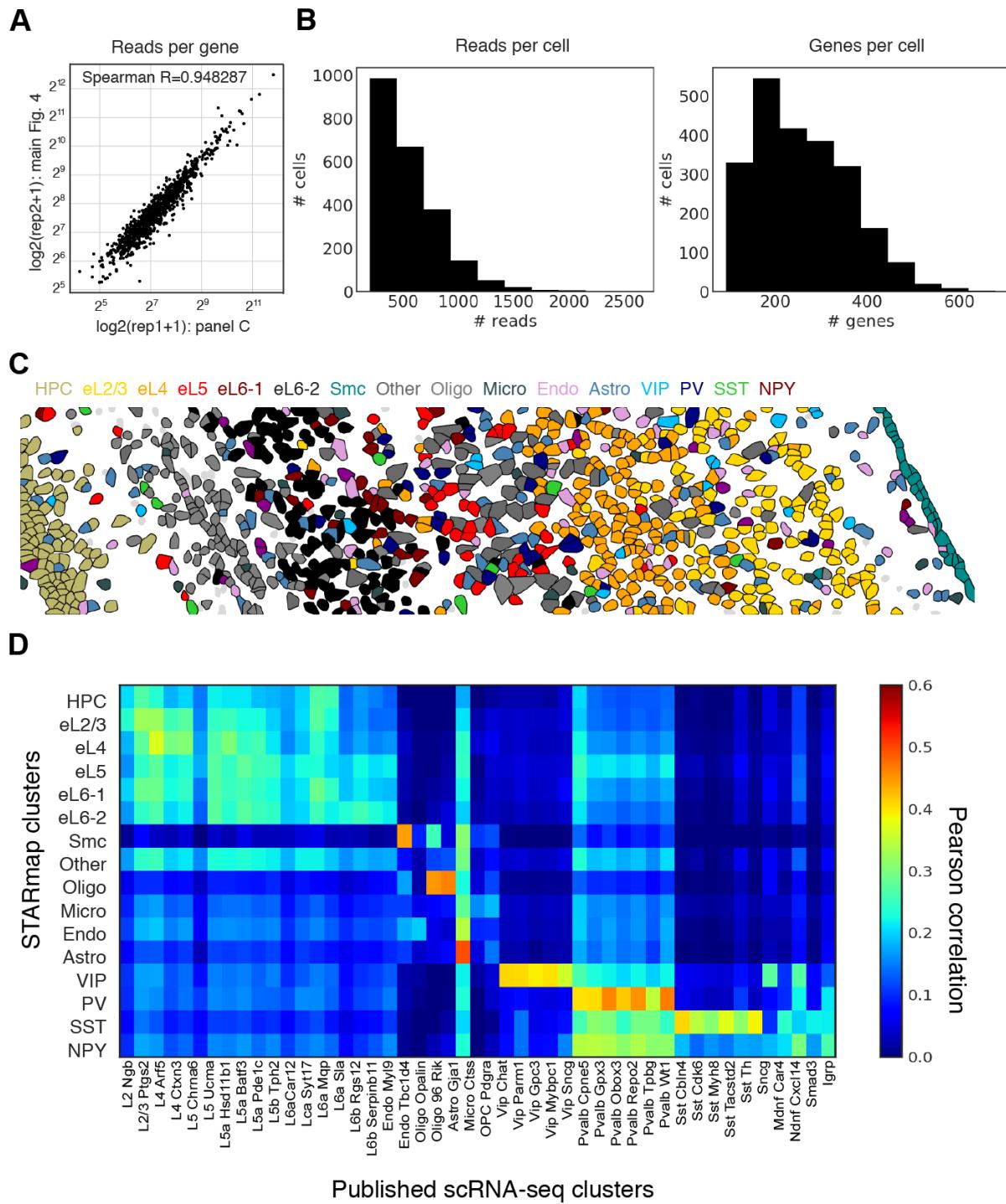
Fig. S14. Reproducibility and cross-method comparison of 1,020 genes measurements in mouse primary visual cortex by STARmap. (**A**) Correlation of reads per gene between two 1,020 gene replicates in visual cortex. (**B**) Histogram of detected reads per cell (left) and detected genes per cell (right). (**C**) Spatial map of cell types in other replicate of 1,020 gene visual cortex experiment. (**D**) Pearson correlation of average gene expression across all genes within identified STARmap 1,020 gene clusters, and corresponding clusters identified by single-cell RNA-seq from Allen Brain Institute (*24*).
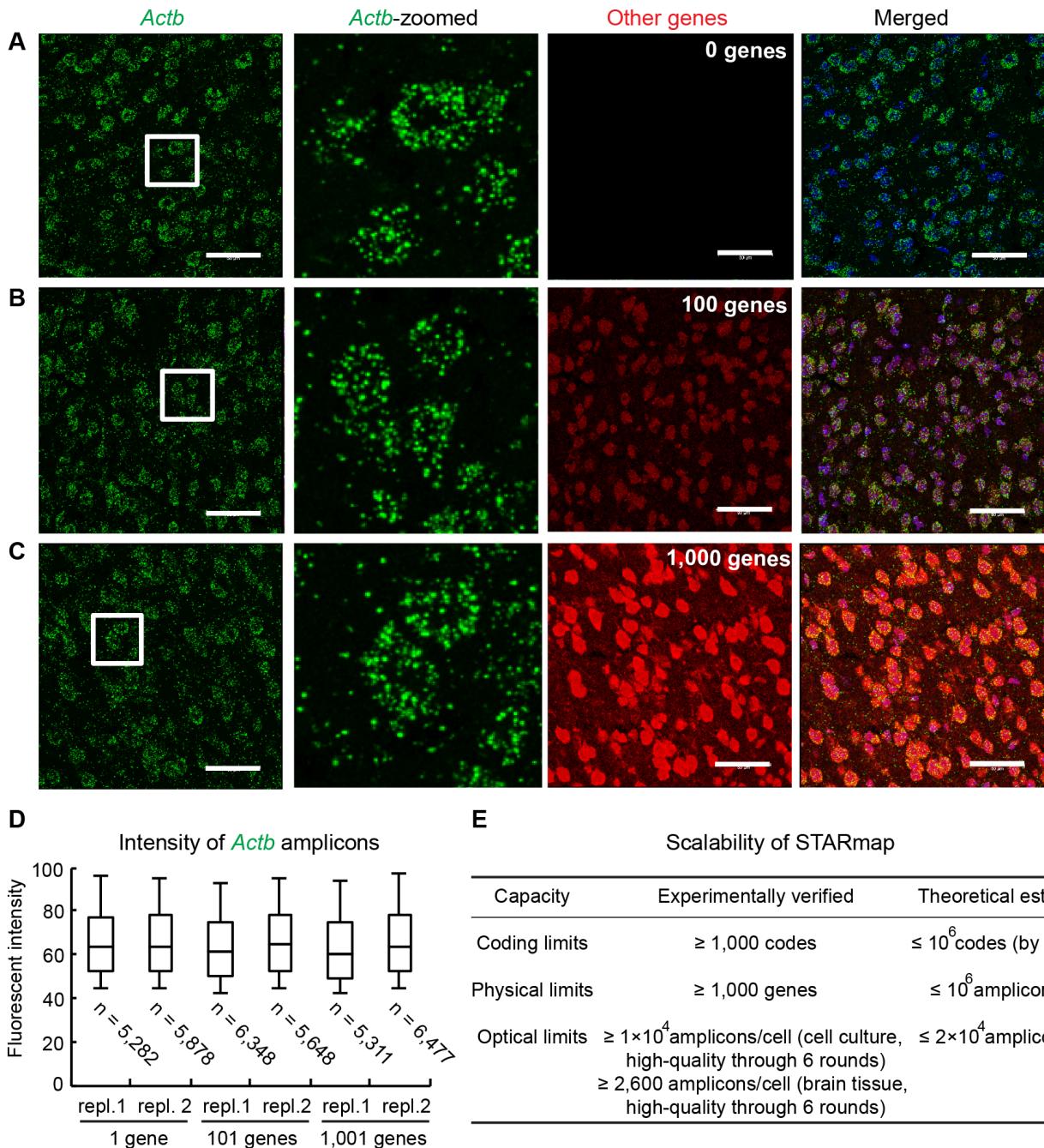
Fig. S15. Scalability of STARmap. (**A-C**) Detection of *Actb* mRNA when increasing amounts of other RNAs are co-detected, to test for potential dilution effect of probe mixing along with the physical capacity of cells for SNAIL DNA amplicons. The SNAIL probe of *Actb* was designed with an orthogonal DNA sequence for detection and was spiked into the mix with probes of 0 (**A**), 100 (**B**) or 1,000 (**C**) other genes. If the SNAIL probe were less efficient when working in a mixture than working as a single probe, or if there were not enough space for rolling circle amplification, the *Actb* spike-in would have resulted in fewer amplicons, and/or the intensity of each amplicon would be reduced. Fluorescence images were acquired in mouse visual cortex; green: Alexa546 channel of *Actb* amplicons; red: Alexa 647 channel of all other genes; blue: DAPI staining of cell nuclei. (**D**) Quantification of panel **A-C**. Box plots show that

any effects of dilution and cell space limitation is insignificant at least up to the scale of 1,000 genes. Box: first and third quartiles; middle line: median; whisker: 5% and 95% data points; n: number of *Actb* amplicons across the 228×228×2 μm imaging volume; y axis: absolute fluorescence intensity. (**E**) Experimental and theoretical estimation of STARmap scalability. Coding: The 5-nt code can encode 1,024 genes; the SNAIL probe has 35-nt coding space in addition to the RNA-complementary sequence; SEDAL requires 17-nt as a sequencing unit (11-nt docking region for reading probe plus 5-nt code and 1-nt flanking base) and thus the SNAIL probe can hold two such units and allow 4^{10} (10^6) codes; with other sequencing methods for longer reads (e.g. SOLiD, 18-nt for primer binding and 17-nt for coding), the upper limit will approach 10^{11}. Physical capacity has been verified here in mammalian neurons for up to 1,000 genes; since the physical size of the DNA amplicon is around 100-200 nm as determined by AFM and TEM (*7*), given that the diameter of a cell is around 15 μm and using a close-packed model (space efficiency 74%), estimated maximum capacity is 10^6 amplicons per cell. Optical volume has been validated with the 1,020-gene experiments in mouse hippocampal cell culture and visual cortex experiments; amplicons/cell refers to those that had been successfully registered through all 6 sequencing rounds. As imaged by confocal microscopy, the mean diameter of the DNA amplicons is 400-600 nm; applying the same model as used in physical limits, the maximum capacity is $2×10^4$ amplicons per cell. The experimental data for 1,020 genes approaches this bound. Any numerical differences between cell culture and tissue slices may be attributed to the following considerations: (1) whole cells were imaged in cell culture while cell fractions were imaged in tissue slices (8 μm, < 1 cell thickness); (2) the hippocampal cell culture was less-differentiated compared to adult mouse brain, thus exhibiting a larger diversity of RNAs (more genes) per cell; (3) in contrast to the dense 3D packing of cells in brain tissue, cells cultured *in vitro* spread out considerably in the *xy* plane and become thinner in the z direction, while images in the *xy* plane exhibit higher optical resolution compared with in the *z* direction (voxel size 78×78×250 μm).

A

Sample dimension | Thin tissue (z < 16 μm, cell monolayer) | Thicker tissue (z >100 μm, multiple cell layers)

Day 1 — Freshly frozen mouse brain → Cryostat slicing → PFA fixation

Day 1 — PFA-fixed mouse brain → Vibratome slicing

Library preparation

Permeablization → Hybridization

Day 1-2 — Permeablization & hybridization (modified probes)

Day 2 — Ligation & amplification → Hydrogel embedding & tissue clearing

Day 3 — Hydrogel embedding & tissue clearing

Day 4-5 — Ligation & amplification

Imaging & sequencing

Day 3-5 — Single-amplicon resolution; High NA oil-immersed objective; Imaging 200 cells per hour; SEDAL reaction with degenerate probes; Exponential readout from cycles to genes

Day 6-9 — Single-cell resolution; Low NA water-immersed objective; Imaging 10, 000 cells per hour; SEDAL reaction with orthogonal probes; Linear readout from cycles to genes

Data Output

3D amplicons 2D cell typing (main Figure 1-5)

3D cell typing (main Figure 6)

B

DNA probes   AA-NHS   modified DNA probes

C

Time vs. # genes

Tissue Type	Genes	Time
Thin tissue	4	2 days
Thin tissue	160	5 days
Thin tissue	1020	5 days
Thick tissue	1	5 days
Thick tissue	28	9 days

D

Comparisons of STARmap and single-cell PCR/RNA sequencing

| Method	RNA species	Spatial resolution	Quantification	#Cells
STARmap thin-tissue	≥ 40 nt	single-amplicon (250 nm)	absolute RNA copies	$10^2 \sim 10^3$
STARmap thick-tissue	≥ 40 nt	single-cell (1 μm)	relative intensities	$10^4 \sim 10^5$
single-cell PCR	≥ 100 nt	No	relative amount or absolute RNA copies	$10^1 \sim 10^2$
single-cell RNA sequencing	poly(A)+	No or 100 μm	absolute RNA copies	$10^2 \sim 10^5$

Fig. S16. Experimental flowcharts and cost estimates of STARmap for thin and thick tissue sections. (**A**) Experimental flowcharts of STARmap for thin and thick tissues. (**B**) Preparation of modified primer probes for large volume experiments: DNA probes were ordered with 5'amine modification, pooled, and converted to a polymerizable moiety by AA-NHS. (**C**) Experimental duration of different experimental design with various numbers of genes. (**D**) Comparison of RNA species, spatial resolution and throughput of STARmap with other single-cell approaches (*48*). Single-cell RNA sequencing may be combined with recently developed spatial transcriptomics methods to gain regional spatial resolution (100 μm; *49*).
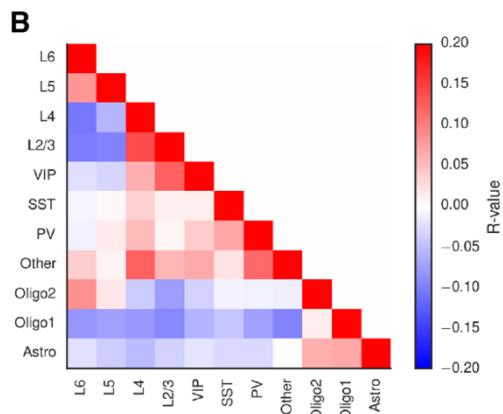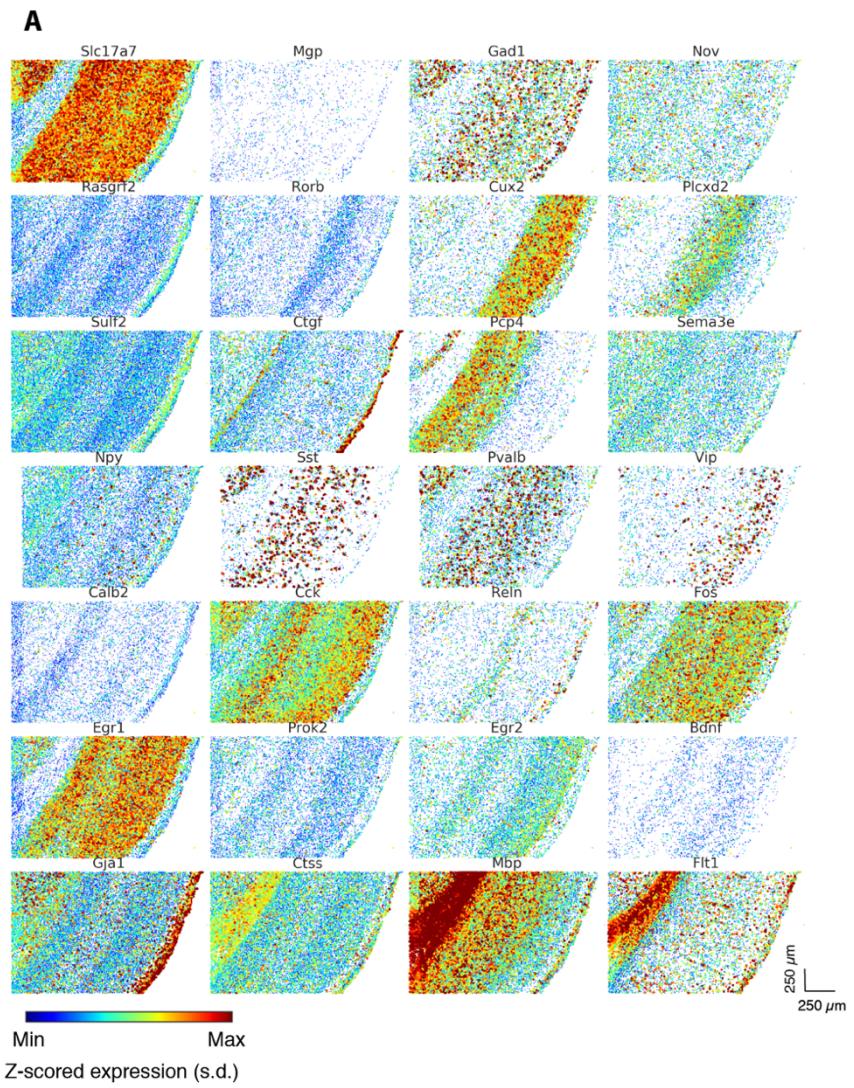
Fig. S17. Expression of multiple marker genes: sequential SEDAL readout. (**A**) Expression of each gene in 3D max projected in XY plane, showing spatial distributions of per-cell extracted gene expression values used for later clustering. Each gene z-scored per cell across all genes. (**B**) Voxel-wise correlation coefficient between distributions of each cell type, binned on 25 μm grid.

A Nearest neighbour distances in 2D: maximum projection of z stacks

z=24-32 (8 μm)

z=32-40 (8 μm)

z=64-72 (8 μm)

z=48-56 (8 μm)
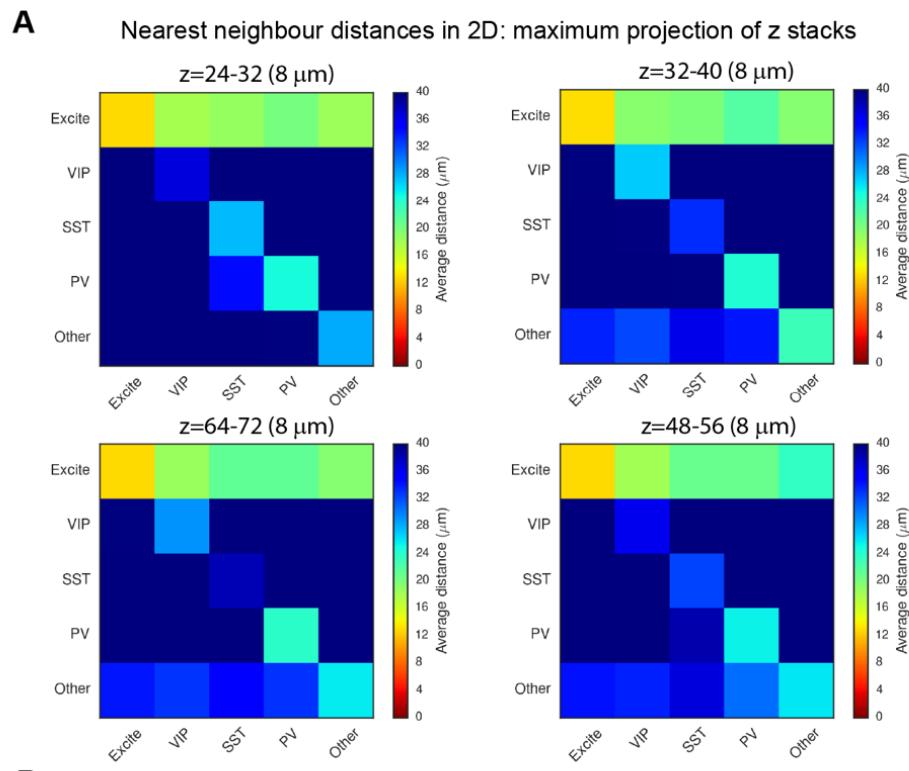
B Short-range clusters of inhibitory neurons revealed by STARmap
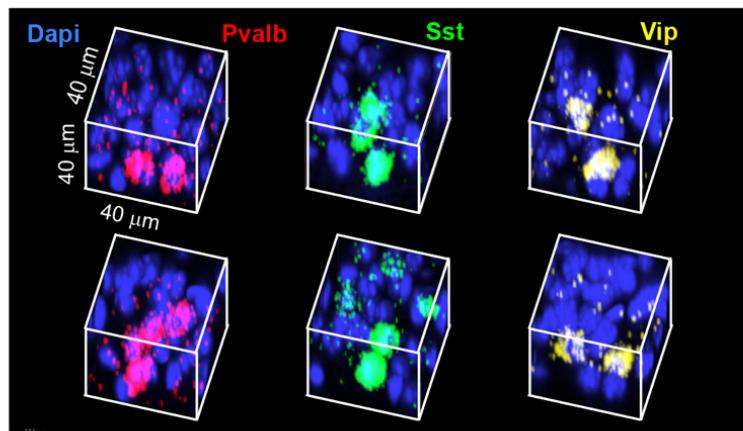
**C** Short-range clusters of Pvalb neurons revealed by transgenic labeling
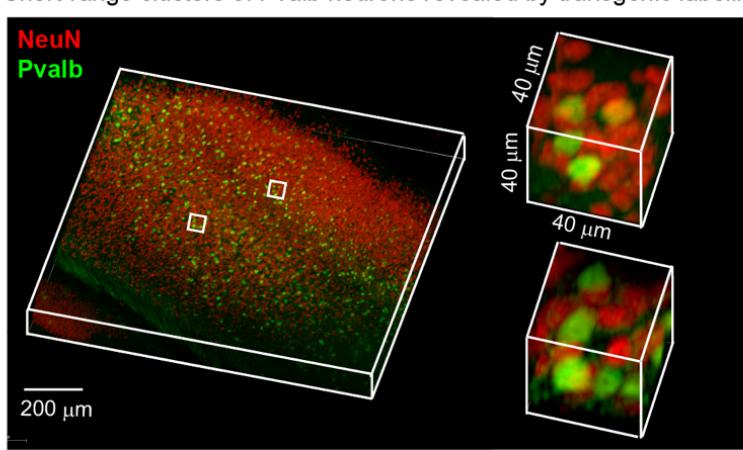
Fig. S18. 2D nearest neighbor analysis of short-range inhibitory clusters in mouse primary visual cortex and cross-method validation. (**A**) Average nearest neighbor distances between excitatory and different inhibitory cell types computed in the 2D projection of 8 μm (thinner than one cell) slices along *z* direction taken within the same 3D volume shown in Fig 6. The 2D nearest-neighbor distances cannot accurately estimate (overestimate) the 3D distances for the same cell types shown in Fig. 6G. (**B**) Examples of 3D short-range clusters of inhibitory neurons, zoomed view from Fig. 6C. (**C**) Short-range inhibitory neuron clusters were observed in the primary visual cortex of transgenic mice (generated by crossing Parv-IRES-Cre and Ai14): *Pvalb*-positive cells was by labeled tdtomato (green) and all neuronal nuclei were immunostained with Alexa 647-conjugated anti-NeuN (red).