Methods

CASS Barcode generation

The initial 20-base pair (bp) barcode candidates were generated through random combinations of the four nucleotides (A, T, C, G) in silico. All candidate sequences were then filtered using the online BLAST tool by comparing them against the RefSeq database. Sequences with a similarity score within the lowest 0.1% of all hits were selected. Subsequently, a second round of BLAST was performed on the chosen candidates to ensure that no pair of sequences shared more than seven identical base pairs, and to confirm the absence of significant similarity between any sequences.

This in silico selection process resulted in 46 candidate sequences. To further refine these, an in vitro selection was conducted. As outlined in panel 2 of Extended Data Fig. 1A, plasmids expressing the selected candidates were transfected into 293T cells, and corresponding probes were used for hybridization. Data from these experiments are also presented in Extended Data Fig. 2. Candidates exhibiting non-specific binding or weak fluorescence signals were discarded.

To assess the non-specificity of the probes with respect to endogenous RNA, additional hybridization experiments were performed on wild-type mouse brain sections. Only probes that failed to produce any signal were selected.

Following both the in silico and in vitro selection processes, 24 short sequences were identified and selected as the final elements of the CASS barcode.

Construction of barcoded viral plasmids

To generate the barcode plasmids, we employed commercial chemical gene synthesis. However, synthesizing the entire pool using this method would be prohibitively expensive. Therefore, we divided the CASS barcode into two parts: Part 1 (containing sequences ABCD) and Part 2 (containing sequences EFGH). Each part consisted of $3^4 = 81$ possible combinations, and 81 fragments from each part (162 total) were synthesized using commercial gene synthesis services (Sangon and Dynegene).

The plasmid for Part 1 was constructed by inserting it into the SAD-RV-dG-BFP vector (generated in our laboratory). Part 2 was amplified by PCR (Takara, R045A) and subsequently digested with restriction enzymes (NheI and SbfI, NEB, R3131 and R3642). The digested PCR product from Part

2 was then ligated into the plasmid vector of Part 1 via traditional restriction digestion and ligation (NEB, M0202).

Both parts were divided into nine equimolar groups, and 81 ligation reactions were performed, each containing one of the 81 potential products. The ligation products were introduced into competent E. coli cells (Accurate Biology, AG11804) through transformation. The transformed E. coli cells were cultured on 15 cm diameter plates (JET BIOFIL, TCD010150) containing low-salt LB medium (0.5% NaCl, 1% yeast extract, 1% tryptone, 2% agar) and 0.01% ampicillin (BBI Life Sciences , A610028-0025). Plasmids were then extracted from bacterial cultures using the plasmid extraction kit (Macherey-Nagel, 740426.5).

One-step rabies virus packaging and tittering

BHK-EnVA cells were cultured in DMEM (Yeasen, 41401ES76) supplemented with 10% FBS (Yeasen, 40131ES76) at 37℃ in a 5% CO_2 atmosphere. Cells were grown to 80% confluence prior to transfection.

The plasmid library of RV-BFP-CASS barcode was divided into nine groups, each containing 273 theoretical variants. Each transfection reaction included 45.6μg RV backbone, 5.8μg CAG-T7 polymerase, 9μg CAG-N, 4.7μg CAG-P, 4μg CAG-L, 2.5μg CAG-TVA, 4.8μg CAG-EnVA. The plasmid mixes were purified using isopropanol precipitation with an ethanol wash and then dissolved in electroporation buffer (120mM KCl, 0.15mM $CaCl_2$, 1.3Mm KH_2PO_4, 8.7mM K_2HPO_4, 25mM HEPES, 2mM EGTA, 5mM $MgCl_2$, 2mM Na_2ATP, 5mM Glutathione, filtered through a 0.23μm filter).

For electroporation, BHK-EnVA cells were trypsinized and resuspended in DMEM with 10% FBS. Cells were pelleted by centrifugation, then resuspended in ice-cold electroporation buffer. The plasmid mixture was added to the cell suspension, followed by additional electroporation buffer to reach a final volume of 100 μL. This mixture was transferred to an ice-cold electroporation cup (Bio-Rad, 1652088-1). Electroporation was performed using an electroporator (BEX, CUY21EDIT II) with the following settings: 250 V for 10 ms (ON) and 5 ms (OFF), followed by a 40 V square wave pulse for 50 ms repeated 10 times. Before electroporation, the resistance was verified to be approximately 200 Ω. After electroporation, the cells were placed on ice, and fresh DMEM with 10% FBS at 37℃ was added to resuspend the cells. The cells were then transferred to a 75 cm^2

culture flask (JET BIOFIL, TCF012250) and cultured at 37℃ in 5% CO_2 for 48 hours. The day of electroporation was designated as Day 0.

On Day 2, the culture medium was replaced. On Day 4, the medium was replaced again, and the cells were moved to a 34℃, 3% CO_2 incubator. By Day 5, BFP^+ cells were observable, and the medium was collected and stored at 4℃, with fresh DMEM added. Cells were trypsinized and transferred to a 15 cm diameter culture plate (JET BIOFIL, TCD010150) pretreated with poly-D-lysine (Beyotime, ST508). Medium was collected daily from Day 6 to Day 8 and stored at 4℃ with fresh medium added to the cells each day.

On Day 8, the collected medium from Days 5 to 8 for each transfection group was pooled, treated with 30 U/mL Benzonase (Servivebio, G3406-50KU), and incubated at 37℃ for 20 minutes. The supernatant was filtered through a 0.45 μm filter (JET BIOFIL, FPV403250) to remove debris.

For ultracentrifugation, the filtered supernatant was added to an ultracentrifuge tube (Beckman Coulter, 344058), followed by 5 mL of 20% sucrose, and the mixture was centrifuged at 4℃ at 20,000 rpm for 2 hours (Beckman Coulter, SW32Ti rotor). After centrifugation, the supernatant was discarded, and the resulting pellet was resuspended in 30 μL of PBS per tube, with gentle shaking at 4℃ overnight. The virus solution was then collected for titering, and the remaining solution was stored at 4℃. At this stage, nine separate virus libraries were obtained, each containing 273 variants in theory.

For virus titering, serial dilutions (10^{-1}, 10^{-2}, etc.) of the virus were added to the culture medium of 293T-TVA cells in a 24-well plate (JET BIOFIL, TCP011024). After 2 days, the dilution yielding approximately 100 sparse BFP^+ cells was selected. BFP^+ cells were counted in 10 randomly selected fields under a 10X objective, and the virus titer was calculated. The nine virus splits were then pooled in equimolar amounts, aliquoted, and stored at -80℃ for quality control and subsequent injections.

Virus quality control

Library construction

To extract RNA from virus samples, 4μL of virus, 200μL of TRIzol (Invirogen, 15596018CN), and 1μL of glycogen (Thermo Scientific, R0551) were mixed thoroughly by pipetting. After a 5-minute incubation, 40 μL of chloroform ($CHCl_3$) was added, and the mixture was centrifuged at 14,000 rpm

for 15 minutes at 4°C. The aqueous phase was carefully collected, and RNA was precipitated by adding 100μL of isopropanol. The solution was then centrifuged again at 14,000 rpm for 15 minutes at 4°C. Afterward, the RNA pellet was washed twice with 75% ethanol, followed by a final resuspension in 10μL of RNase-free water.

Reverse transcription was performed using the Maxima H Reverse Transcriptase Kit (Invitrogen, EP0753) according to the manufacturer's protocol. The reverse transcription primer, which included unique molecular identifiers (UMIs) with an eight-nucleotide random sequence (NNNNNNNN), was as follows:

CCTACACGACGCTCTTCCGATCTNNNNNNNNGTGGCCATTACGGCCGGCGCGC.

The reverse transcription product was then subjected to the first PCR amplification for NGS library construction using the PrimeSTAR Kit (Takara, R045A) with 25 cycles. The annealing temperature was set to 58°C. The forward primer used was CCTACACGACGCTCTTCCGATCT, and the reverse primer was

TCAGACGTGTGCTCTTCCGATCTCGACTGAAAAGCTAGTGCTAAGCGGCCGC. PCR products were purified using the SPRIselect reagent (Yeasen, 12601ES08) according to the manufacturer's instructions (0.7X, collecting DNA fragments >550 bp).

The purified products were then subjected to a second PCR amplification using the PrimeSTAR Kit (Takara, R045A) with 10 cycles. The annealing temperature was again 58°C. The forward primer used was

AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGC, and the reverse primer was

CAAGCAGAAGACGGCATACGAGATTGAATGTAGTGACTGGAGTTCAGACGTGTGCT.

After amplification, the products were purified using SPRIselect (0.5X + 0.2X, collecting DNA fragments within the range of 550–900 bp).

Finally, the purified NGS library was sent for next-generation sequencing (Mingmatechs, Illumina Nova6000 S4, index i7 = GAAACACA).

Data analysis

The raw ".fastq" sequencing data was utilized for virus quality control. All data analysis was performed using custom scripts written in MATLAB R2018b (MathWorks), and the corresponding

code is provided in the supplementary files (folder NGS).

The data were initially filtered to ensure that the Q30 score was greater than 90. The unique molecular identifiers (UMIs), as specified in the reverse transcription primer, were then extracted. Due to the distinct design of the CASS barcode (Extended Data Fig. 1b), the 5' 150 bp sequence captured the ABCD region, while the 3' 150 bp sequence covered the EFGH region. The barcode information was subsequently extracted, and the frequency of each barcode was calculated based on UMI counts (Fig. 2b, c).

For uniqueness estimation (Fig. 2d), barcodes were randomly selected from the pool in multiple iterations (particles). Each barcode was checked for uniqueness within the selected group, and the final result was based on the average of 1000 simulations.

For optimization of injection titer and volume (Fig. 2e), we simulated the random selection of barcodes from 1 to 1000, evaluating whether all the selected barcodes were unique. The probability of achieving uniqueness was calculated from 1000 simulation trials.

Mice and injection

All animal procedures were conducted in strict accordance with the approved protocols from the Animal Care and Use Committee of the Center for Excellence in Brain Science and Intelligence Technology, Institute of Neuroscience, Chinese Academy of Sciences, and the committee at Lingang Laboratory. The mice were housed in groups under standard laboratory conditions. *Neurod6*-cre mice were maintained as heterozygotes on a C57BL/6 background.

As outlined in Extended Data Fig. 1d, three mice were used in the connectome study of the visual cortex. The injection coordinates for each mouse were as follows: mouse #1, AP -1.99 mm, ML -3.22 mm, DV 0.40 mm; mouse #2, AP -1.75 mm, ML -2.95 mm, DV 0.40 mm; and mouse #3, AP -1.91 mm, ML -3.11 mm, DV 0.40 mm. The Bregma point was defined as AP 0 and ML 0, and the pial surface was defined as DV 0.

Sample pre-treatment

Three or four brain slices were processed at a time. All solutions used were RNase-free, and procedures were conducted in a clean, RNase-free environment, with most steps carried out under a laminar flow hood. The brain slices were first incubated separately in 4% PFA (Coolaber, SL1830)

in 24-well plates for 15 minutes, followed by three washes with PBS. The slices were then incubated in 8% SDS (dissolved in PBS) at room temperature with gentle shaking for 40 minutes, followed by three washes with PBS.

Subsequently, the slices were transferred to chambered glass slides (Cellvis, C1-1.5H-N), which had been pre-treated with 0.02 mg/ml Poly-D-lysine (Beyotime, ST508) overnight at room temperature. The slices were gently attached to the chambered glass using a clean brush in PBS. After aspirating the PBS, the chambered glass was quickly dried using a hair dryer. The sample (three or four slices on the chambered glass) was then incubated in 4% PFA with gentle shaking at room temperature for 10 minutes, followed by three washes with PBS.

Fluorescence signals for BFP, mCherry, and E2Crimson were captured using confocal microscopy (Olympus, FV3000) to identify the starter cells. Following imaging, the sample was incubated in Wash Buffer 1 (30% deionized formamide (BBI, A600211), 2 mM Ribonucleoside-Vanadyl Complexes (RVC, BBI, B644221) in 2X SSC solution) at 50℃ for 5 minutes. The wash buffer was aspirated, and the sample was then incubated with Hyb Buffer 1 (30% deionized formamide, 2 mM RVC, 0.2 mg/ml salmon sperm DNA (Sigma, D7656), 1 mg/ml yeast tRNA (Acmec, T48660), and 10% w/v dextran sulfate (Santa Cruz, sc-203917) in 2X SSC solution). A HybriSlip hybridization cover (Electron Microscope Sciences, 70329) was placed over the sample to prevent evaporation. The sample was incubated at 50°C for 4 hours, then allowed to cool to room temperature and stored overnight.

On the following day, the sample was washed in Wash Buffer 1 at 55℃, and the HybriSlip was removed. After aspirating Wash Buffer 1, the sample was incubated with fresh Wash Buffer 1 at 55℃ for 30 minutes, followed by three washes with 2X SSC at room temperature. The sample was then incubated in 4% PFA with gentle shaking at room temperature for 10 minutes, followed by three additional washes with 2X SSC. The sample was then placed in 2X SSC for pre-bleaching.

The pre-bleaching step was performed using an LED laser (FluoCa, FC904) with a 420 nm long-pass optical filter on a BX51 microscope. Using a 10X air objective, a region of interest was bleached for 15 minutes. Once all regions of interest had been bleached, the sample was ready for barcode FISH.

Barcode fluorescence in situ hybridization

The probes were chemically synthesized with a 5' fluorescent dye (Alexa488 for 488, Cy3 for 561, and Cy5 for 647) by a commercial primer synthesis service (Sangon). The probes were dissolved in RNase-free water to a 100μM stock concentration and stored at -80°C. For temporary use, the probes were diluted to 10μM and stored at 4°C. A mixture of three probes was prepared in Hyb Buffer 2 (5% deionized formamide, 2 mM RVC, and 10% w/v dextran sulfate in 2X SSC) at a final concentration of 10 nM for each probe.

For each cycle, the liquid was first aspirated, followed by the incubation of probes (approximately 500μL per chambered glass) at 37°C for 20 minutes. After incubation, the probes were aspirated, and the sample was washed three times with Wash Buffer 2 (5% deionized formamide in 0.2X SSC) at 37°C, with each wash step lasting 2 minutes. Finally, approximately 2 mL of Wash Buffer 2 was left to infiltrate the sample. Following washing, fluorescent probes hybridized to their corresponding RNA targets, and a Z-stack image was captured using an inverted confocal microscope (Olympus, FV3000) with a 20X air objective (NA 0.8). An additional bright-field image of the entire sample was captured during the first cycle to facilitate subsequent alignment with the Allen Brain Atlas.

The Z-axis was set to the middle of the sample, and a photobleaching step was performed using the "stimulation" module in FV3000 software with the same objective. In this step, the selected lasers (488, 561, and 647 nm) were turned on, while the GaAsP PMT detector was turned off. Typically, the entire imaging area was selected as the bleaching region; however, if the signal was only present in a part of the view, the bleaching region was defined as a region of interest (ROI). Photobleaching was carried out for 15 minutes per view (optical power is around 3.0mW/mm^2 for each channel at the objective lens), which was sufficient to quench the fluorescence of the probes. After photobleaching, the next cycle was initiated.

For manual 8-cycle FISH, the sample may be removed from the microscope, so a manual check of the view is necessary between cycles. For automated FISH, the sample remains in place until the completion of all 8 cycles, with only micron-level shifts occurring. These shifts can be corrected during the subsequent data analysis process.

Decoding accuracy is defined as the fraction of cells that are fully decoded in each cycle. The accuracy in each cycle remains constant (Extended Data Fig. 4d), suggesting that there was no interference of preexisting probes on newly added ones.

Immunofluorescence labelling

Immunofluorescence labeling was applied to samples after the completion of the 8-cycle FISH procedure. Initially, the samples were washed three times with PBS, with each wash lasting 10 minutes. The samples were then incubated in blocking buffer (5% BSA (Sigma, V900933), 0.5% Triton X-100 (Aladdin, T109027) in PBS) for 2 hours at room temperature with gentle shaking. Following blocking, the samples were incubated overnight at 4°C with rabbit anti-somatostatin antibody (Invitrogen, PA585759) at a 1:1500 dilution in blocking buffer, with gentle shaking.

On the following day, the samples were rinsed three times with PBS, with each rinse lasting 10 minutes. Subsequently, the samples were incubated with donkey anti-rabbit secondary antibody conjugated to a 647 nm fluorescence dye (A-31573) for 4 hours at room temperature, with gentle shaking. Afterward, the samples were washed three times with PBS, each wash lasting 10 minutes, prior to imaging.

The images were collected using an inverted confocal microscope (Olympus, FV3000). The same imaging region for FISH was manually selected based on BFP signals. Somatostatin-positive cells were identified manually in the resulting images.

Barcode decoding

Coordinate Alignment with Allen Brain Atlas

Each brain slice was aligned to the Allen Brain Atlas (Allen CCF v3)[1] using the QuickNII[2] software (NITRC). The bright-field image, captured during the first cycle of each sample, was manually aligned to the standard mouse brain template. QuickNII provides a linear coordinate system, allowing for the calculation of a transformation matrix that enables the mapping of all pixels to the standard brain coordinate system for each slice.

The primary source of potential alignment error occurs during this step. We estimated that the maximum positional error for each cell was approximately 100μm, particularly along the anterior-posterior axis, due to the limitations of the coordinate transformation method.

Cellpose3-based cell segmentation

Cycle 5 was selected as the template, with the 405 nm channel (BFP) of the cycle 5 image chosen for cell mask generation. A 3D cell mask was generated using Cellpose3[3] with a pre-trained model, achieving approximately 80% accuracy. Manual verification was performed to refine and output the

cell masks for each image.

Shift correction

Micron-level XYZ shifts in the sample during the 8-cycle FISH were corrected. As described earlier, cycle 5 was used as the template. The MATLAB function "imregtform" was employed to estimate and correct these shifts. Following this correction, the cell mask generated from cycle 5 was applied consistently across all cycles for decoding.

Decoding score and reliability

Initially, a small region devoid of fluorescence was designated as the background ROI. Pixel values for all 3D cell ROIs (including the background ROI) were calculated for each channel and cycle. The decoding score was defined as the ratio of the mean intensity of the cell ROI to the standard error of the background ROI. Statistical significance was determined using a Student's t-test with FDR correction (Benjamini-Hochberg method). A cycle was flagged as a "missing cycle" for a cell ROI if no channel in that cycle exhibited significant differences from the background (Extended Data Fig. 4c, d). Only ROIs with significant differences were used for outputting decoding results and reliability assessments.

For decoding, a single channel (e.g., a1) was identified as the decoding result if its mean intensity exceeded the sum of the other two channels (e.g., a2 and a3) and showed a significant difference from the background. If no channel exhibited a dominant intensity, the cell was classified as containing "multiple barcodes."

To guide manual verification, a reliability value of 0 or 1 was assigned to each cell ROI in each cycle. 1 indicated no need for manual checking and required strict criteria: decoding score for channel a1 must satisfy the following conditions:

1. $\mathrm{a1} > 2 \times (a2 + a3)$
2. $\mathrm{a1} > 2$
3. $\mathrm{a2} < 1$
4. $\mathrm{a3} < 1$

Approximately 60% of ROI*cycles passed this stringent criterion, with the remainder subjected to manual review. Cycles with a reliability score of 1 were also verified, and no mis-decoding errors were identified.

Location of cell

The XYZ coordinates (image location) of each cell were determined using the 3D ROI "centroid" property via the MATLAB function "regionprops3". The corresponding image region (a square) within the entire coronal section was manually identified. Each cell's location was then mapped to the coronal section (coronal section location) based on the image region. Subsequently, the coronal section location was transformed into the Allen Brain Atlas coordinate system using the transformation matrix described previously. A 3D model was generated using the BrainMesh[4] app in MATLAB.

Connectivity detection

After decoding, the location and barcode information for each cell were obtained. Starter cells were manually identified based on marker expression: cells expressing RV, TVA, and RVG were classified as starters; cells expressing RV and TVA were classified as "TVA only"; and cells expressing RV and RVG were classified as "G only". These starter cells were then linked to their respective barcode and location information. Unique barcode identification was performed within each experimental set. All RVG-expressing cells (including both starter and "G only" cells) were considered for unique barcode identification. Barcodes expressed in only one RVG-expressing cell were classified as "unique". Starter cells were subsequently divided into three groups: those with multiple barcodes, unique barcodes, and non-unique barcodes (Extended Data Fig. 4b).

Next, input cells were analyzed for the expression of unique barcodes. This included input cells with one or two missing cycles, which were accounted for using wildcard matching (fault-tolerant algorithms). For instance, if an input cell expressed the barcode 01111112 (with 0 indicating a missing cycle) and a starter cell expressed 31111112 while no RVG-expressing cells expressed 111111112 or 21111112, the input cell was matched to the starter cell, and the corresponding barcode was identified as unique or non-unique. Fully decoded cells were further matched against the pool of unique starter barcodes. Input cells were categorized into four groups: matched with starter cells expressing unique barcodes, matched with starter cells expressing non-unique barcodes, unmatched (unable to find corresponding starter cells), and those with multiple barcodes.

In Extended Data Fig. 4e, unmatched input cells were further subdivided based on decoding completeness. Cells that could not be fully decoded due to missing cycles were labeled as "not fully decoded". The information for all connection-detected cells is provided in the supplementary file,

T_connection_3mice.xlsx.

Connectivity data processing

Spatial distribution analysis

The anterior-posterior (AP) distance for each input-starter cell pair was calculated by directly subtracting the x-axis coordinate values of the two cells.

For lateral-medial (LM) distance calculations, a perpendicular line was drawn from each starter cell to the pial surface. The LM distance was defined as the shortest distance from the input cell to this perpendicular line. Additionally, the length of this line, extending from the starter cell to the pial surface, was recorded as the depth.

The territory of each starter cell was determined as the smallest ellipsoid encompassing 90% of the input cells, with the centroid of the ellipsoid aligned with the location of the starter cell.

Regional connectivity analysis

All connected starter and input cells were automatically annotated to the Allen Brain Atlas based on their spatial coordinates. To ensure consistency, several subregions were manually merged (e.g., LGd-co, LGd-ip, and LGd-sh were combined into dLGN). The connections between these regional pairs were quantified (Extended Data Fig. 6 and 7a).

Randomized shuffle data were generated by randomly pairing existing starter cells with input cells, with 1,000 iterations of shuffling performed. For each regional pair, the statistical difference was calculated by comparing the observed connection value to the distribution of the 1,000 shuffled values using a Z-test. To account for multiple comparisons across all input-starter pairs, p-values were adjusted using the False Discovery Rate (FDR) correction with the Benjamini-Hochberg method.

Co-innervation analysis

When two input cells projected to the same region, they were considered a co-innervation pair. These pairs were counted and categorized according to the region of the input cells (Extended Data Fig. 8a).

To assess the statistical significance of these co-innervation pairs, randomized shuffle data were generated under the same conditions: existing starter cells were randomly paired with input cells. For each input-input pair, the statistical difference was calculated by comparing the observed co-

innervation counts to the distribution of the 1,000 shuffled counts using a Z-test. P-values were adjusted using FDR correction with the Benjamini-Hochberg method.

Automated FISH system

The system was designed and set up in our laboratory. Key components include a metal sand bath (DLAB, HB120S), 24V peristaltic pumps and controller (Runze Fluid, MC20A), electric linear actuators (24V, 4mm/s, 50N, with a 50mm work distance), and syringes for the micro-fluid injection system. The system is controlled by a microcontroller (Arduino, Mega2560) and a real-time clock module (DS3231) for precise timing. A four-digit digital tube display (TM1637) provides additional information for debugging. The custom-designed PCB board (design files provided in supplementary materials, manufactured by Jeipei) controls the system, and an electromagnetic valve (Wokun Technology, WK07-308-3/4-NO) regulates fluid flow. The syringe brackets were 3D printed using #8200 plastic (Wenext). The controller is housed in a custom-designed stainless-steel enclosure (SUS304). All components are mounted on a 33U removable cabinet for easy access and maintenance.

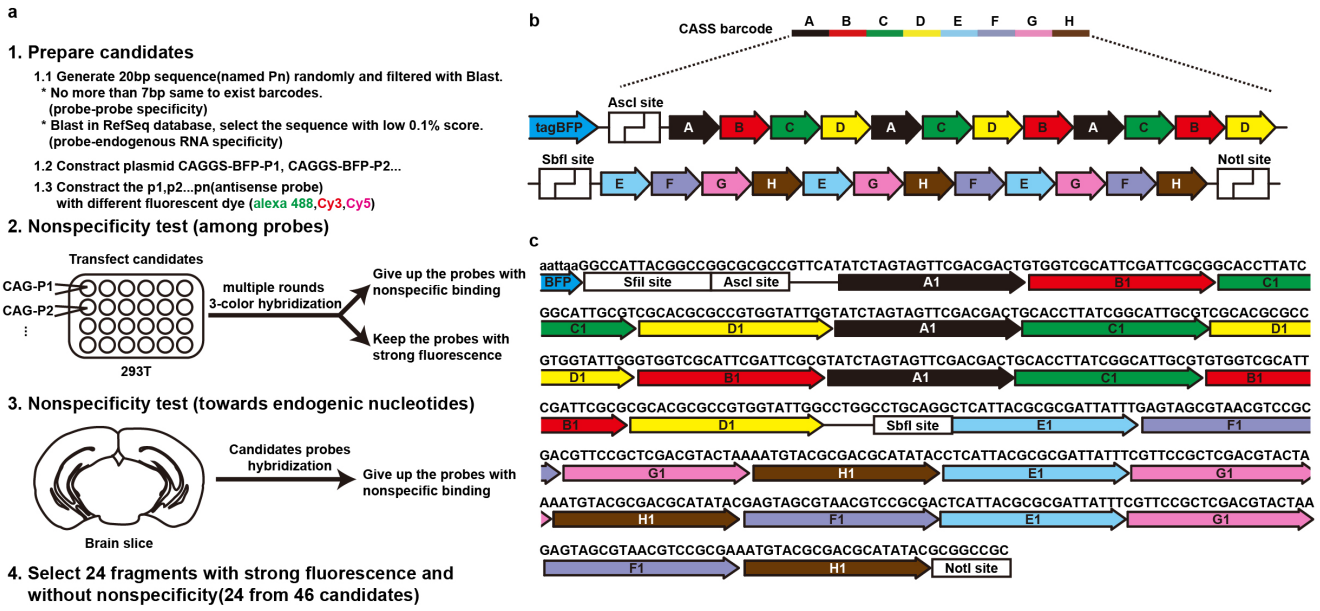
1 Wang, Q. et al. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. Cell **181**, 936-953 e920, doi:10.1016/j.cell.2020.04.007 (2020).

2 Puchades, M. A., Csucs, G., Ledergerber, D., Leergaard, T. B. & Bjaalie, J. G. Spatial registration of serial microscopic brain images to three-dimensional reference atlases with the QuickNII tool. PLoS One **14**, e0216796, doi:10.1371/journal.pone.0216796 (2019).

3 Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular segmentation. 2024.2002.2010.579780, doi:10.1101/2024.02.10.579780 %J bioRxiv (2024).

4 Hao, Y. BrainMesh: A Matlab GUI for rendering 3D mouse brain structures, < https://github.com/Yaoyao-Hao/BrainMesh/> (2020).

357 **Supplementary figures:**



d

mouse id	gender	birth date	genotype	AAV9-EF1a-DIO-mCherry-F2A-TVA-WPRE	AAV9-hSyb-DIO-oRVG-p2A-E2Crimson-WPRE	helper mix injection volume	helper injection date	RV titer	RV volume	RV injection date
#1	M	2024/1/31	*Neurod6+/+*	5E10	1E12	50nl	2024/5/2	5E6	50nl	2024/5/16
#2	M	2023/11/18	*Neurod6+/-*	5E10	1E12	50nl	2024/6/11	5E6	50nl	2024/6/25
#3	M	2024/4/20	*Neurod6+/+*	1E10	5E11	30nl	2024/7/30	5E6	50nl	2024/8/21

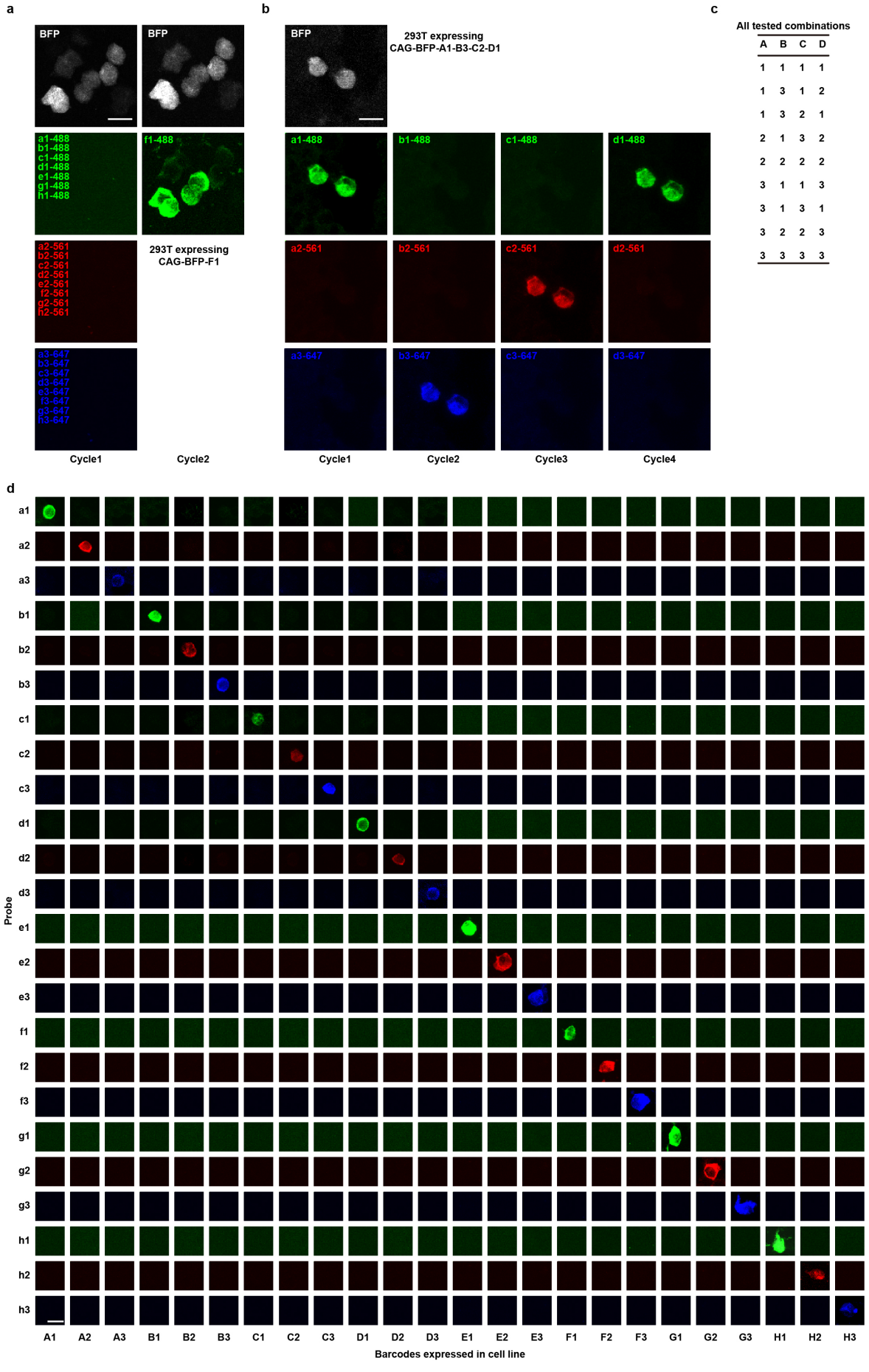
Extended Data Fig. 1 Details of CASS Barcode Design and Experimental Setup

a. Process for generating 24 short artificial nucleotide sequences.

b. Full structure of a single barcode located in the 3' UTR region of BFP. Each short sequence appears in triplicate, arranged to minimize potential secondary structures. The middle SbfI site facilitates plasmid pool construction.

c. Sequence and annotation of a representative barcode (A1B1C1D1E1F1G1H1).

d. Details of mouse experiments and virus injection information.


a
BFP
BFP
a1-488
b1-488
c1-488
d1-488
e1-488
g1-488
h1-488
f1-488
293T expressing
CAG-BFP-F1
a2-561
b2-561
c2-561
d2-561
e2-561
f2-561
g2-561
h2-561
a3-647
b3-647
c3-647
d3-647
e3-647
f3-647
g3-647
h3-647
Cycle1
Cycle2
b
BFP
293T expressing
CAG-BFP-A1-B3-C2-D1
a1-488
b1-488
c1-488
d1-488
a2-561
b2-561
c2-561
d2-561
a3-647
b3-647
c3-647
d3-647
Cycle1
Cycle2
Cycle3
Cycle4
c
All tested combinations
A B C D
1 1 1 1
1 3 1 2
1 3 2 1
2 1 3 2
2 2 2 2
3 1 1 3
3 1 3 1
3 2 2 3
3 3 3 3
d
Probe
a1 a2 a3 b1 b2 b3 c1 c2 c3 d1 d2 d3 e1 e2 e3 f1 f2 f3 g1 g2 g3 h1 h2 h3
A1 A2 A3 B1 B2 B3 C1 C2 C3 D1 D2 D3 E1 E2 E3 F1 F2 F3 G1 G2 G3 H1 H2 H3
Barcodes expressed in cell line

Extended Data Fig. 2 In Vitro Specificity Testing of Barcodes

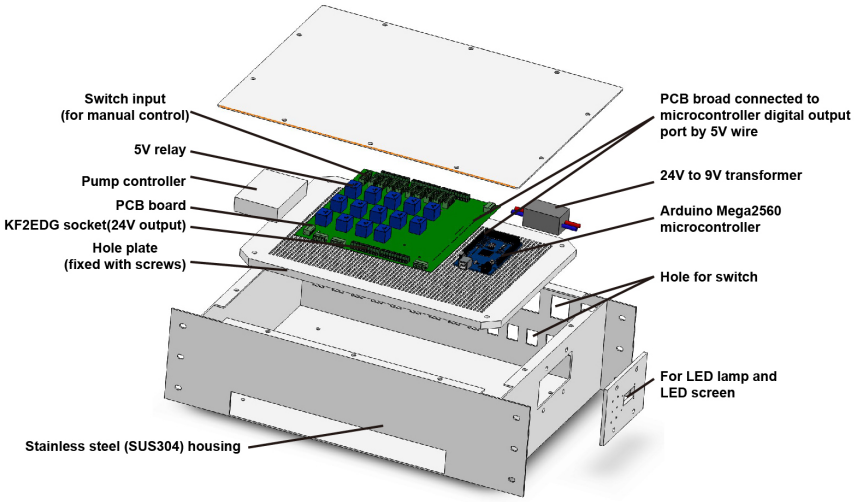
a. Example of a 2-cycle hybridization test. CAG-BFP-F1 was expressed in 293T cells; the first cycle excluded probe f1, while the second cycle included it. Scale=20 μm.

b. Specificity testing with a partial CASS barcode (AnBnCnDn). The barcode was expressed in 293T cells and analyzed via a 4-cycle FISH decoding process, demonstrating high specificity. Scale=20 μm.

c. Comprehensive testing with nine variants of partial CASS barcodes to assess probe specificity. Related to b.

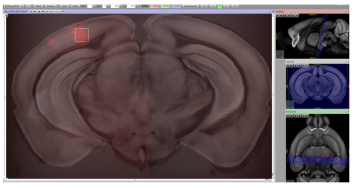
d. Summary of specificity testing, including results from 2-cycle hybridization (a) and 4-cycle FISH decoding (b, c), confirming high specificity for all probes to their corresponding sequences.

a



b

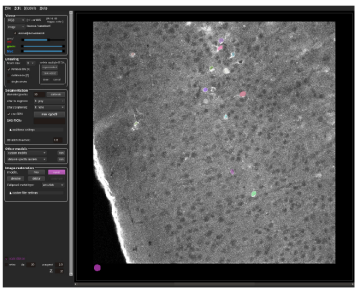
1. Raw data from confocal microscope
2. Register to Allen brain atlas(mouse V3 2017) by QucikNII(NITRC tools)



Get the transform matrix (for each slice)

$$[xq \quad yq \quad zq] = [x/w \quad y/h \quad 1] \times \begin{bmatrix} ux & uy & uz \\ vx & vy & vz \\ ox & oy & oz \end{bmatrix}$$

3. Cellpose3 Z-stack segmentation (for each view) using pre-trained custom model



Get the mask

4. 8-cycle images registering by Matlab

Multicycle XYZ shift correction

5. Semi-automatic barcode decoding by matlab

a. Select the background ROI
b. Calculate the score of all channels in each cell = mean intensity/background Std and get the xyz position of each cell
c. Compare the score of each channel and get the decoding result(1,2,3,0,or multiple) and reliability
 * high reliability(example a1)= [$Score_{a1}$>2] and [$Score_{a1}$>2*($Score_{a2}$+$Score_{a3}$)] and [$Score_{a2}$<1] and [$Score_{3}$<1]
d. Manual check the decoding result

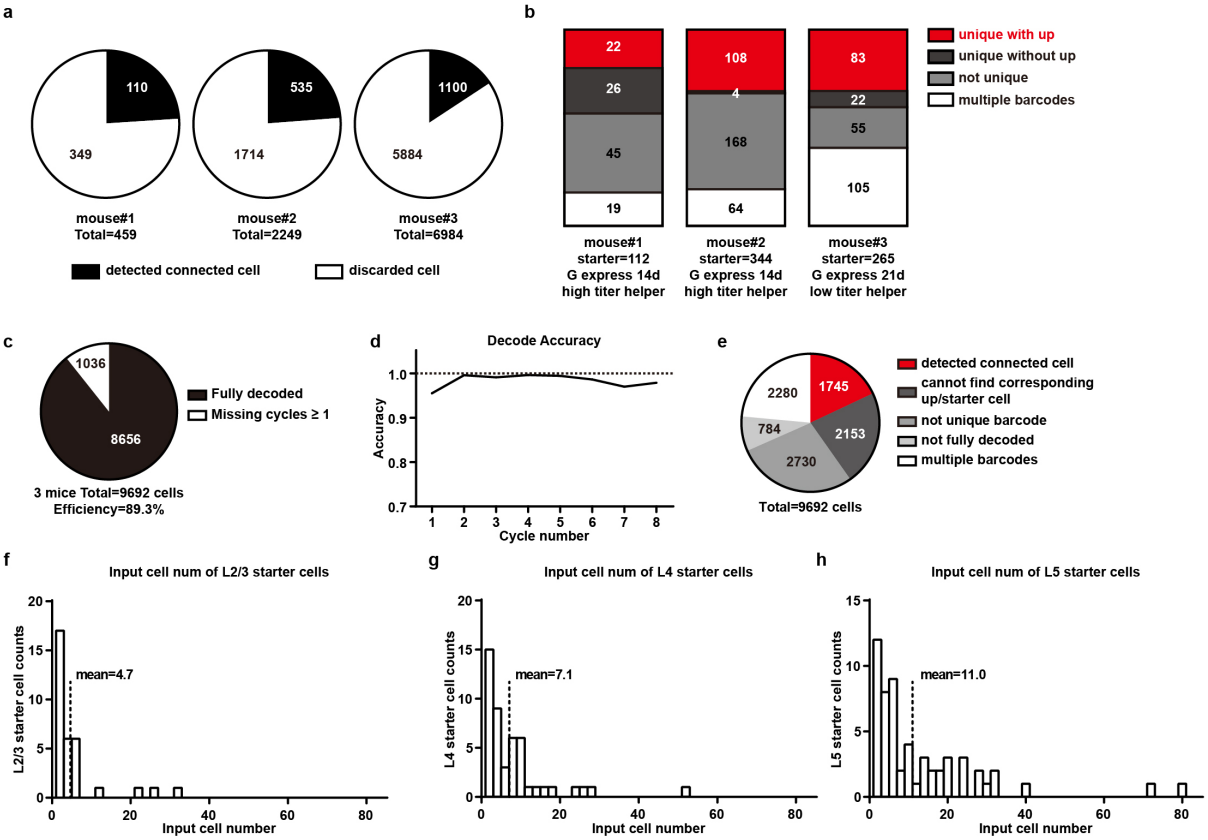
6. Transform the xyz position to Allen brain atlas

$$[xa \quad ya \quad za \quad 1] = [xq \quad yq \quad zq \quad 1] \times \begin{bmatrix} 0 & 0 & 25 & 0 \\ -25 & 0 & 0 & 0 \\ 0 & -25 & 0 & 0 \\ 13175 & 7975 & 0 & 1 \end{bmatrix}$$

Extended Data Fig. 3 Semi-Automated FISH System and Data Processing

a. Schematic of the microcontroller-based control system used in the semi-automated FISH system. Wires and switches are not shown.

b. Details of cell detection and atlas alignment. The XYZ shift in our setup was around 10 μm, basic imaging registration algorithm in Matlab was used and yielded reliable results.



Extended Data Fig. 4 Overview of Detected Cells

a. Proportion of connected and discarded cells for each mouse sample.

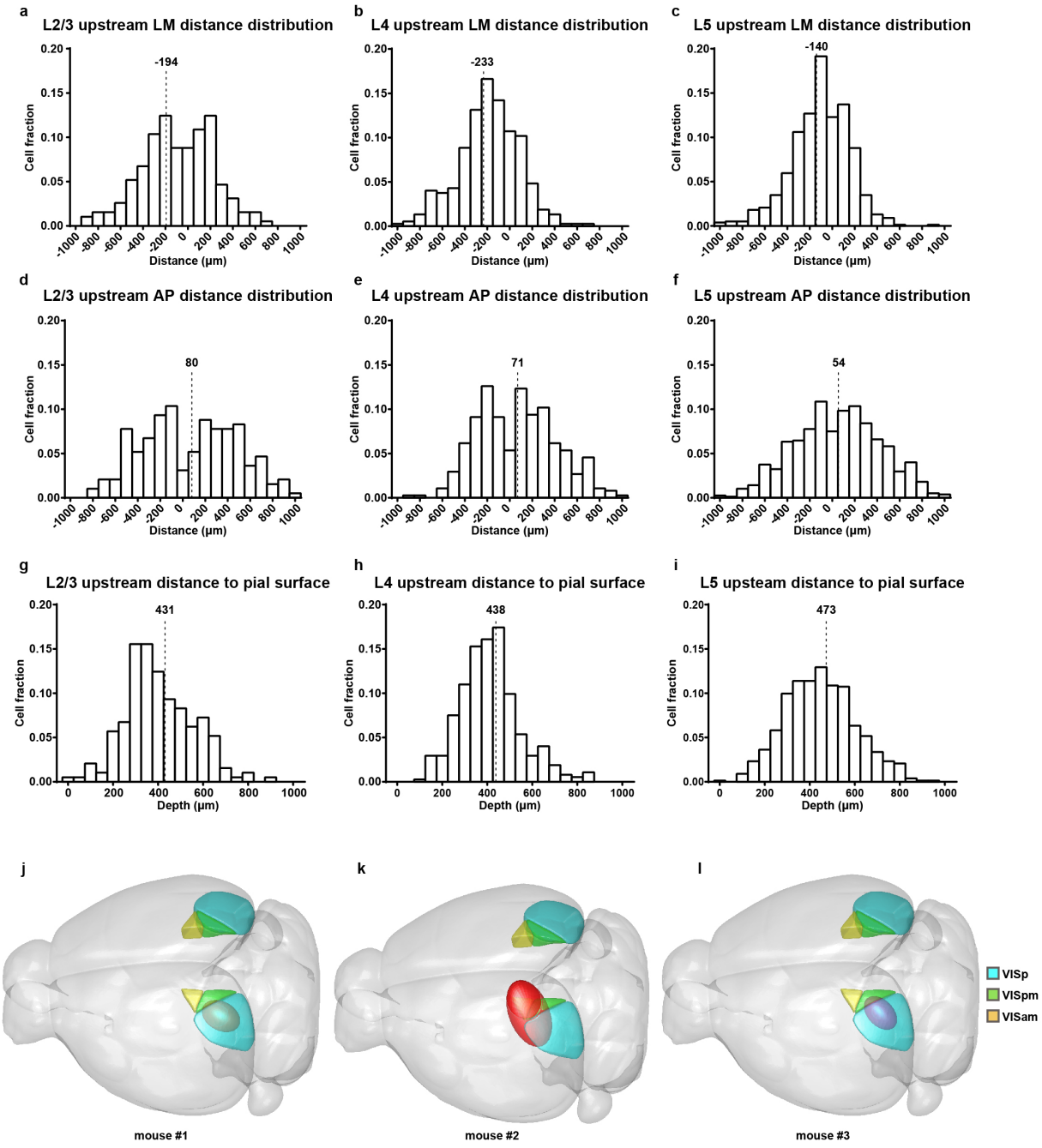
b. Distribution of starter cell categories across mouse samples.

c. Decoding efficiency for all detected cells.

d. Decoding accuracy for all detected cells in each cycle. Accuracy is defined as the fraction of cells that are fully decoded in each cycle.

e. Classification of all detected cells into various conditions. Only cells in the black group (“detected cells”) were included for connection detection. Fault-tolerant algorithms allowed some partially decoded cells to be identified as connected cells or carrying not unique barcodes (see method). The “cannot find corresponding input/starter cells” group here only includes fully decoded cells.

f-h. Distribution of input cell numbers across starter cells, grouped by cortical layers.



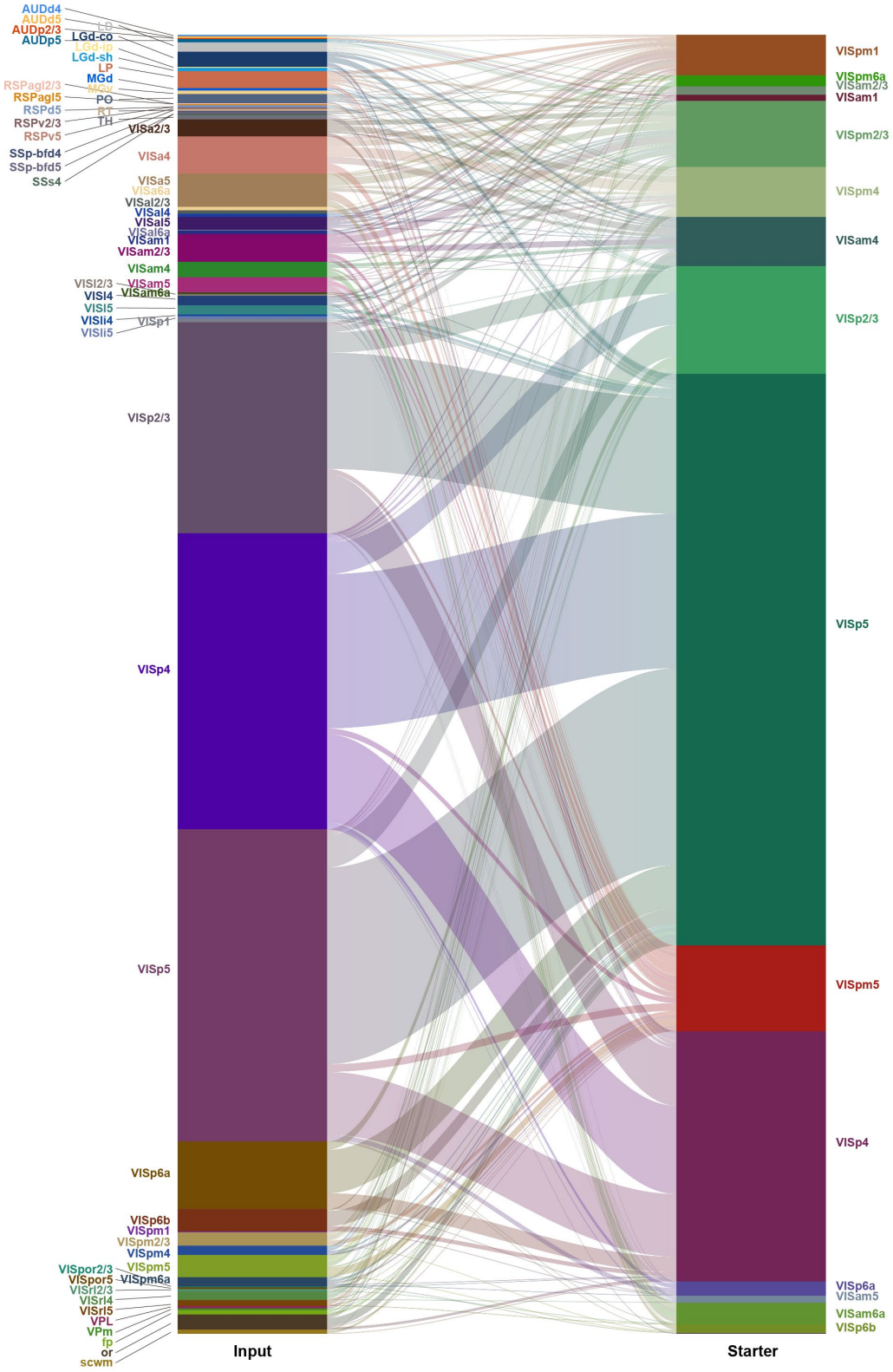
Extended Data Fig. 5 Spatial Organization of Input Cells

a-c. Input cell distribution along the lateral-medial axis, grouped by starter cell layers.

d-f. Input cell distribution along the anterior-posterior axis, grouped by starter cell layers.

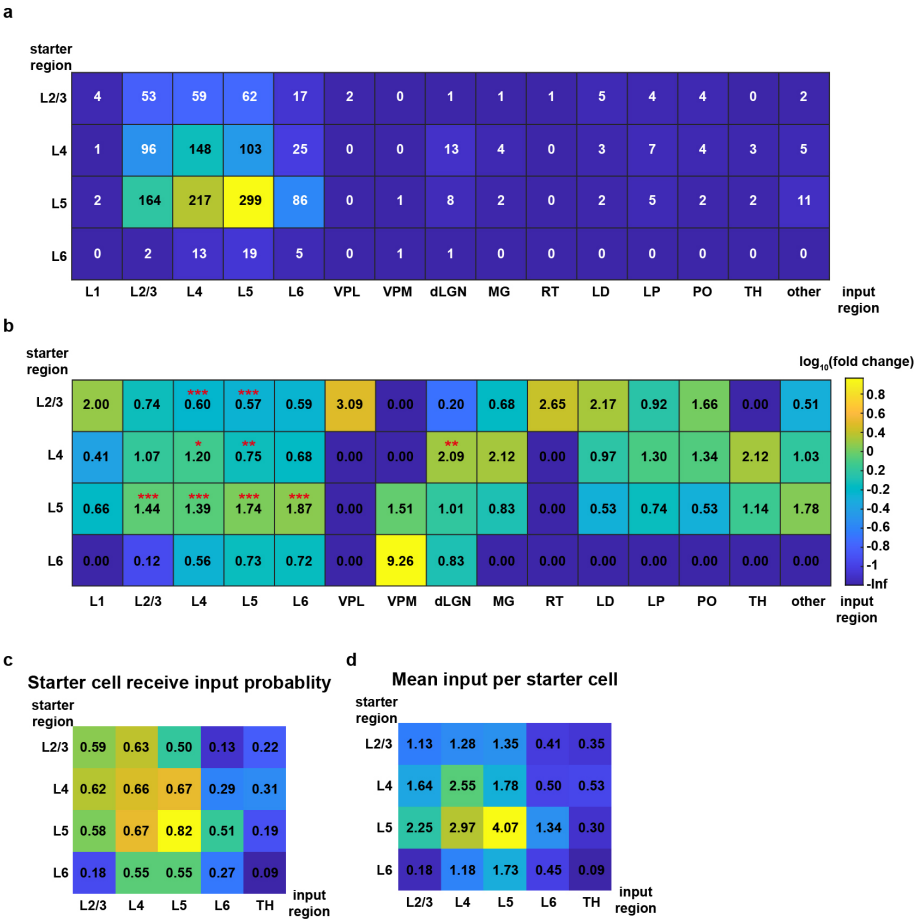
g-h. Distribution of input cell depths relative to the pial surface, grouped by starter cell layers.

i-l. Mean input territories of starter cells for individual experiments, including mouse #2 (injection site covering V1 and V2) and mice #1 and #3 (injection sites in central V1).



Extended Data Fig. 6 Sankey Diagram of Detected Connections

Connections between upstream and starter cells are represented as lines, with line thickness corresponding to connection strength (i.e., the number of connections). Brain regions are annotated according to the Allen Brain Atlas and are color-coded.



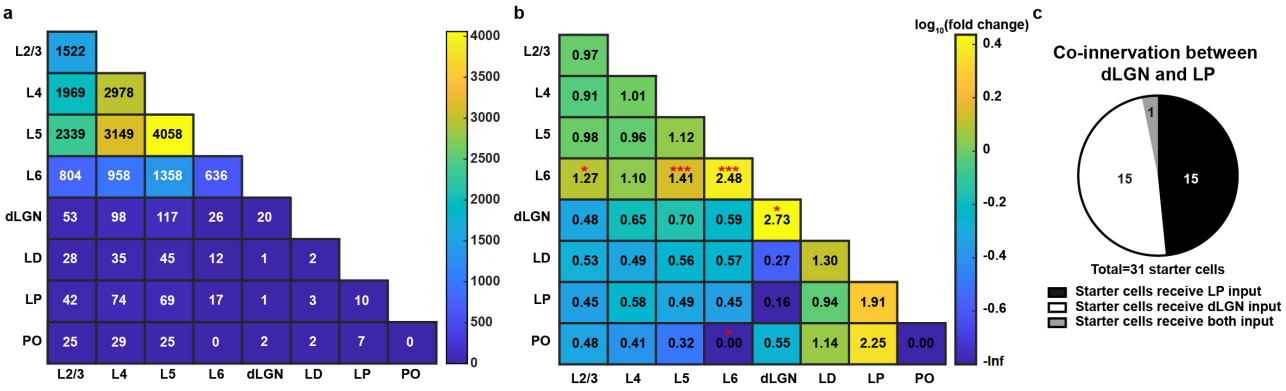
Extended Data Fig. 7 Connectivity Matrix

a. Observed counts of connections between brain regions.

b. Extended connectivity matrix, showing fold changes (observed/shuffled) as numbers and log10 fold changes as colors. Statistical analysis used Z-tests and FDR correction (Benjamini-Hochberg); *p < 0.05, **p < 0.01, ***p < 0.001. Shuffle iterations = 1000.

c. Probability of starter cells receiving input from various brain regions, with TH representing thalamic subregions.

d. Average number of input cells per starter across brain regions.

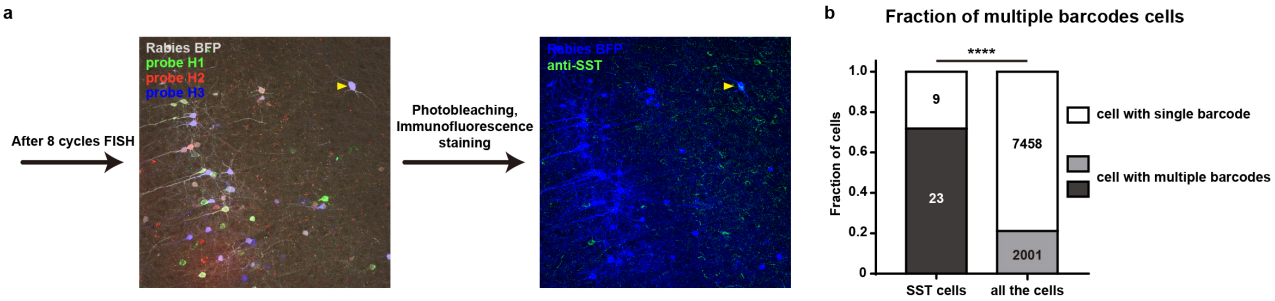


Extended Data Fig. 8 Co-Innervation Preferences of Input Cell Pairs

a. Counts of observed co-innervated pairs.

b. Observed co-innervation pairs compared with randomized data. Fold changes are shown numerically, and log10 fold changes are indicated by color. Statistical analysis used Z-tests and FDR correction; *p < 0.05, **p < 0.01, ***p < 0.001. Shuffle iterations = 1000.

c. Proportion of starter cells receiving input from dLGN and LP regions.



Extended Data Fig. 9 SST Neuron Identification by Immunofluorescence after barcode decodings

a. Immunofluorescence labeling performed after 8-cycle FISH.

b. Proportions of cells expressing multiple barcodes among SST neurons and all detected cells.