

In the format provided by the authors and unedited.

# Probabilistic cell typing enables fine mapping of closely related cell types in situ

Xiaoyan Qian <sup>1,4</sup>, Kenneth D. Harris <sup>2,4\*</sup>, Thomas Hauling<sup>1,2</sup>, Dimitris Nicoloutsopoulos<sup>2</sup>, Ana B. Muñoz-Manchado<sup>3</sup>, Nathan Skene <sup>2,3</sup>, Jens Hjerling-Leffler <sup>3</sup> and Mats Nilsson <sup>1\*</sup>

---

<sup>1</sup>Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden. <sup>2</sup>Institute of Neurology, University College London, London, UK. <sup>3</sup>Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. <sup>4</sup>These authors contributed equally: Xiaoyan Qian, Kenneth D. Harris. \*e-mail: [kenneth.harris@ucl.ac.uk](mailto:kenneth.harris@ucl.ac.uk); [mats.nilsson@scilifelab.se](mailto:mats.nilsson@scilifelab.se)

## Supplementary Discussion

### **Correspondence of identified cell classes with previously established ground truth**

Cell type assignments conformed closely to known combinatorial patterns of gene expression in CA1 interneuron subtypes. The identification of *Sst*<sup>+</sup> cells as O/LM or hippocamptoseptal correlated with further expression of *Reln* or *Npy*<sup>1,2</sup> (examples: **Figure 2A**, cells 1,2). Identification of *Pvalb* cells as axo-axonic, basket or bistratified correlated with further expression of *Pthlh*, *Satb1/Tac1*, or *Sst/Npy*<sup>1,3,4</sup> (Cells 3-5). Identification of neurogliaform (NGF) cells as caudal ganglionic eminence (CGE)-derived or medial ganglionic eminence (MGE)-derived/Ivy correlated with further expression of *Ndnf/Kit/Cxcl14* or *Lhx6/Nos1*<sup>5-8</sup> (Cells, 7,8). Identification of projection GABA neurons as trilaminar or radiatum-retrohippocampal correlated with expression of *Chrm2* or *Ndnf/Reln*<sup>2,9</sup> (Cells 8,9). *Cck* cells were identified as two subtypes correlated with expression of *Cxcl14*, with both expressing *Cnr1* and further subdivided by *Vip* expression<sup>6,10,11</sup> (Cells 10-11). Finally, interneuron-selective (IS) cells were divided into three classes correlated with the combinatorial expression of *Calb2* and *Vip*<sup>12,13</sup> (Cells 12-14).

The laminar distributions of identified cell types were consistent with ground truth established by previous work. Amongst *Sst*<sup>+</sup> neurons, O-Bi, O/LM or hippocamptoseptal were preferentially located in *stratum oriens* (*so*), while bistratified cells could also be found in *stratum pyramidale* (*sp*)<sup>14,15</sup> (*Sst/Nos1* cells were too rare to be reliably localized; **Supplementary Figure S14**). *Pvalb*<sup>+</sup> basket cells were found in *sp* and less often *so*, while rarer *Pvalb*<sup>+</sup> axo-axonic cells were found in the pyramidal layer<sup>16</sup>. Amongst NGF cells, those identified as having developmental origin in MGE, including Ivy cells, were found throughout all layers, while those having origins in CGE were found in *stratum lacunosum-moleculare* (*slm*)<sup>7,8</sup>. The two classes identified with long-range projecting GABAergic neurons were found in the expected layers: trilaminar cells primarily in *so*<sup>2,17,18</sup>, and radiatum retrohippocampal at the border of *stratum radiatum* (*sr*) and *slm*<sup>2,9,19,20</sup>. *Cck* interneurons were divided into two primary classes, with the *Cxcl14*<sup>+</sup> class located primarily in *sr*, close to the *slm* border, and the *Cxcl14*<sup>-</sup> class in all layers, as previously predicted<sup>6</sup>. Amongst interneuron-selective subtypes, cells identified as IS1 were found in all layers as expected<sup>13</sup>, while IS3 cells were located primarily in *sp* and *sr*, but very rare in *slm*<sup>10</sup> (IS2 cells were too rare for reliable quantification of their laminar distribution).

## Supplementary Methods

### **Padlock probe design**

Except for *Sst* and *Npy*, each padlock probe contained a 40 nucleotide (nt) recognition site, a 4nt barcode, a 20nt hybridization site, and a 20nt anchor sequence (with the latter being the same for all probes). The 4nt DNA barcode and the four possibilities for the hybridization site together define a length 5 barcode allowing each probe to be identified in five imaging rounds. The set of barcodes used were designed such that every pair differed in at least two positions. When multiple probes were used against a single gene, they typically all had the same gene-specific barcode sequence. However, for technical validation, three genes (*Cxcl14*, *Reln*, *Htr3a*) were

equipped with multiple barcodes (allowed to have only one-base difference), and in few other cases where previously ordered oligos were reused (*Calb2*, *Cdh13*, *Pde1a*, *Plcx2*, *Rorb* had two barcodes).

Probes were designed with an in-house Python software package which utilizes ClustalW and BLAST+ and supports parallel computing. Mouse transcriptome sequences were downloaded from NCBI RefSeq database, using gene name as search criterion. For genes with multiple isoforms, a multiple sequence alignment by ClustalW was first performed to find consensus regions, and any region shorter than 40nt was discarded. All the remaining target sequences were cut into 40nt sequence fragments, and only fragments with melting temperature between 65°C and 75°C were kept. Candidate fragments were then aligned against the mouse whole transcriptome, only considering the same strand polarity, using BLAST+ to test specificity. In addition to itself, if a fragment matched to another transcript or non-coding RNA with more than 50% coverage, 80% homology, and the coverage spanned the center 10nt, it was considered unspecific and discarded. All remaining candidates being at least 20nt apart along a transcript were considered final target candidates.

All target candidates were then converted into padlock probe sequences by cutting the target into two halves of 20nt each and by inserting a backbone sequence which contains a 20nt hybridization sequence, a 20nt anchor sequence, a 4nt barcode, a 5nt stabilizer sequence for sequencing-by-ligation (SBL) and a 6nt linker sequence. When designing *Sst* and *Npy* padlock probes, the 20nt anchor sequence in the backbone was omitted. Finally, probe sequences were selected manually from padlock probe candidates, taking into consideration the number of probes needed for a gene in relation to its expected expression level, and the distribution of target sequences along the transcript. All padlock probe sequences are shown in **Supplementary Table S2**. Probes were ordered as ultramer oligos from Integrated DNA Technologies (IDT) with 5'-phosphorylation modification. Detection-, anchor- and SBL oligos, as well as oligos for detection of *Sst* and *Npy* were also ordered from IDT with fluorophores conjugated (sequence and fluorophore modification in **Supplementary Table S2**).

### **Mouse sample preparation**

We used fresh frozen brain tissue from a CD1 male mouse, aged postnatal day 25. The brain was sliced into 10 µm coronal sections on cryostat (Leica) and were collected onto SuperFrost Plus (VWR) slides. The slides were kept at -80°C until use. All experimental procedures performed followed the guidelines and recommendations of local animal protection legislation and were approved by the local committee for ethical experiments on laboratory animals (Stockholms Norra Djurförsöksetiska nämnd, Sweden) under file N282/14.

### **In situ rolling circle products (RCP) generation**

Slides were taken out from -80°C and thawed at room temperature for 10 minutes. The sections were pre-fixed for 5 minutes in fresh 4% (w/v) paraformaldehyde (Sigma) in DEPC (Sigma)-treated PBS at room temperature, followed by one wash in DEPC-PBS-T (DEPC-treated PBS containing 0.05% Tween-20 (Sigma)). The tissue sections were then permeabilized with 0.1 M HCl (Sigma) for 5 minutes at room temperature, followed by two washes in DEPC-PBS-T. An ethanol (VWR) series of 70% (v/v), 85% (v/v) and ethanol absolute, 2 minutes each at room

temperature, was performed to remove fat and further permeabilize tissue. The sections were let dry in air and SecureSeal hybridization chambers (Grace Bio-Labs) were mounted onto slides.

Reverse transcription mix was added to the sections after a brief wash in PBS-T to rehydrate slides. The mix contained 0.5 mM dNTP mix (Thermo), 5  $\mu$ M random decamer (IDT), 0.2  $\mu$ g/ $\mu$ L BSA (NEB), 1 U/ $\mu$ L RIBOPROTECT RNase Inhibitor (Blirt) and 20 U/ $\mu$ L TranscriptMe reverse transcriptase (Blirt) in 1x reverse transcription buffer (Blirt). Slides were stored in a humid chamber and the reaction last overnight at 37°C. The mix was removed and fresh 4% (w/v) paraformaldehyde in DEPC-PBS was added to the sections without any wash in between. This post-fixation step aimed to cross-link newly synthesized cDNA to the cellular matrix and was carried out at room temperature for 30 minutes, followed by two washes in DEPC-PBS-T.

RNaseH digestion and padlock probing were performed in a single reaction mix. The mix contained 0.05 M KCl (Sigma), 20% formamide (Sigma), 20 nM of each padlock probe (638 probes for 84-gene panel, 755 probes for 99-gene panel), 0.2  $\mu$ g/ $\mu$ L BSA, 0.5 U/ $\mu$ L Ampligase (epicenter) and 0.4 U/ $\mu$ L RNase H (Blirt) in 1x Ampligase buffer (epicenter). The sections were first incubated at 37°C for 30 min for RNaseH digestion and moved to 45°C for 60 minutes for stringent hybridization and optimal Ampligase activity. The sections were washed twice in DEPC-PBS-T.

For rolling circle amplification, the sections were incubated in a mix containing 5% glycerol (Sigma), 250  $\mu$ M dNTP mix, 0.2  $\mu$ g/ $\mu$ L BSA, 1 U/ $\mu$ L Phi29 polymerase (Thermo Fisher Scientific) and 1x Phi29 buffer (Thermo Fisher Scientific) for overnight at room temperature, followed by three washes in DEPC-PBS-T.

### **RCP labeling**

A Lab Vision Autostainer 360 (AH Diagnostics) was used for SBL and detection oligo hybridization reactions. Reaction chambers were removed and tissue sections dehydrated by taking the slides through an ethanol series. The reaction area was lined out by ImmEdge Hydrophobic Barrier PAP Pen (Vector Labs). The slides were mounted in the autostainer, and a program carried out the following steps at room temperature: 1) wash once in DEPC-PBS-T and air-blow to remove residual reagent, 2) add anchor stain reaction mix with 2x SSC, 20% formamide and 0.1  $\mu$ M AlexaFluor 750-labeled anchor oligo and incubate for 15 minutes, 3) wash three times in DEPC-PBS-T and air-blow, 4) add SBL mix with 1 mM ATP (Thermo Fisher Scientific), four different base-interrogating oligos (0.1-0.3  $\mu$ M each), 0.5  $\mu$ g/ml DAPI (VWR), 0.2  $\mu$ g/ $\mu$ L BSA and 0.1 U/ $\mu$ L T4 DNA ligase (Blirt) and 1x T4 ligation buffer (Blirt) and incubate for 60 minutes, 5) wash three times in DEPC-PBS-T. The autostainer was kept in a dark room and the reaction mixes were prepared and loaded at the beginning of each run. To prepare for imaging, small amount of SlowFade Gold antifade mountant (Life Technologies) was added onto the sections and coverslips were mounted.

For subsequent cycles, a UNG-treatment step with 0.02 U/ $\mu$ L UNG (Thermo Fisher Scientific) and 0.2  $\mu$ g/ $\mu$ L BSA in 1x UNG buffer (Thermo Fisher Scientific) for 15 minutes followed by three washes with 60% formamide were performed before step 1) in the autostainer program. All staining cycles were identical except for that the base-interrogating oligos were changed for each reaction cycle. Moreover, in reaction cycle 5, no ligation was required. Instead, following UNG

treatment and formamide wash, a mix with 2x SSC, 20% formamide, four hybridization oligos (H1-H4) 0.1  $\mu\text{M}$  each, 0.1  $\mu\text{M}$  AlexaFluor 750-labeled anchor oligo and 0.5  $\mu\text{g/ml}$  DAPI was used in step 2), incubated for 30 minutes, and the program finished after step 3). For reaction cycle 6, detection of *Sst* and *Npy*, again no ligation was required. Similar to cycle 5, a mix with 2x SSC, 20% formamide, *Sst* and *Npy* sandwich probes 0.1  $\mu\text{M}$  each, two corresponding labeled oligos 0.1  $\mu\text{M}$  each, 0.1  $\mu\text{M}$  AlexaFluor 750-labeled anchor oligo and 0.5  $\mu\text{g/ml}$  DAPI was added to the sections, followed by 30 minutes incubation.

## **Microscopy**

After each round of labeling, all slides from an experiment were mounted onto an epifluorescence microscope AxioImager.Z2 (Zeiss) equipped with multi-slide stage and mercury short-arc lamp (HXP R 120 W/45 C VIS). First, only DAPI images were acquired using a 2.5x/0.075 objective in order to define tissue regions and to record coordinates outlining each tissue. After switching to a 20x/0.8 objective, images were acquired in 6 channels using Zeiss filter set 49 for DAPI, Chroma filter set 49020 for AF488 (base T), Chroma filter set SP102v2 for Cy3 (base G), Chroma filter set SP103v2 for TexasRed (base C), Chroma filter set SP104v2 for Cy5 (base A) and Chroma filter set 49007 for AlexaFluor 750 (anchor oligo). The images were taken using a 16-bit camera (C11440-22CU, Hamamatsu) and each field of view image is 2048 x 2048 pixels. The resolution is determined by the camera pixel size and magnification, therefore 0.33  $\mu\text{m}$  in our setup. At each tile (field of view), the image software ZEN (Zeiss) first performed automatic focusing based on DAPI channel, and stacks of 7 z layers were acquired for each channel; as we used widefield imaging followed by software focus stacking (rather than 3d confocal microscopy), this axial resolution sufficed to obtain good 2d images. An RCP has an estimated diameter of 0.5-1  $\mu\text{m}$ , so the sampling frequency is slightly below Nyquist limit. However, due to optical point spread, there is no risk of RCPs not being detected. 10% tile overlap was used to guide stitching in the analysis step. Imaging data was saved in ZEN's native czi format, which can be read by Bio-Formats (<https://www.openmicroscopy.org/bio-formats/>). In the next round of imaging, the slides were inserted into the same position in the stage as in the previous cycles and the sections were located by retrieving saved coordinates for each slide.

## **Simulations**

To estimate the accuracy of cell calling, and how this depends on the depth of classification required and the error rates of gene detection, we performed a simulation analysis.

To make the simulation, we discarded all information from the *in situ* dataset except the modal assigned class of each cell  $\hat{k}(c)$ , and each cell's segmented DAPI outline. We then simulated a dataset where each cell  $c$  was known *a priori* to be of class  $\hat{k}(c)$ . To do so, for each cell  $c$  we picked a random cell from the scRNA-seq database of class  $\hat{k}(c)$ . This random sampling captured the biological cell-to-cell variability of gene expression without any assumptions about its distribution, and therefore allowed us to test whether the assumed negative binomial distribution was suitable to model this variability parametrically. To model false-positive errors (misreads) in the *in situ* method we replaced a fraction  $\beta$  of the reads with randomly-chosen genes (the miscall rate  $\beta$  therefore ranges between 0 and 1); to model false-negative errors (inefficiency), we kept only a fraction  $\alpha\eta_g$  of the reads of gene  $g$  (drawn from a binomial

distribution), where  $\eta_g$  is the gene efficiency parameter estimated by the cell-calling algorithm, and the relative inefficiency rate  $\alpha$  controls the rate of false-negative errors:  $\alpha = 1$  would be the same false-negative rate as in our current method;  $\alpha \leq 1$  would simulate lower efficiency, and  $\alpha \geq 1$  would simulate higher efficiency than we obtained with the current sequencing chemistry. The reads were arranged spatially according to a Gaussian distribution of width equal to the cell's width, which allowed them to be located also outside the DAPI boundary.

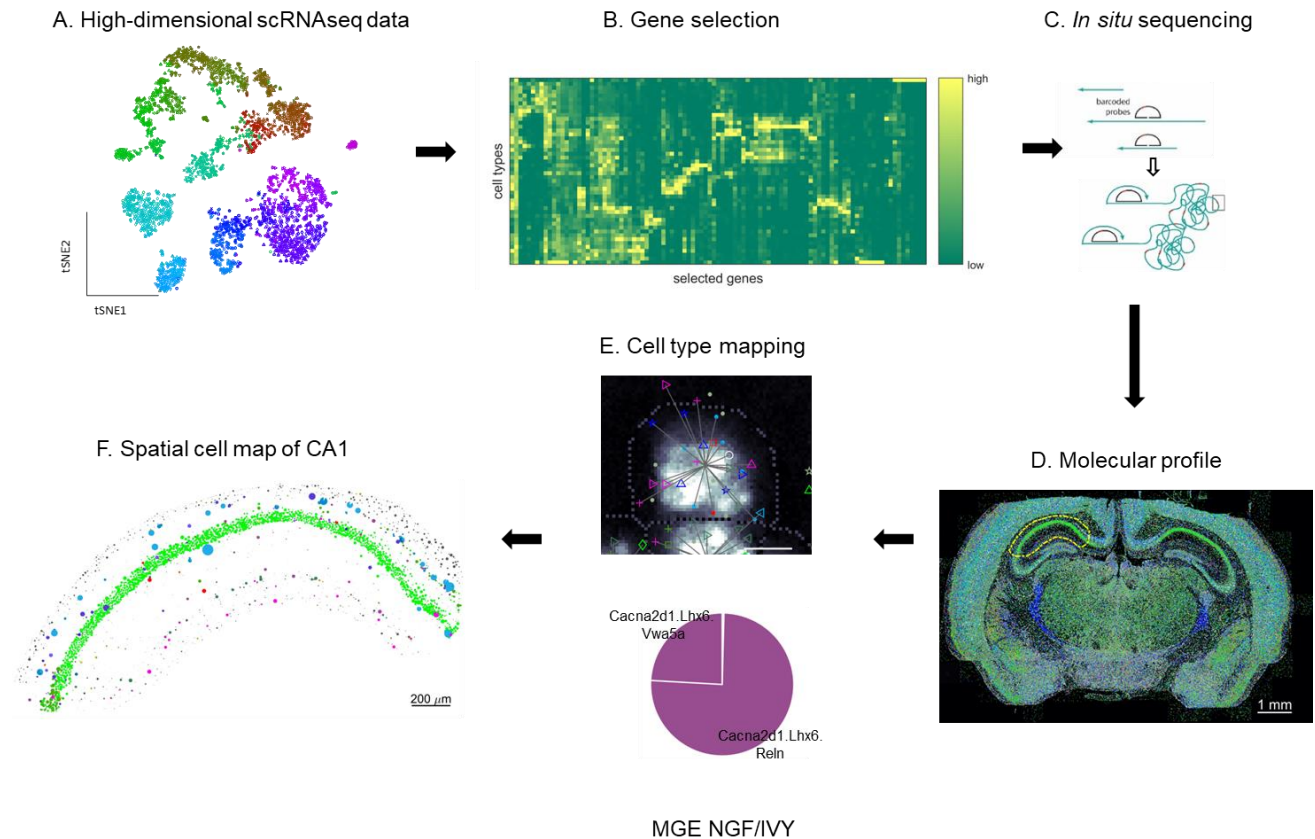
The performance of the algorithm was estimated for four different levels of required cell-type distinction, focusing only on inhibitory cell classes. For each level, we merged cell types according to the hierarchical classification scheme defined in Ref <sup>6</sup>. For example, at level 2, cells from both MGE-NGF subclasses *Cacna2d1.Lhx6.Refn* and *Cacna2d1.Lhx6.Vwa5a* are merged into a single class *Cacna2d1.Lhx6*, while cells from the CGE-NGF classes *Cacna2d1.Ndnf.Cxcl14* and *Cacna2d1.Ndnf.Rgs10* would be merged into a single class *Cacna2d1.Ndnf*; at level 1, all four fine types would be merged into a NGF superclass *Cacna2d1*. To assess the fineness of these distinctions, we computed the mean fraction of cells each class comprised. Because interneurons themselves only comprise 5% of the full population, these classes are very small: even at level 1, each interneuron subtype comprises on average 1.24% of all cells; while at level 3 they comprise on average 0.3% of all cells.

We assessed the quality of assignments the algorithm made by computing the median posterior probability assigned over cells simulated from an actual source class, to be assigned to each possible predicted class. This data was displayed as a matrix, for each division level (**Supplementary Figure S15A**). At division level 1, performance was nearly perfect; at lower division levels however, some cells were classified as belonging to related types. For example, at level 3, the algorithm was unable to accurately identify the fine subtypes of inhibitory-selective interneurons (*Calb2* classes).

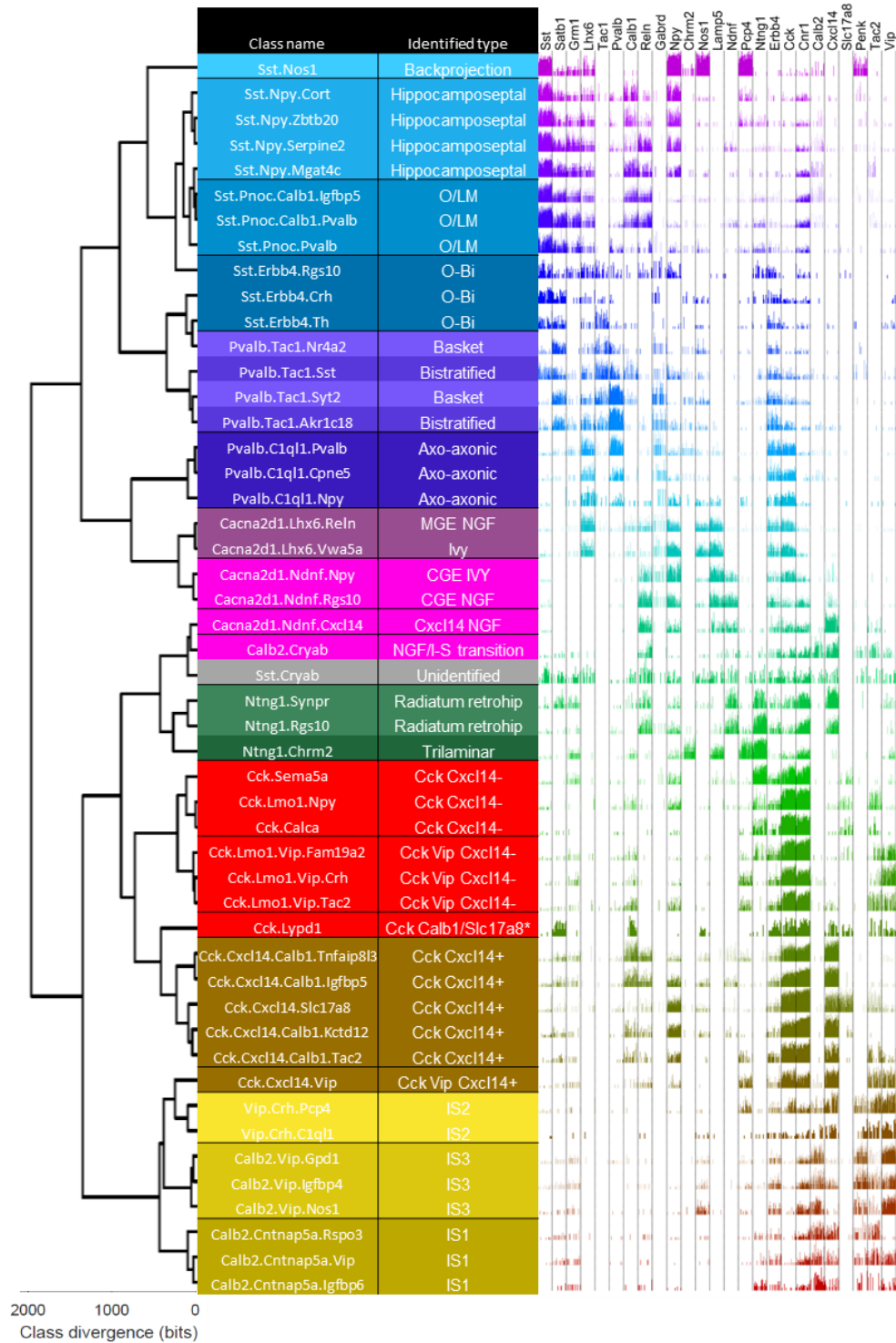
To quantify the performance of the algorithm, we computed the mean probability that a cell is assigned to the correct interneuron class, as the weighted mean of the diagonal elements in these matrices. At level 1, where each class comprised on average 1.24% of total cells, the correct class probability was 87%; at level 2 (class size 0.65% of cells) gave accuracy of 72%, while levels 3 and 4 (class sizes ~0.3% of cells) gave 53% and 51% accuracy. We conclude that at current efficiency levels the method gives excellent performance when required to distinguish cells to a level of subclasses comprising ~0.6% of the full population, but is less efficient at distinguishing yet finer subdivisions. However, even at the finest cell type level (level 4), the accuracy (51%) is 150 times better than chance level (0.3%).

To estimate the effects of different error rates, we recomputed the accuracy statistic as a function of the miscall rate and relative inefficiency parameters (**Supplementary Figure S15B**). We found that accuracy dropped rapidly with miscall rate. For example, a miscall rate of 30% led to an accuracy drop from 72% to 58% at subdivision level 2. Our simulations also showed that improved performance would be obtained with greater efficiency than currently possible: with relative efficiency of 2, accuracy increased from 72% to 83% at level 2. We conclude that improvements in the efficiency of gene detection would likely further boost cell calling performance.

## Supplementary figures

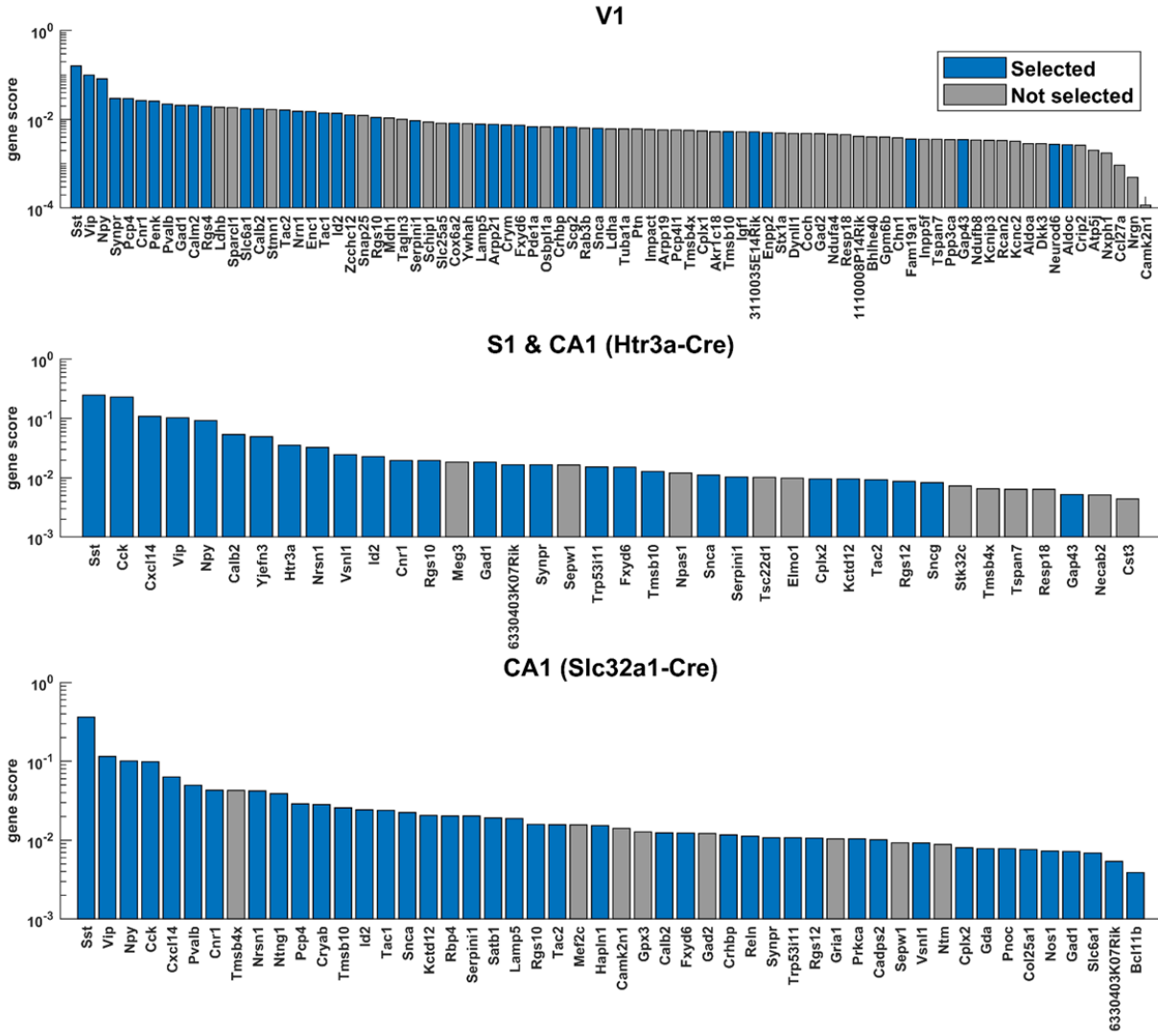


**Supplementary Figure S1.** The pciSeq pipeline. **A)** Cell-types are first defined by scRNAseq clusters. **B)** A set of ~100 marker genes are selected based on their ability to distinguish cell types in scRNA-seq data. **C)** A library of padlock probes targeting these marker genes is synthesized, and applied *in situ* to generate micron-sized barcoded amplification products for each detected RNA molecule. **D)** The gene-specific barcodes are sequenced *in situ*, generating a dense expression map of the targeted genes. **E)** Gene reads are assigned to cells and cells are assigned to cell-types according to a probabilistic model derived from scRNA-seq data. An example is shown for a medial ganglionic eminence (MGE)-derived neurogliaform/Ivy (NGF) cell. **F)** Cell-type maps are produced, with every cell displayed as a pie chart whose segment colors indicating the estimated probability the cell belongs to each class, and the size proportional to the number of assigned gene reads.

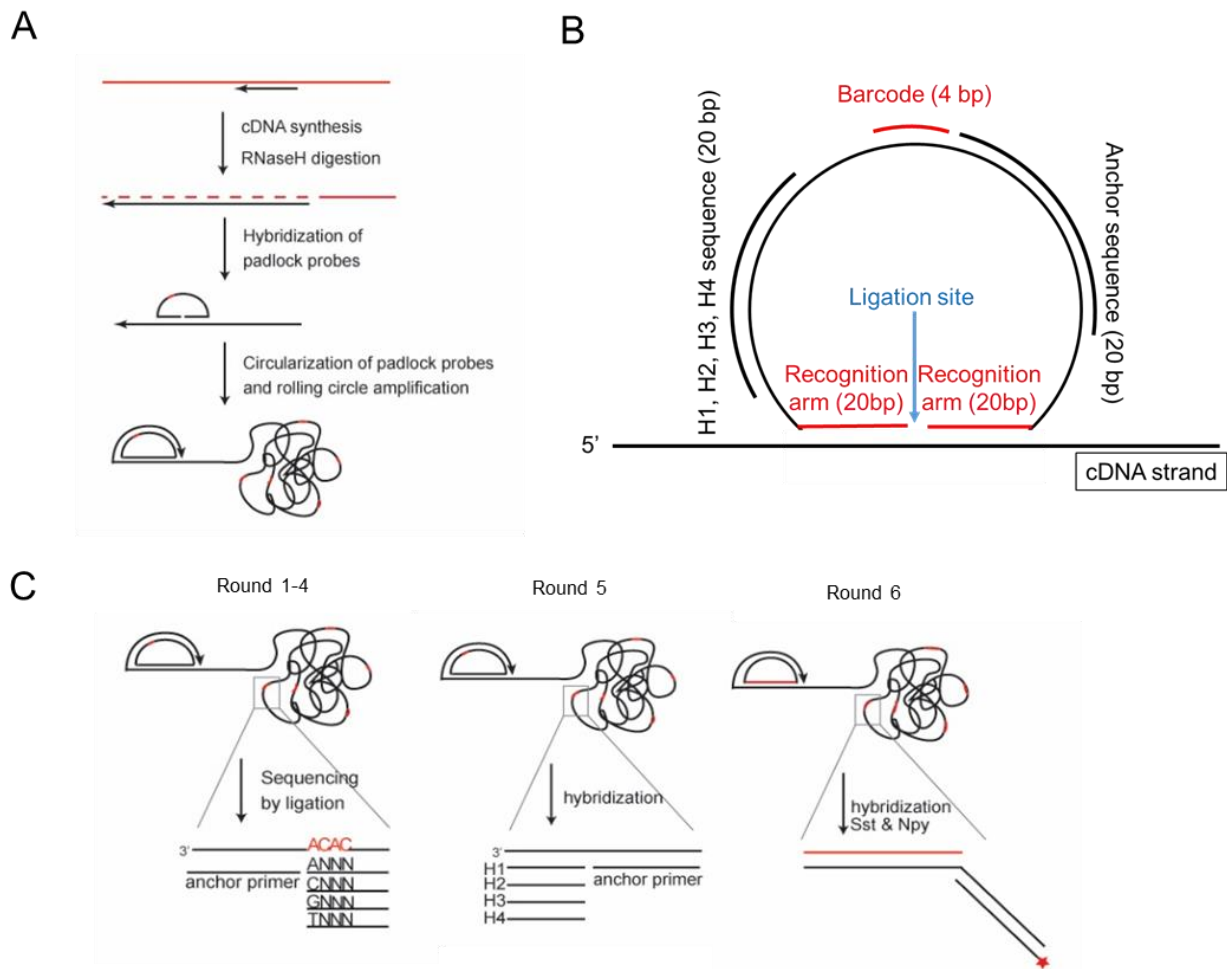


**Supplementary Figure S2.** Hippocampal inhibitory cell classes defined by scRNA-seq data from an independent study (modified from Ref. <sup>6</sup>). Note the genes shown on the right are an illustrative subset allowing identification of clusters with previously described biological classes, and are not the gene panel that was used to make cell class definitions either in the previous scRNA-seq study <sup>6</sup>, or in the current *in situ* study.





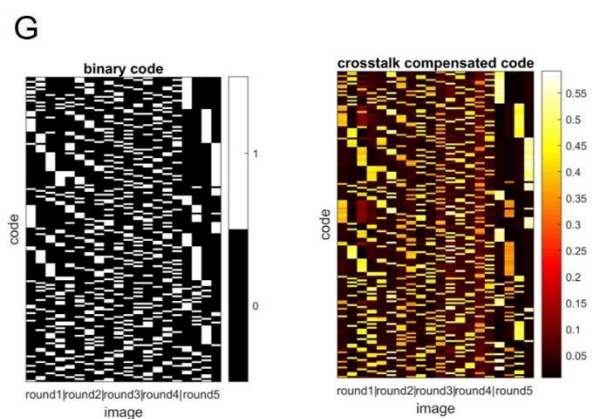
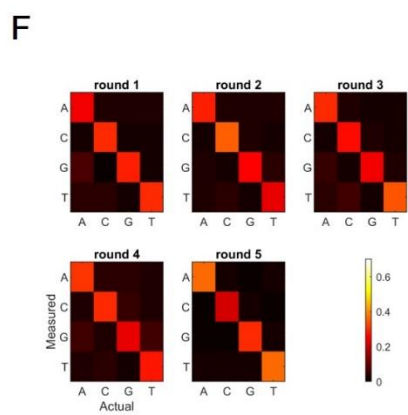
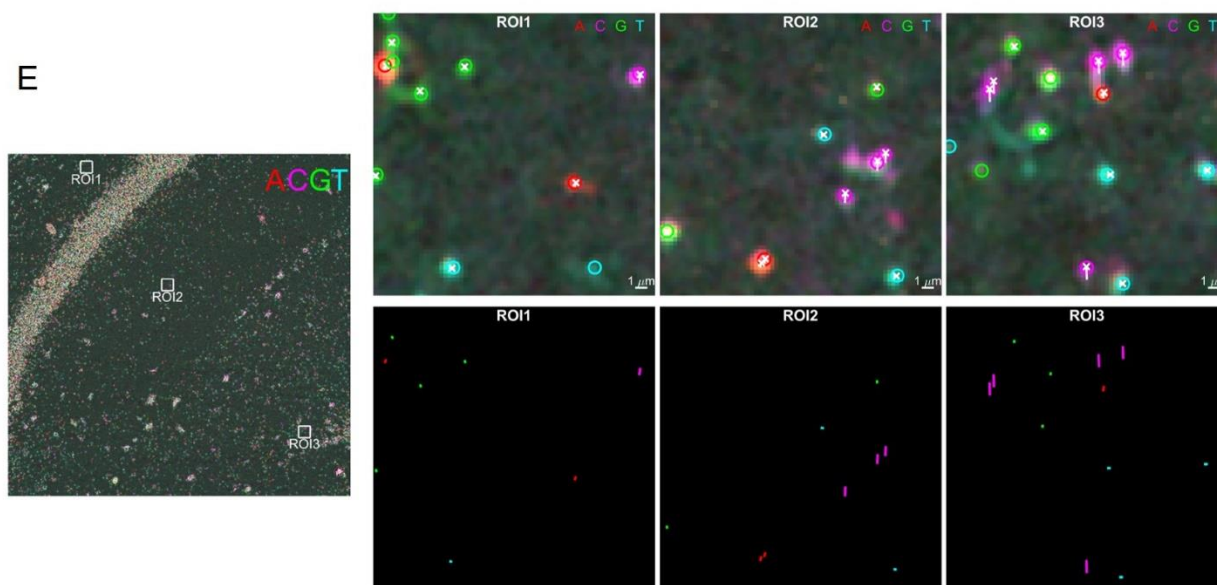
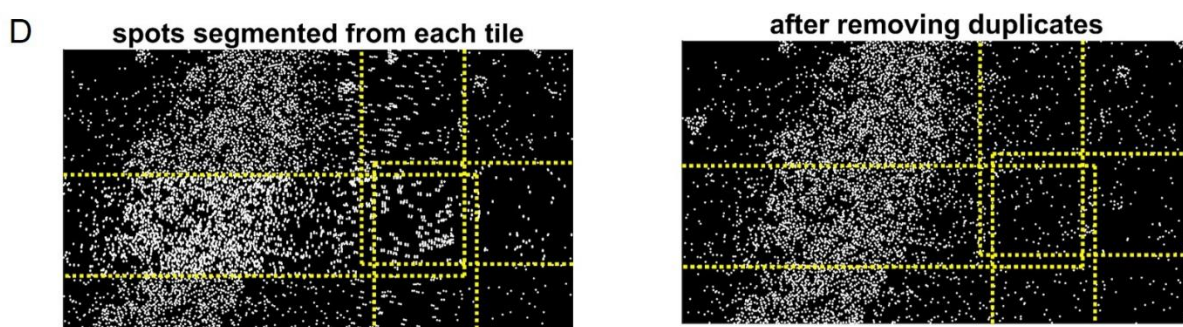
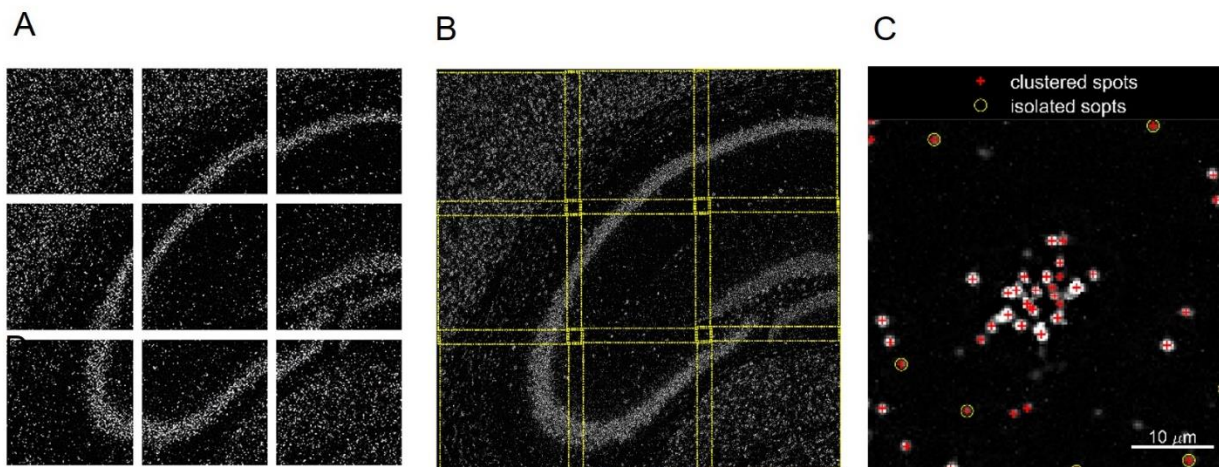
**Supplementary Figure S3.** Gene scores from automatic marker selection algorithm run on scRNA-seq from V1 interneurons<sup>5</sup> (top); S1 and CA1 interneurons from Htr3a-Cre mice<sup>21,22</sup> (middle); and CA1 interneurons from Slc32a1-Cre mice<sup>6</sup> (bottom). Y-axis represents a “gene score” summarizing each gene’s unique contribution to classification in the scRNA-seq database. Gray bars show genes excluded during manual curation.



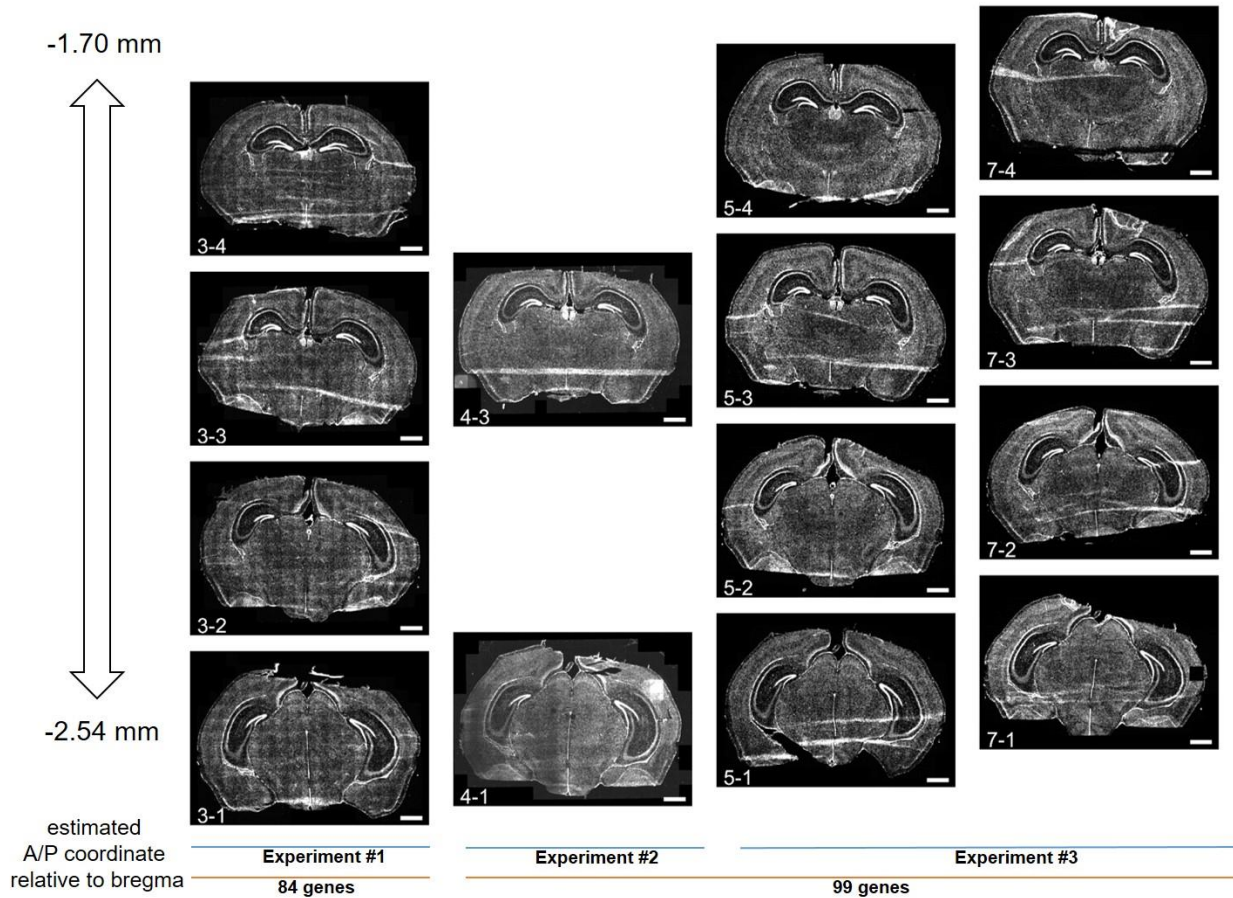
**Supplementary Figure S4.** Details of the *in situ* sequencing method. **A)** Rolling circle product (RCP) generation. An *in situ* cDNA library is generated using random primers. RNA is digested and padlock probes hybridized. Padlock probes are ligated into DNA circles, which are amplified into RCPs containing multiple copies of a barcode sequence specific to each gene. **B)** Padlock probe design. Padlock probes contain two recognition arms of length 20bp; a 4bp barcode; a 20bp hybridization sequence allowing a 5<sup>th</sup> round of imaging; and 20bp anchor sequence that acts as a primer for sequencing by ligation and also allows all RCPs to be labelled simultaneously. **C)** Readout strategy. Barcodes are read *in situ* with four rounds of sequencing by ligation. Each of these four rounds targets a specific position in the barcode sequence, and RCPs containing each of the four basepair possibilities for that position are labeled with a different color fluorophore. In round 5, the four possible hybridization sequences are targeted with one of four colored dyes. In round 6, two additional probes are used to detect RCPs corresponding to the strongly expressed genes *Sst* and *Npy*, which would overwhelm detection of the other genes if used in previous rounds.



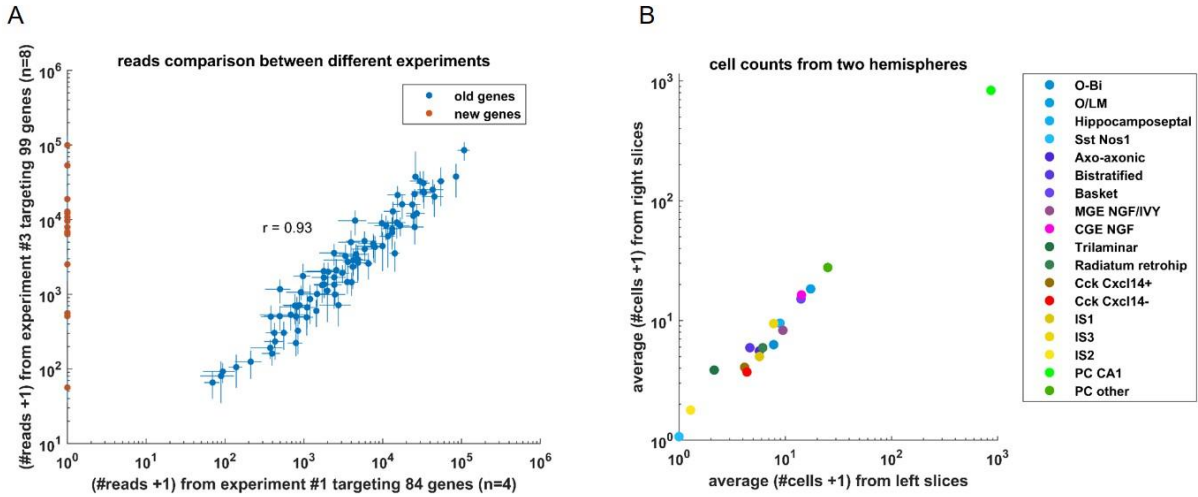
most padlock probes). Text at top (Cxcl14:p2, Cxcl14:p3, Cxcl14:p1) represents promoter sites defined in the FANTOM5 project <sup>23</sup>. These data suggest that the 5' end of the gene, including the target sites for the two 5'-ultimate padlock probes, is not expressed in brain, explaining their lack of detection.



**Supplementary Figure S6.** Image processing and base-calling algorithm. **A)** Tiled images were taken on an upright epifluorescence microscope (20x objective, 2048x2048 pixels). Images for all tiles and color channels are preprocessed by focus stacking from 7 z-slices and top-hat filtering. The panel shows an anchor channel image (labelling all RCPs) for 9 neighboring tiles, out of a total of 197 spanning the whole brain section. **B)** Anchor-channel images of all tiles are registered using a custom algorithm making use of the 10% overlap region between neighboring tiles. However, a single stitched image is never generated – instead, the offsets of each tile’s anchor channel relative to a global coordinate system are saved for later use. The panel shows an overlay of the anchor channel images, according to this alignment. **C)** RCPs are detected by thresholding of tophat-filtered anchor images, and well-isolated spots are identified, for later use estimating the crosstalk compensation matrix. Spot segmentation is performed in each tile individually. **D)** Anchor-channel images of separate tiles are registered into a global coordinate system, using a custom algorithm making use of the 10% overlap region. However, a single stitched image is never generated – instead, the coordinates of each detected spot are shifted into a global coordinate system using the relative offsets of each tile. Spots identified in multiple tiles are counted only once (see area enclosed by yellow dashed lines). This operation is performed only for a single “reference round” (round 2). **E)** Images for each tile, round, and color channel are aligned to that tile’s anchor channel on the reference round (round 2), using point-cloud registration (PCR). PCR can correct for chromatic aberration, as it fits an arbitrary affine transformation including rotations and shears (warping is possible in principle but was found to be unnecessary). PCR transformations are shown for three ROIs from a single tile (left). In the right six panels, the top row shows raw fluorescence images (background) superimposed with the PCR transformation vector joining spot position on the anchor channel/reference round to the position in the appropriate color channel (white lines). The colored circles indicates the position predicted by point cloud registration where fluorescence intensity is read off the image. The overlap between this position and the position of maximum brightness of the corresponding color indicates successful fit by PCR. Bottom row shows PCR vectors for each ROI, colored by color channel. Note that the vectors differ between color channels and ROIs, indicating an affine transformation was required to fit chromatic aberration. **F)** To compensate for optical or chemical crosstalk, a transformation matrix is learned from the well-isolated spots (yellow circles in (C)), using scaled k-means. This matrix encodes the actual fluorescence to be expected in each color channel, as a function of the true base value. **G)** The crosstalk matrix is applied to the binary 20-bit codes (four possibilities in 5 rounds), to generate a compensated code for each gene. The actual fluorescence, read off the point-cloud registered tile images for each spot, is compared to this compensated code using a cosine-angle measure, and each RCP is assigned to the best fit if the cosine exceeds a threshold of 0.9; spots not reaching the threshold are discarded.

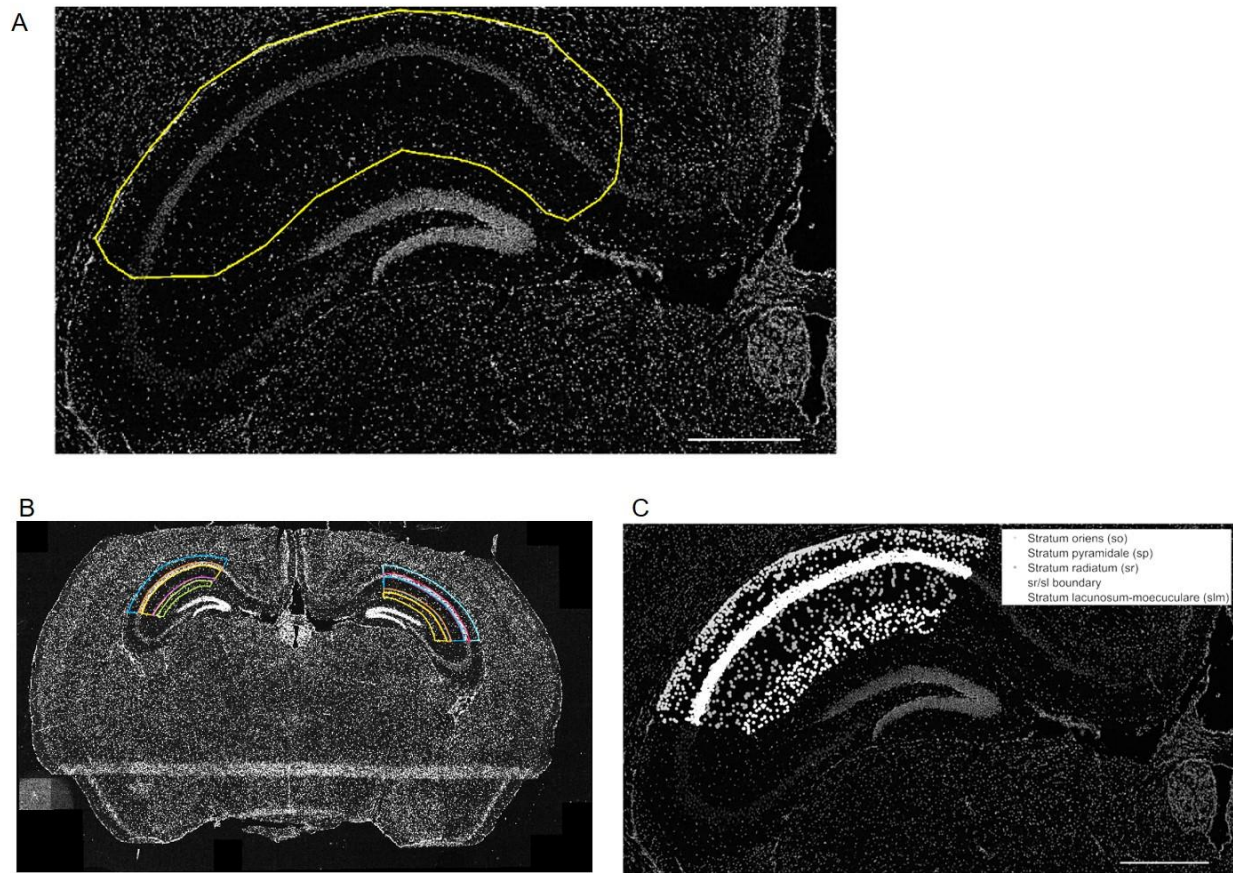


**Supplementary Figure S7.** All 14 sections and three experiments. Section ID in the lower left corner. Scale bars: 1 mm.

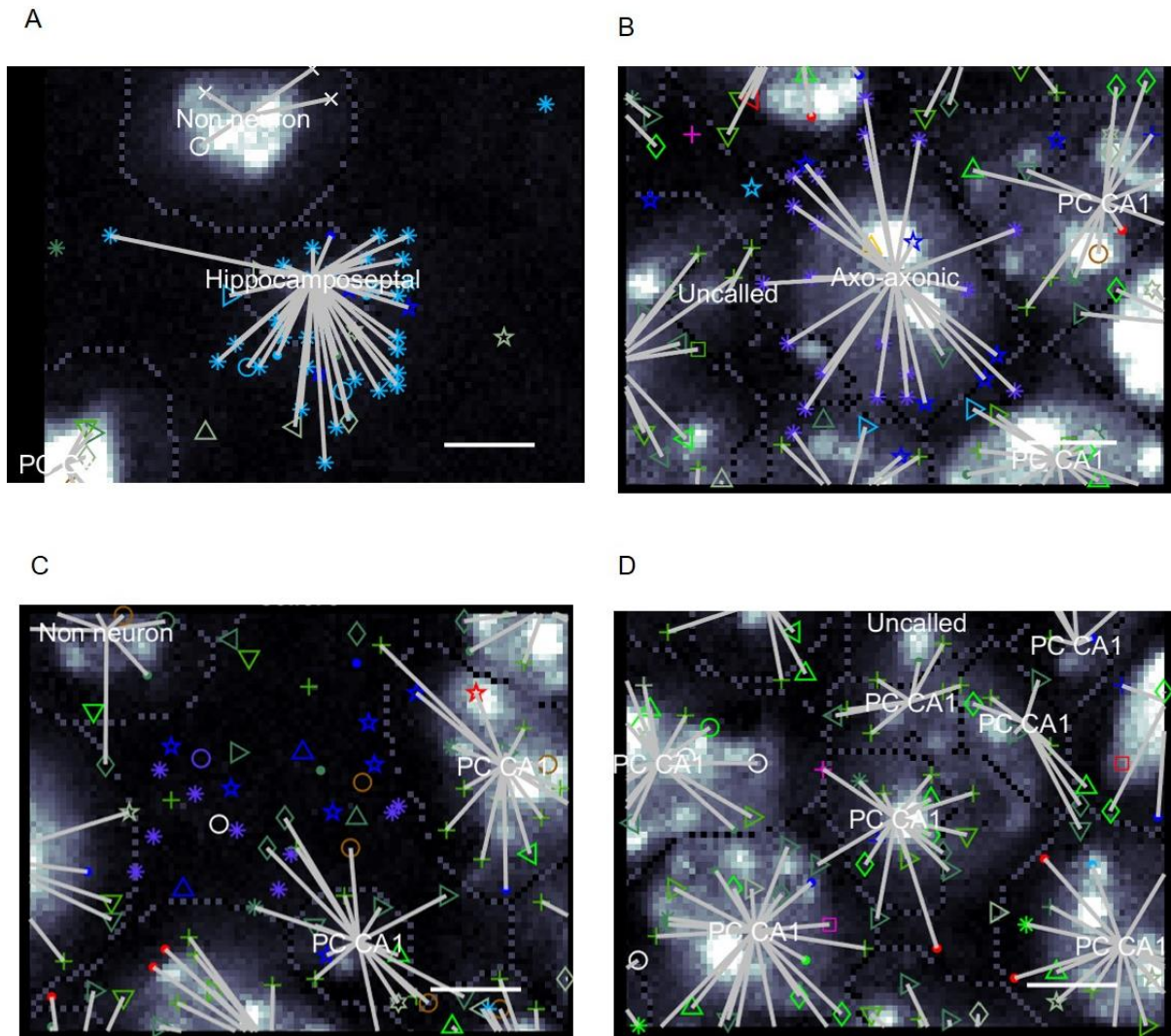


**Supplementary Figure S8.** Consistency of results across experiments. **A)** Mean counts per gene from experiment #1 (which used 84 genes) against mean counts from experiment #3 (which used 99 genes; new genes shown in red; Pearson correlation 0.93). Horizontal and vertical bars: standard deviation. Addition of new probes did not alter detection of original probes. **B)** Mean number of cells of each type detected in CA1 of the right hemisphere closely matches number in the left hemisphere.

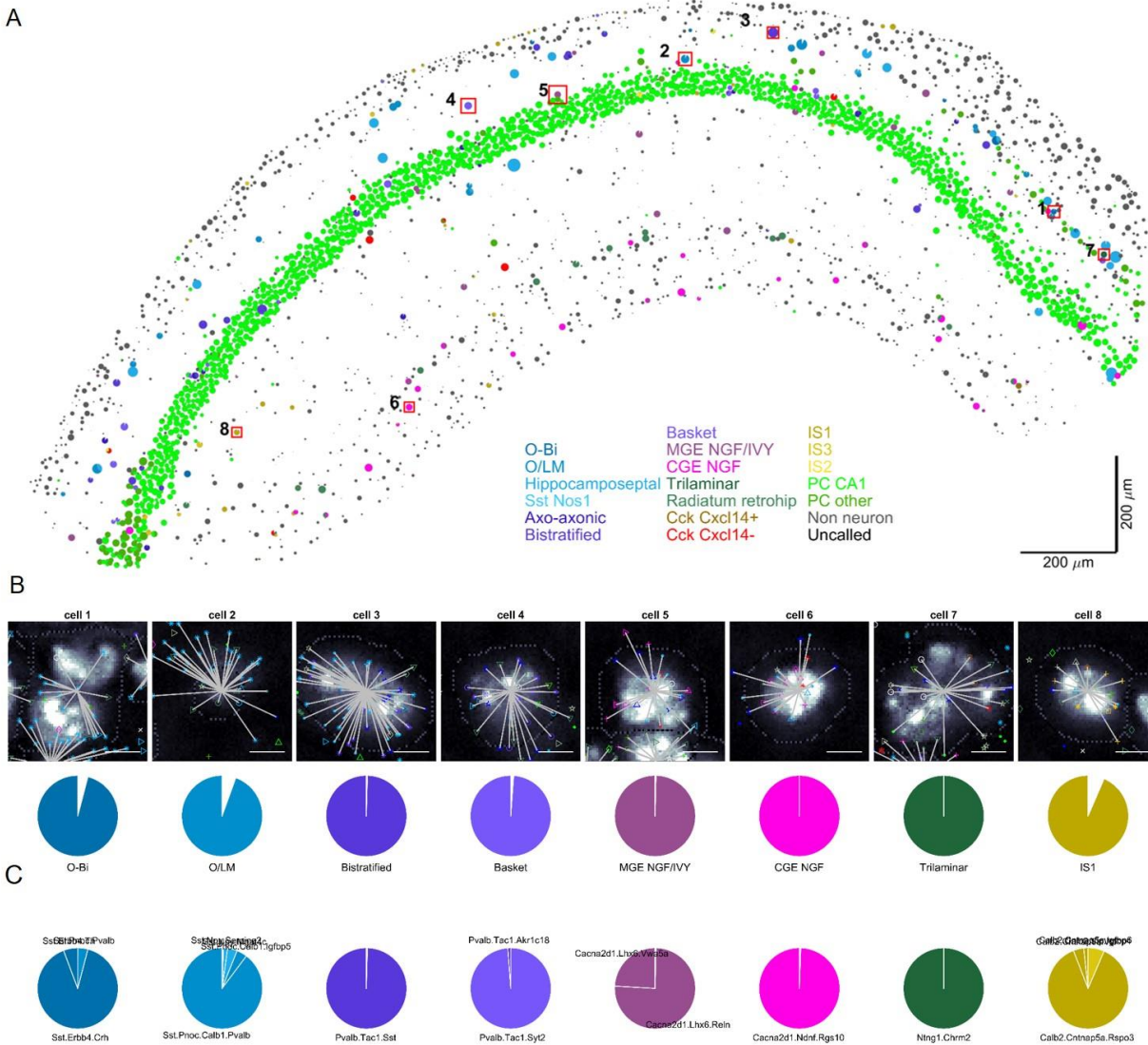




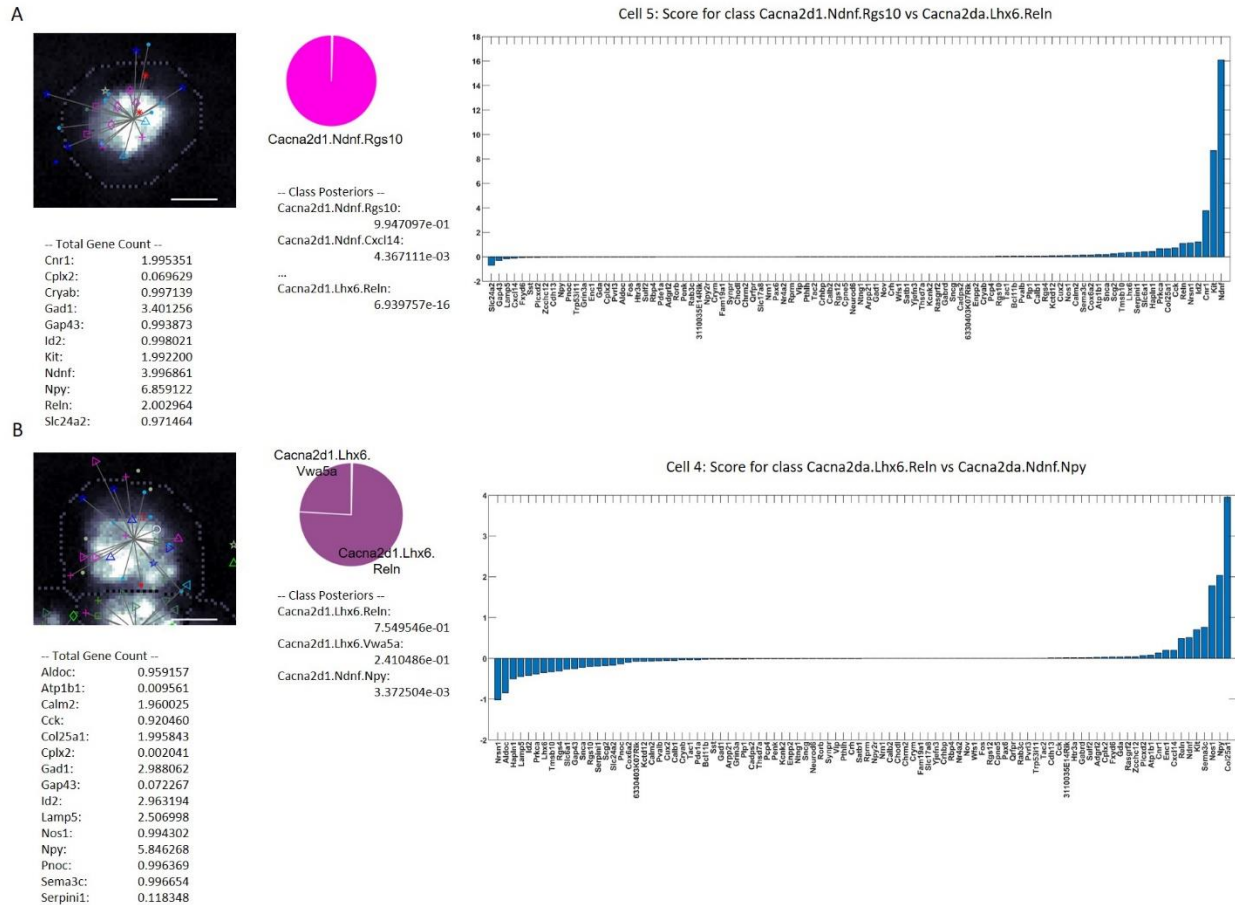
**Supplementary Figure S9.** Identification of area CA1 and its corresponding layers. **A)** For each section, the region corresponding to CA1 in each hemisphere was drawn manually by an operator based on DAPI stain patterns. **B)** Sublayers were identified by manual drawing of boundaries between stratum oriens, stratum pyramidale, stratum radiatum, stratum lacunosum-moleculare, and the border region between strata radiatum and lacunosum-moleculare. These boundaries were drawn using only the DAPI stain, and blind to cell-type identifications. **C)** Location of cells in drawn layers. All scale bars: 500 μm.



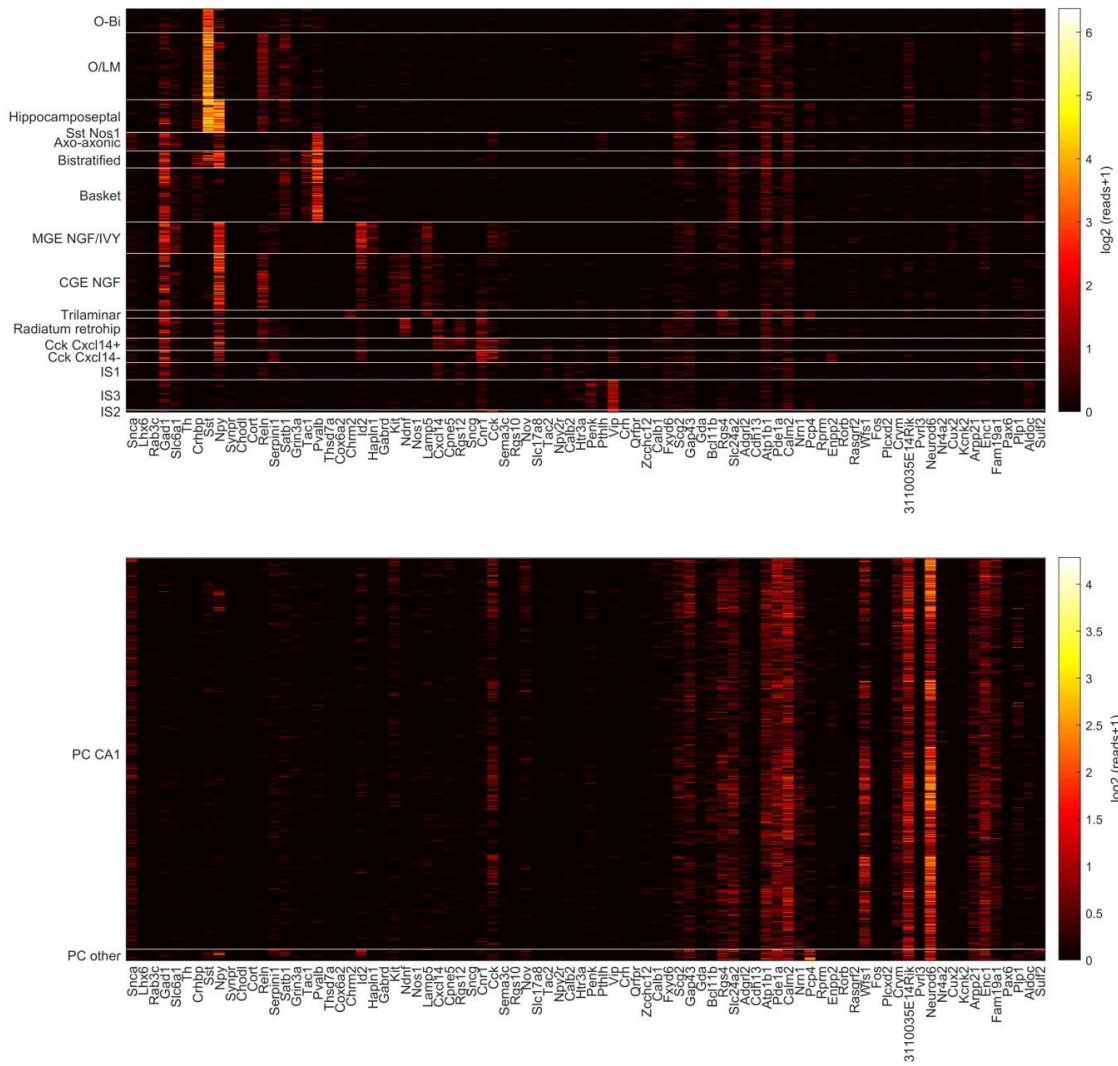
**Supplementary Figure S10.** Erroneous DAPI segmentation has a limited effect on cell calling and read assignment. Representative examples are shown. **A)** A case where the cell boundary assigned by DAPI segmentation (dotted line) was too small. Nevertheless, reads outside the cell were assigned to it by the probabilistic localization algorithm. **B)** A case where dense cell packing creates challenges for read assignment (hippocampal pyramidal layer). Many *Pvalb* detections are found outside the DAPI-segmented region of the identified axo-axonic cell, and some are inside the DAPI boundaries of neighboring pyramidal cells. However, because *Pvalb* is not expressed in pyramidal cells, these reads are assigned to the axo-axonic cell by the probabilistic algorithm. **C)** A section containing genes that must belong to a basket cell (*Pvalb*, *Tac1*), but for which no DAPI is visible, as its nucleus must have been outside the 10  $\mu\text{m}$  section. The cell is missed, but its reads are not erroneously assigned to the surrounding pyramidal cells. **D)** On occasion, pyramidal cells are over-segmented into multiple small cells, which are all assigned to the same cell type. This is the most common case of error in our algorithm. This will lead to overestimating of cell counts, but not assigning of cells to the wrong types. Scale bars: 5  $\mu\text{m}$ .



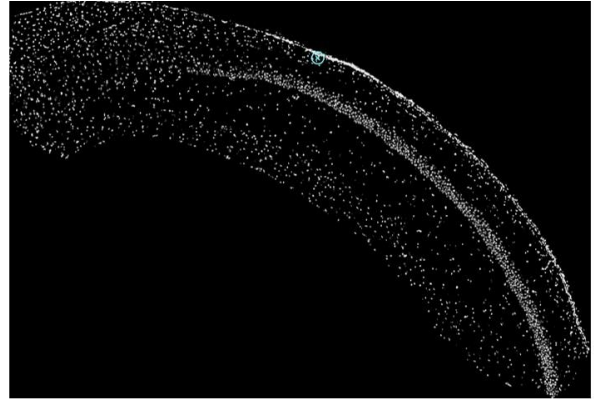
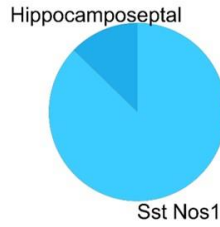
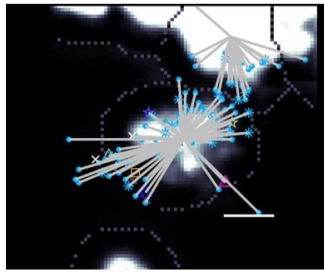
**Supplementary Figure S11.** **A)** Cell calling results from another hippocampus (experiment 4-3 left hemisphere). **B)** Read maps for example cells with pie charts showing posterior probabilities. Scale bars: 5  $\mu$ m. **C)** Pie charts showing posterior probabilities for scRNA-seq subclasses. All 28 maps in this study can be found in Supplementary Results, and in the online viewer at <http://insitu.cortexlab.net>.



**Supplementary Figure S12.** Diagnostics indicating which genes influenced the algorithm's predicted cell types. **A)** Left, weighted gene counts for Cell 5 of Supplementary Figure S10 (non-integer values arise because genes are weighted by their probability of assignment to this cell). Middle, posterior cell-class probabilities for this cell. It was called a Cxcl14-negative subtype of CGE-derived neurogliaform cell with 99.5% probability, while the probability assigned to it being a MGE-derived/Ivy cell was  $7 \times 10^{-16}$ . Right, diagnostics indicating the genes weighting the algorithm's assignment. The presence of *Ndnf*, *Kit*, and *Cnr1* identify this as a CGE-derived neurogliaform cell. **B)** Cell 4 from Supplementary Figure S10 was called as a MGE-derived neurogliaform/Ivy cell with 99.6% probability, but its precise subtype could not be identified confidently. Expression of *Col25a1*, *Nos1*, and *Sema3c*, together with a lack of *Kit* and *Ndnf* identified this cell as MGE-derived.



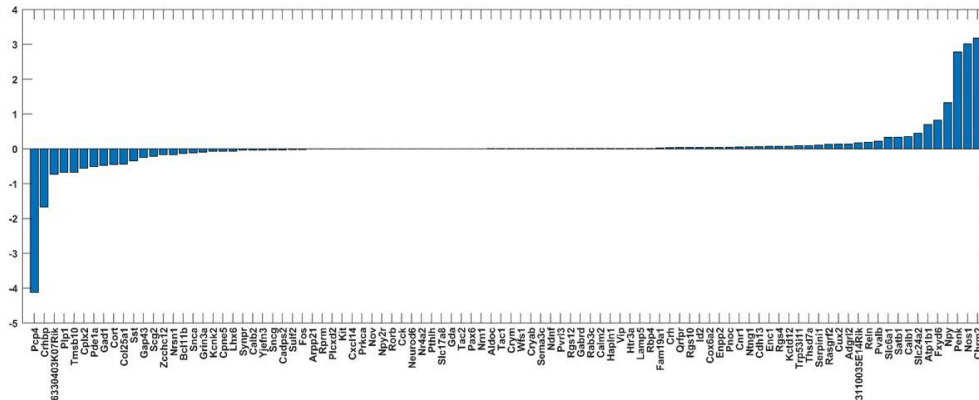
**Supplementary Figure S13.** Pseudocolor representation of read counts from all typed interneurons and pyramidal cells.

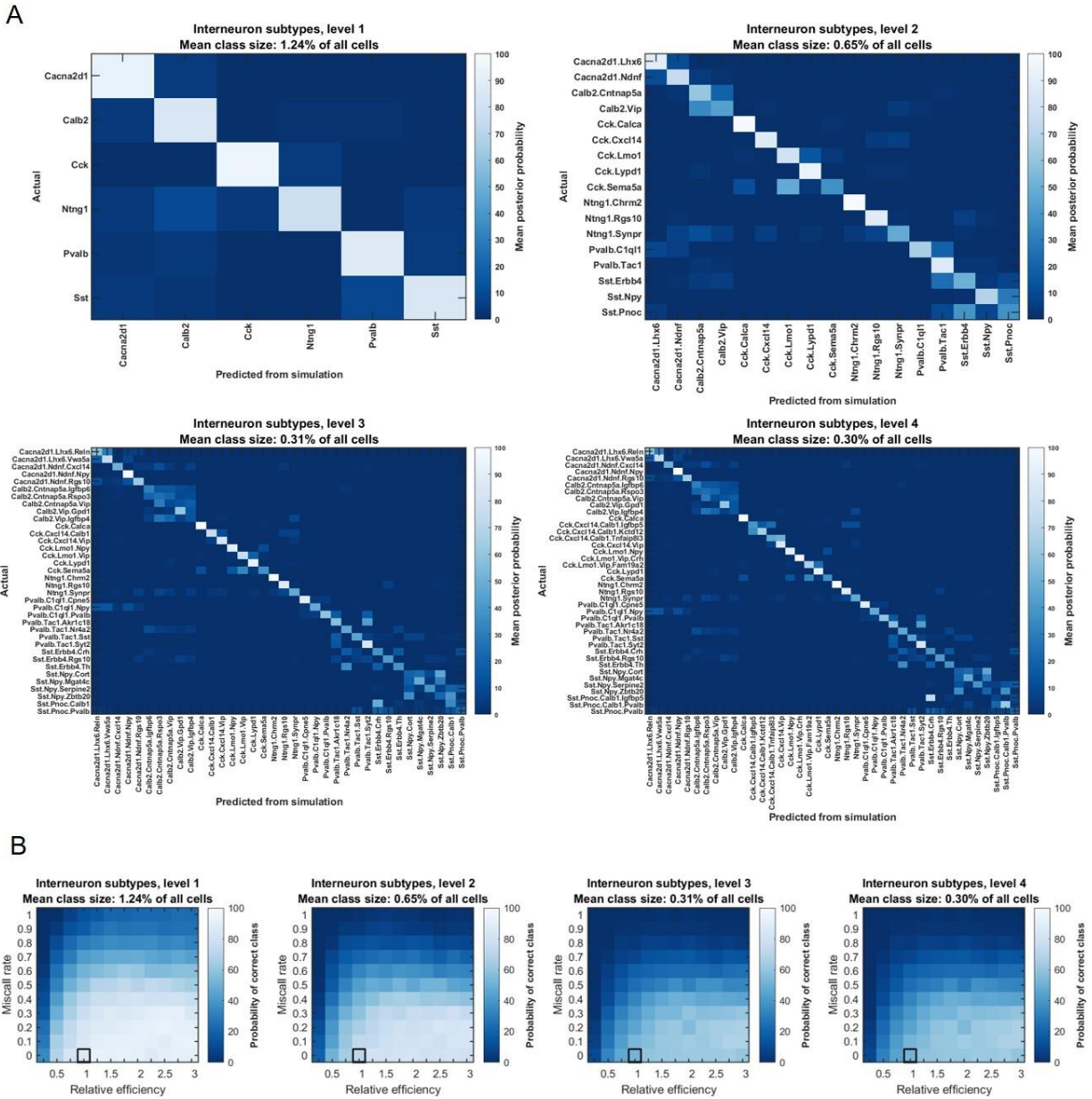


-- Total Gene Count --  
 Chrm2: 0.973348  
 Col25a1: 0.885874  
 Fxd6: 0.835859  
 Gap43: 0.993384  
 Nos1: 0.992781  
 Npy: 50.839523  
 Pde1a: 0.998942  
 Penk: 0.981762  
 Plp1: 1.012358  
 Sst: 19.289720

-- Class Posteriors --  
 Sst.Npy.Zbtb20: 9.947708e-02  
 Sst.Npy.Cort: 2.816030e-02  
 Sst.Nos1: 8.723437e-01

Cell 1784: Score for class Sst.Nos1 vs Sst.Npy.Zbtb20



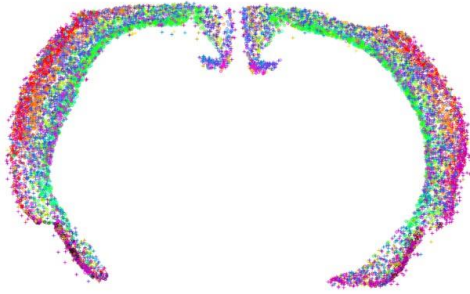


**Supplementary Figure S15.** Performance evaluation of the cell calling algorithm. We estimated the algorithm's accuracy using simulated data with a range of relative efficiencies and miscall rates. The simulation was based on resampling an scRNA-seq database, and did not rely on any assumptions made by our probability models. Accuracy was estimated for four different levels of fine subdivision (see Methods). **A**) For four levels of subdivision, where interneurons were divided into finer subclasses comprising on average 1.24%, 0.65%, 0.31% and 0.30% of all cells, accuracy was 87%, 72%, 53%, and 51%. **B**) Accuracy rates for a range of simulated miscall rates and relative efficiencies (see Methods). Black squares mark the data corresponding to (A), same efficiency and miscall rate as in the original data.

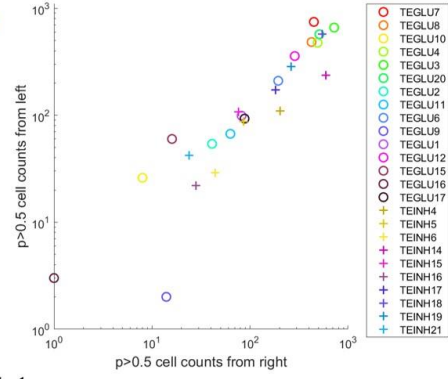




A

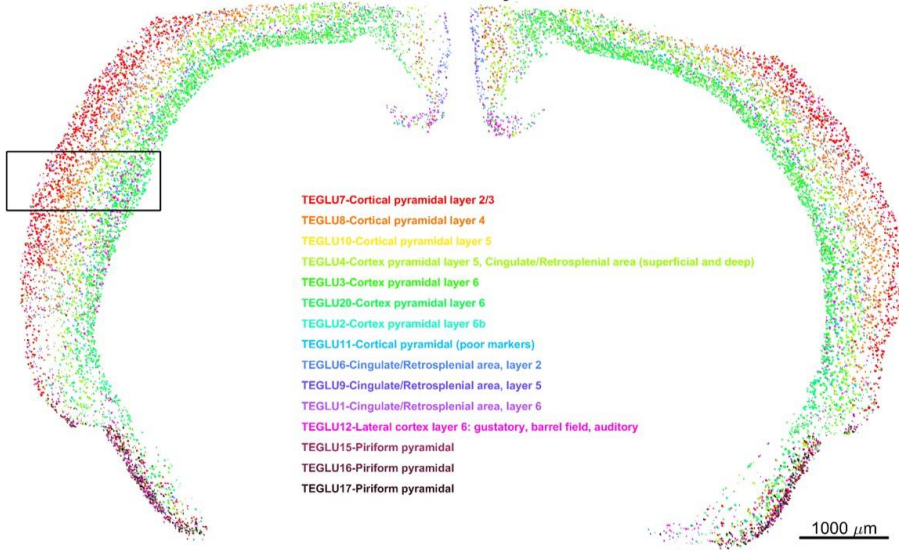


B



C

**cortical excitatory neurons**



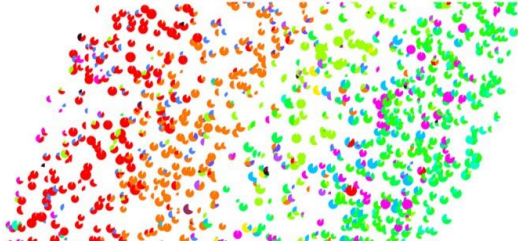
D

**cortical inhibitory neurons**

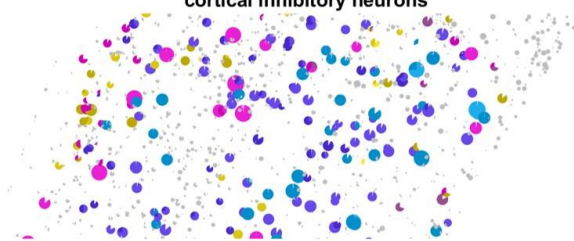


E

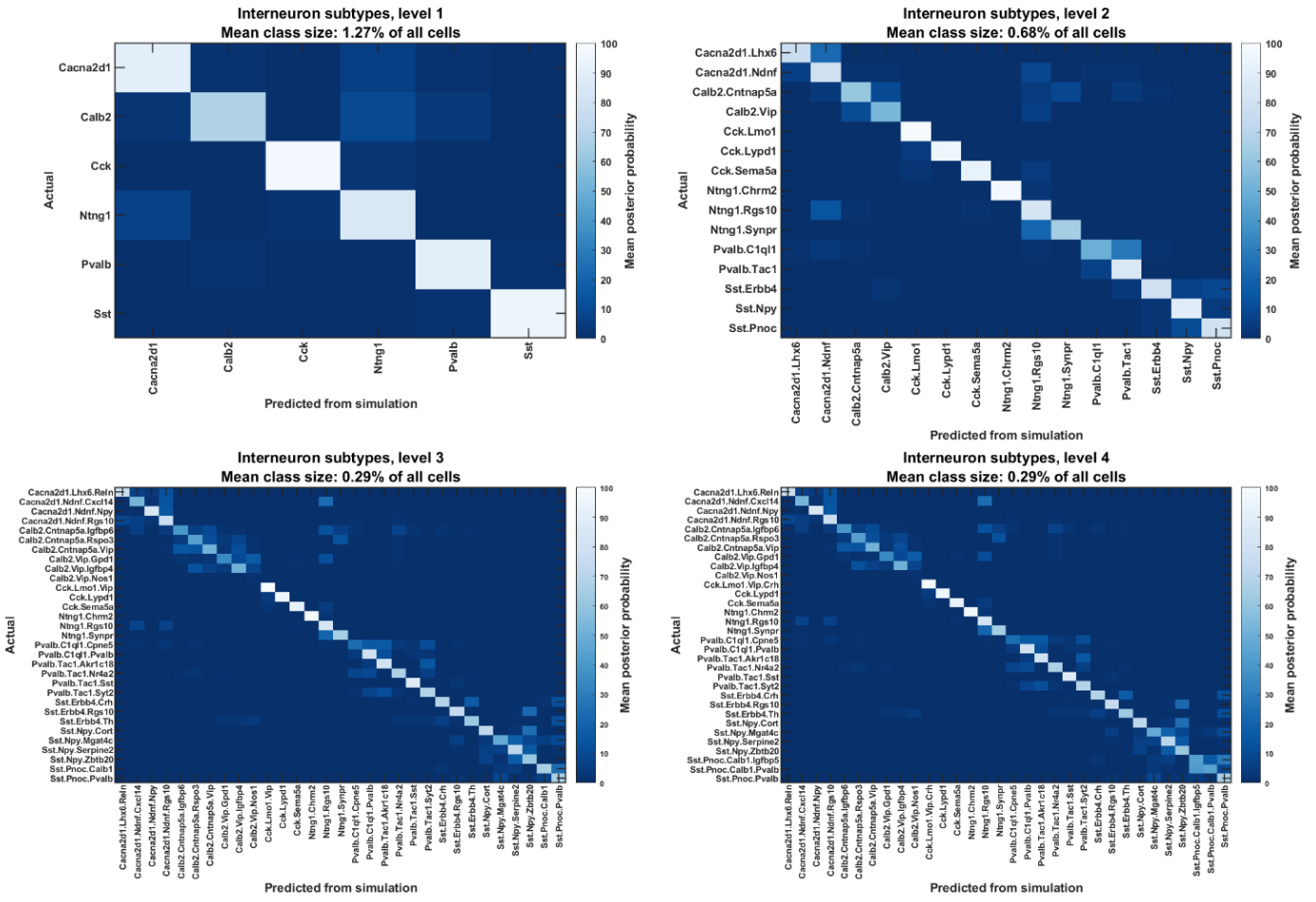
**cortical excitatory neurons**



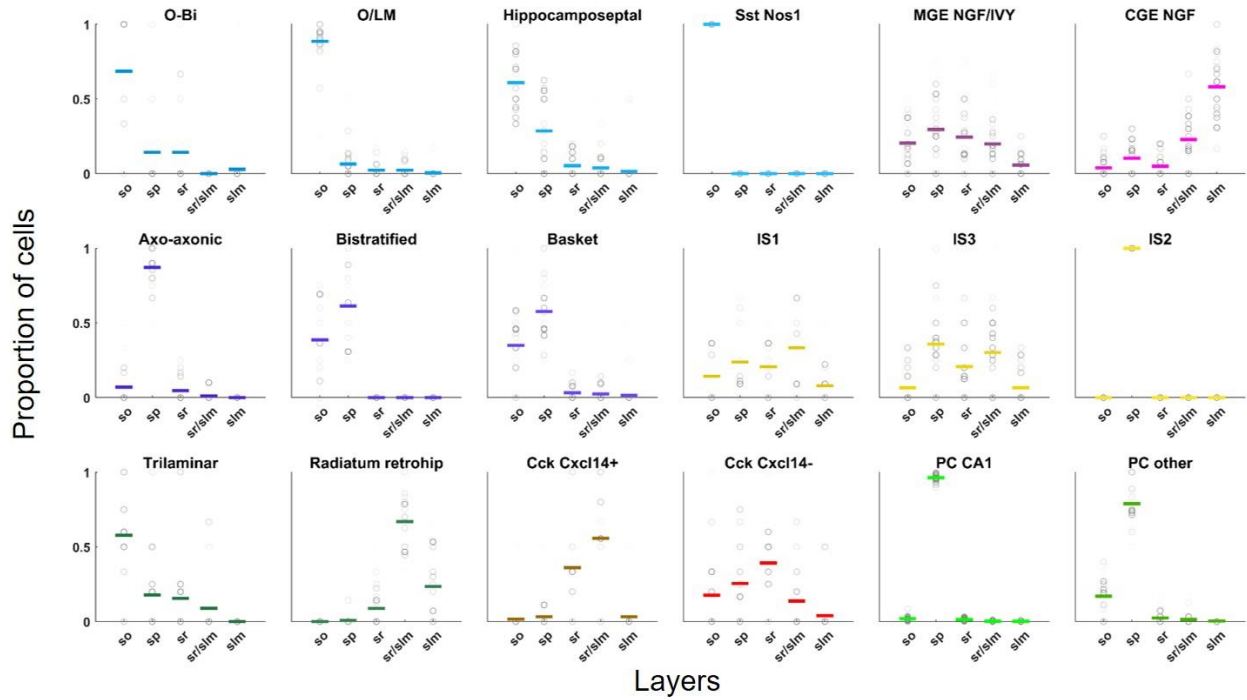
**cortical inhibitory neurons**



**Supplementary Figure S17.** Cell calling in isocortex. **A)** A map of all scRNAseq defined neuron types annotated to be present in isocortex according to Ref. <sup>27</sup> in one coronal section of a mouse brain. **B)** Comparison of frequency of the different cell types in the map between left vs. right hemisphere. **C)** An enlarged view of the cell map in A where only the 15 excitatory neuron types are shown. **D)** The corresponding view of the 10 inhibitory neuron types and non-neurons. **E)** Zoom in of the boxed area in (C) and (D). Note the match of cortical excitatory types to their predicted layers.



**Supplementary Figure S18.** Cell calling performance with only genes suggested by automatic selection algorithm. The relatively high accuracy suggests that manual curation of gene panel is not a pre-requirement for pciSeq. The calculation is carried out the same way as in **Supplementary Figure S15**. The median posterior probability for a cell to belong to the same class as when manually selected genes were also included, was 87%, 76%, 65% and 64%, for four levels of subdivision.



**Supplementary Figure S19.** Stability of cell calling algorithm. The results remain the same with six low-ranking genes (*Atp1b1*, *Slc24a2*, *Tmsb10*, *Calm2*, *Gap43*, *Fxyd6*) excluded from cell calling, as verified by the consistent laminar attribution of each cell class. Circles indicate means of a single experiment with gray level representing number of cells of that class in the experiment; colored lines denote grand mean over all n=28 hippocampi. In each plot, the 5 x-axis positions represent layers: stratum oriens (so), stratum pyramidale (sp), stratum radiatum (sr), border of strata radiatum and lacunosum-moleculare (sr/slm), stratum lacunosum-moleculare (slm). MGE: medial ganglionic eminence. CGE: caudal ganglionic eminence. NGF: neurogliaform. IS: interneuron-selective cells.

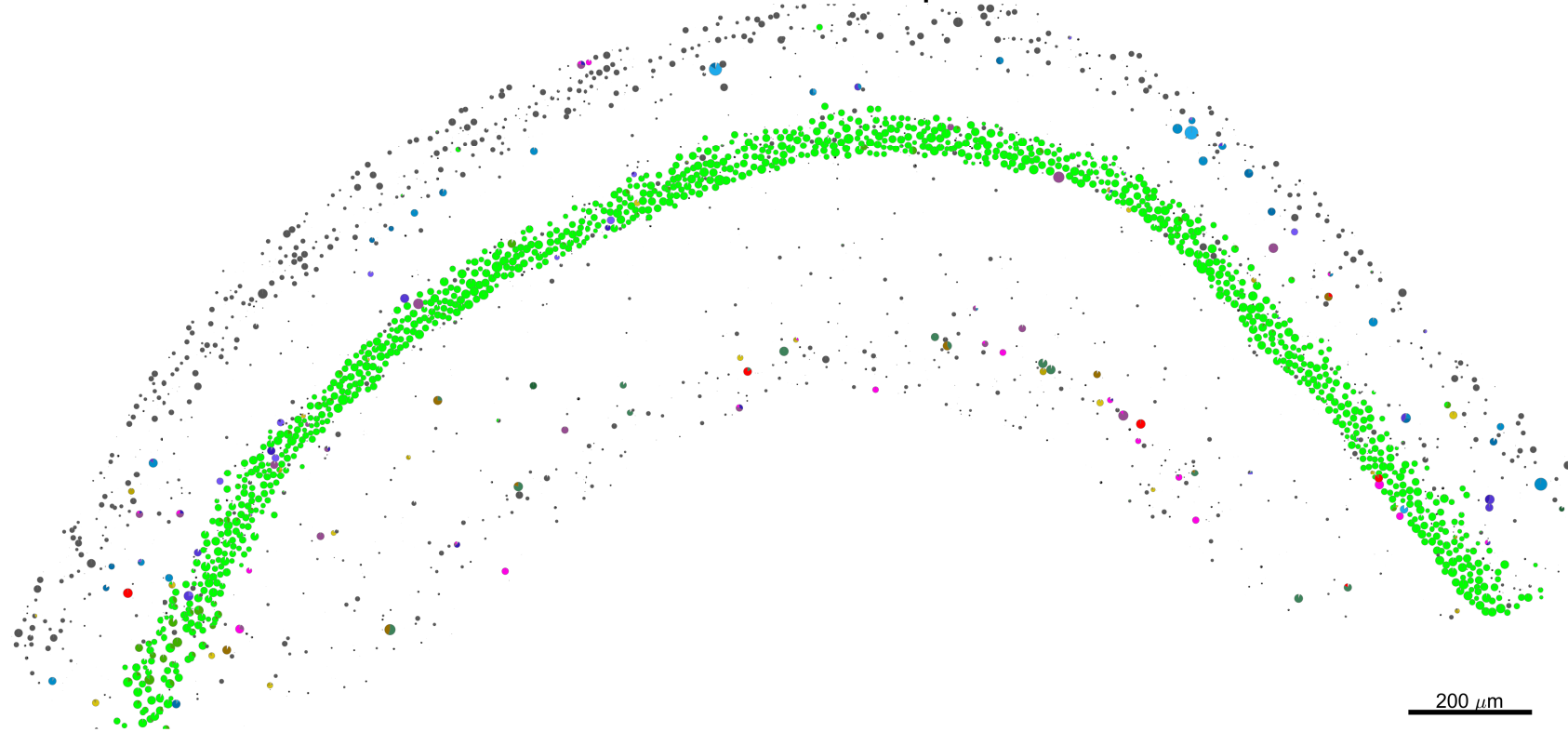
## References

1. Katona, L. *et al.* Sleep and movement differentiates actions of two types of somatostatin-expressing GABAergic interneuron in rat hippocampus. *Neuron* **82**, 872–86 (2014).
2. Jinno, S. *et al.* Neuronal diversity in GABAergic long-range projections from the hippocampus. *J Neurosci* **27**, 8790–804 (2007).
3. Paul, A. *et al.* Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity. *Cell* **171**, 522–539.e20 (2017).
4. Viney, T. J. *et al.* Network state-dependent inhibition of identified hippocampal CA3 axo-axonic cells in vivo. *Nat. Neurosci.* **16**, 1802–1811 (2013).
5. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
6. Harris, K. D. *et al.* Classes and continua of hippocampal CA1 inhibitory neurons revealed by single-cell transcriptomics. *PLoS Biol.* **16**, e2006387 (2018).
7. Fuentealba, P. *et al.* Ivy cells: a population of nitric-oxide-producing, slow-spiking GABAergic neurons and their involvement in hippocampal network activity. *Neuron* **57**, 917–29 (2008).
8. Tricoire, L. *et al.* Common origins of hippocampal Ivy and nitric oxide synthase expressing neurogliaform cells. *J Neurosci* **30**, 2165–76 (2010).
9. Yamawaki, N. *et al.* Long-range inhibitory intersection of a retrosplenial thalamocortical circuit by apical tuft-targeting CA1 neurons. *Nat. Neurosci.* **22**, 618–626 (2019).
10. Acsady, L., Gorcs, T. J. & Freund, T. F. Different populations of vasoactive intestinal polypeptide-immunoreactive interneurons are specialized to control pyramidal cells or interneurons in the hippocampus. *Neuroscience* **73**, 317–34 (1996).

11. Katona, I. *et al.* Presynaptically located CB1 cannabinoid receptors regulate GABA release from axon terminals of specific hippocampal interneurons. *J Neurosci* **19**, 4544–58 (1999).
12. Blasco-Ibanez, J. M., Martinez-Guijarro, F. J. & Freund, T. F. Enkephalin-containing interneurons are specialized to innervate other interneurons in the hippocampal CA1 region of the rat and guinea-pig. *Eur J Neurosci* **10**, 1784–95 (1998).
13. Gulyás, A. I., Hájos, N. & Freund, T. F. Interneurons containing calretinin are specialized to control other interneurons in the rat hippocampus. *J. Neurosci. Off. J. Soc. Neurosci.* **16**, 3397–3411 (1996).
14. Klausberger, T. *et al.* Spike timing of dendrite-targeting bistratified cells during hippocampal network oscillations *in vivo*. *Nat. Neurosci.* **7**, 41–47 (2004).
15. Losonczy, A., Zhang, L., Shigemoto, R., Somogyi, P. & Nusser, Z. Cell type dependence and variability in the short-term plasticity of EPSCs in identified mouse hippocampal interneurons. *J. Physiol.* **542**, 193–210 (2002).
16. Buhl, E. H. *et al.* Physiological properties of anatomically identified axo-axonic cells in the rat hippocampus. *J Neurophysiol* **71**, 1289–307 (1994).
17. Ferraguti, F. *et al.* Metabotropic glutamate receptor 8-expressing nerve terminals target subsets of GABAergic neurons in the hippocampus. *J Neurosci* **25**, 10520–36 (2005).
18. Sik, A., Penttonen, M., Ylinen, A. & Buzsáki, G. Hippocampal CA1 interneurons: an *in vivo* intracellular labeling study. *J Neurosci* **15**, 6651–65 (1995).
19. Miyashita, T. & Rockland, K. S. GABAergic projections from the hippocampus to the retrosplenial cortex in the rat. *Eur J Neurosci* **26**, 1193–204 (2007).

20. Jinno, S. Structural organization of long-range GABAergic projection system of the hippocampus. *Front Neuroanat* **3**, 13 (2009).
21. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
22. Harris, K. *et al.* Molecular organization of CA1 interneuron classes. *bioRxiv* 034595 (2015) doi:10.1101/034595.
23. The FANTOM Consortium and the RIKEN PMI and Clst (dgt) *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).
24. Jinno, S. & Kosaka, T. Patterns of colocalization of neuronal nitric oxide synthase and somatostatin-like immunoreactivity in the mouse hippocampus: quantitative analysis with optical disector. *Neuroscience* **124**, 797–808 (2004).
25. Sik, A., Ylinen, A., Penttonen, M. & Buzsaki, G. Inhibitory CA1-CA3-hilar region feedback in the hippocampus. *Science* **265**, 1722–4 (1994).
26. Fuentealba, P. *et al.* Rhythmically active enkephalin-expressing GABAergic cells in the CA1 area of the hippocampus project to the subiculum and preferentially innervate interneurons. *J Neurosci* **28**, 10017–22 (2008).
27. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999-1014.e22 (2018).

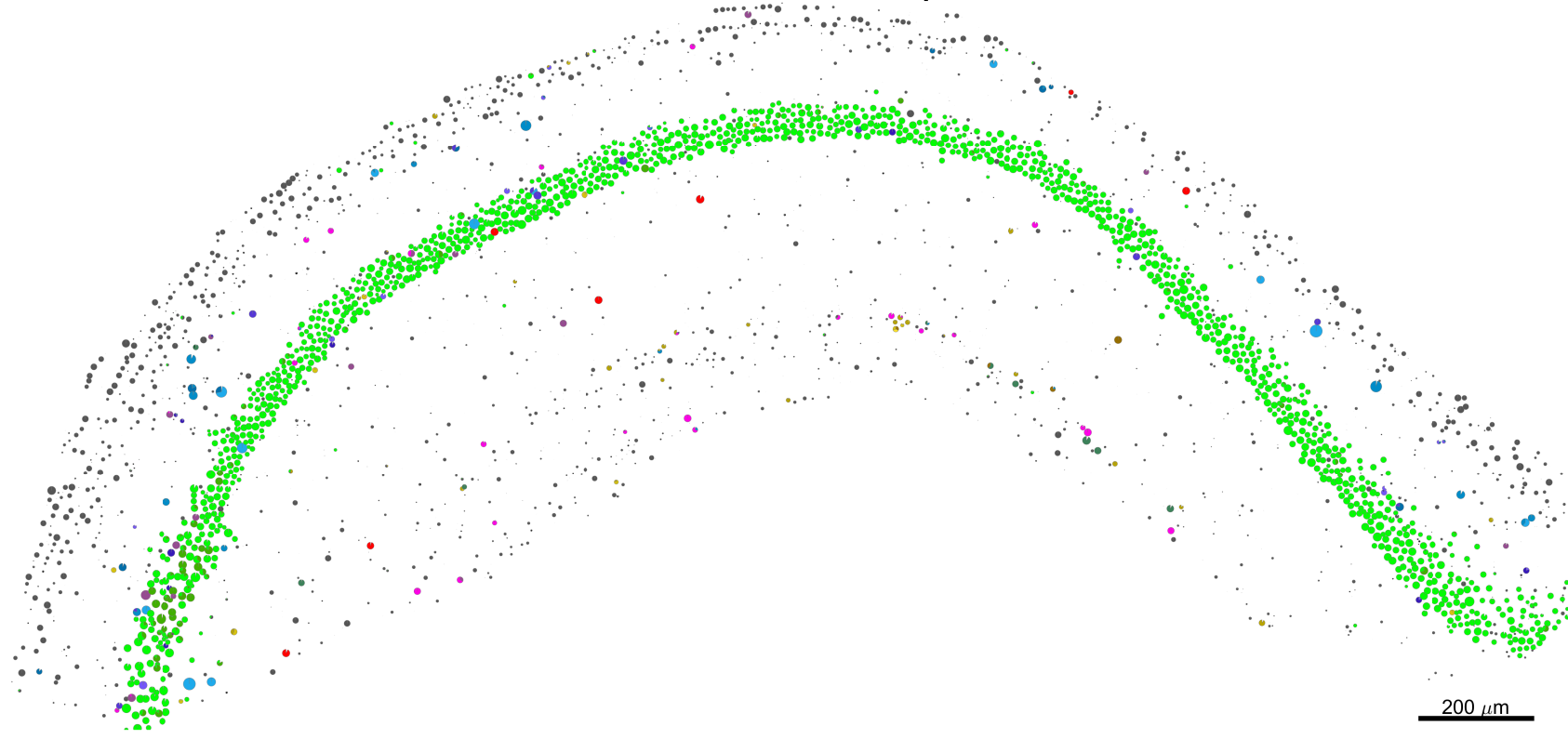
section 7-4 left hemisphere



- O-Bi
- O/LM
- Hippocamposeptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled



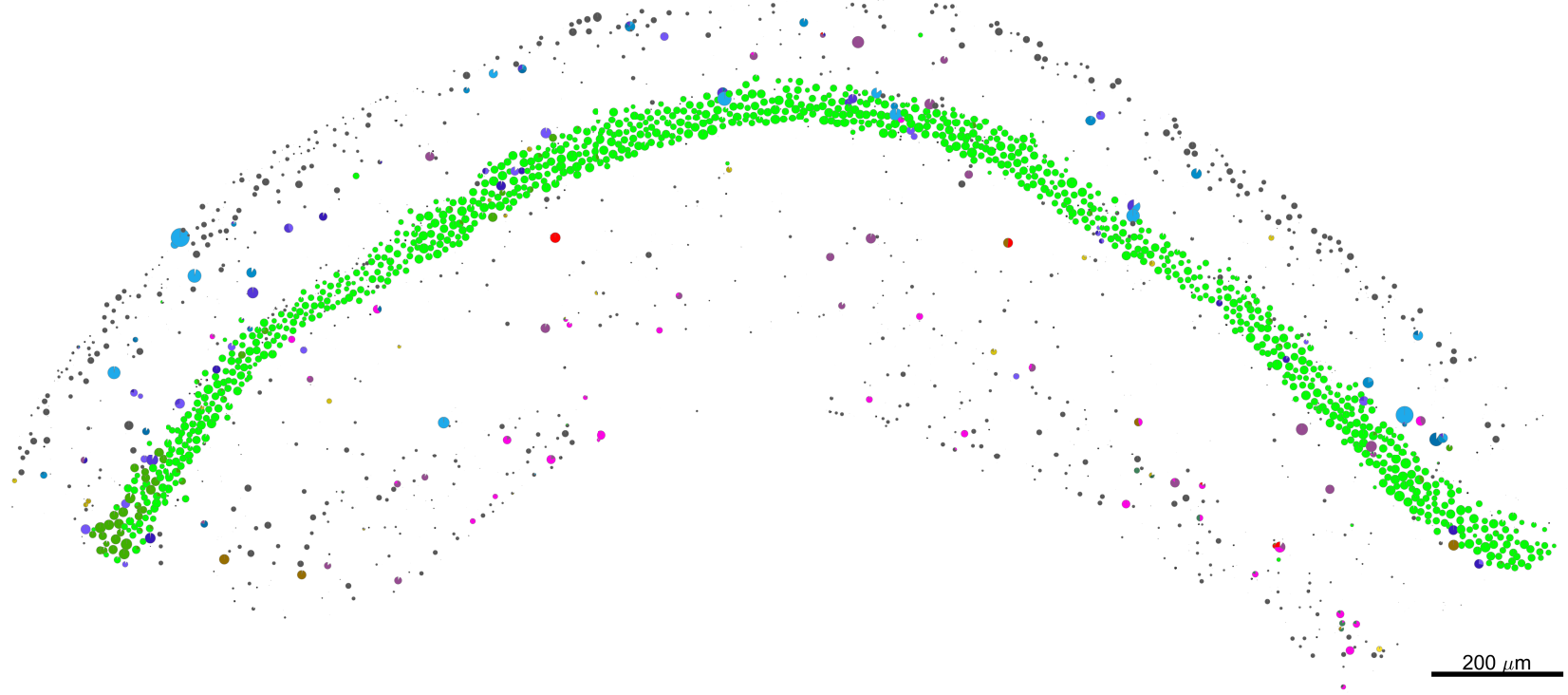
section 5-4 left hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

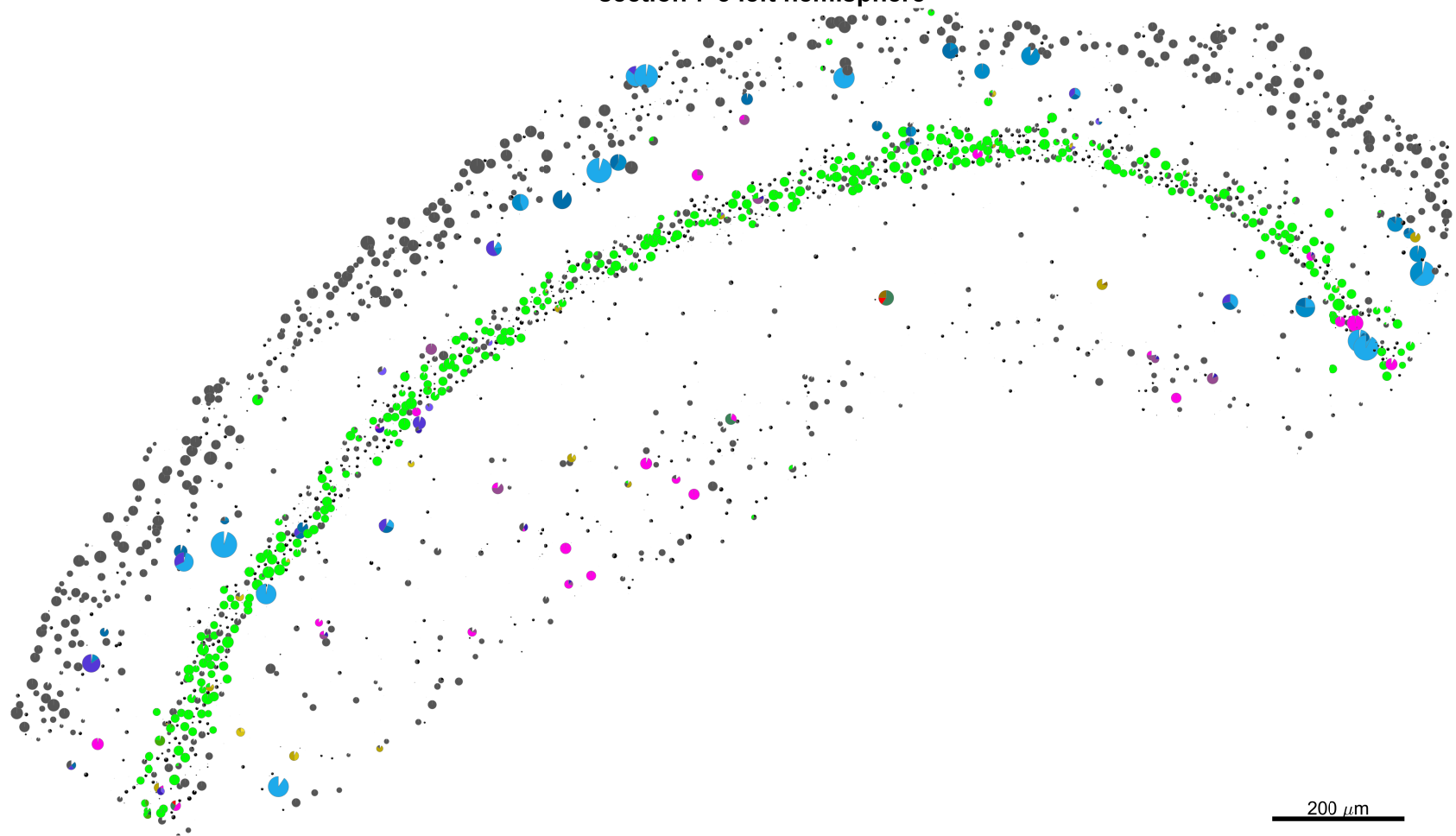
200  $\mu$ m

section 3-4 left hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

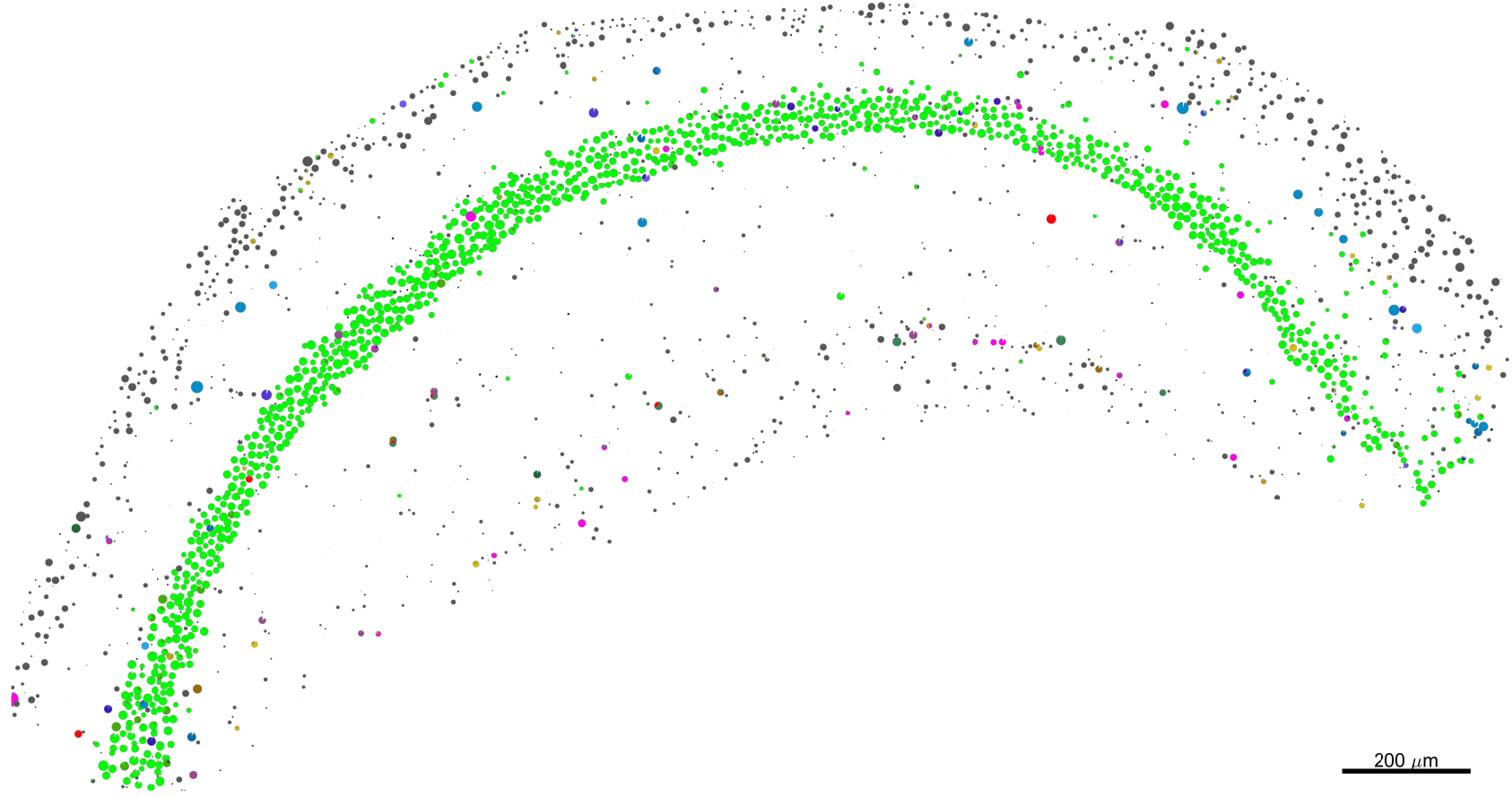
section 7-3 left hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

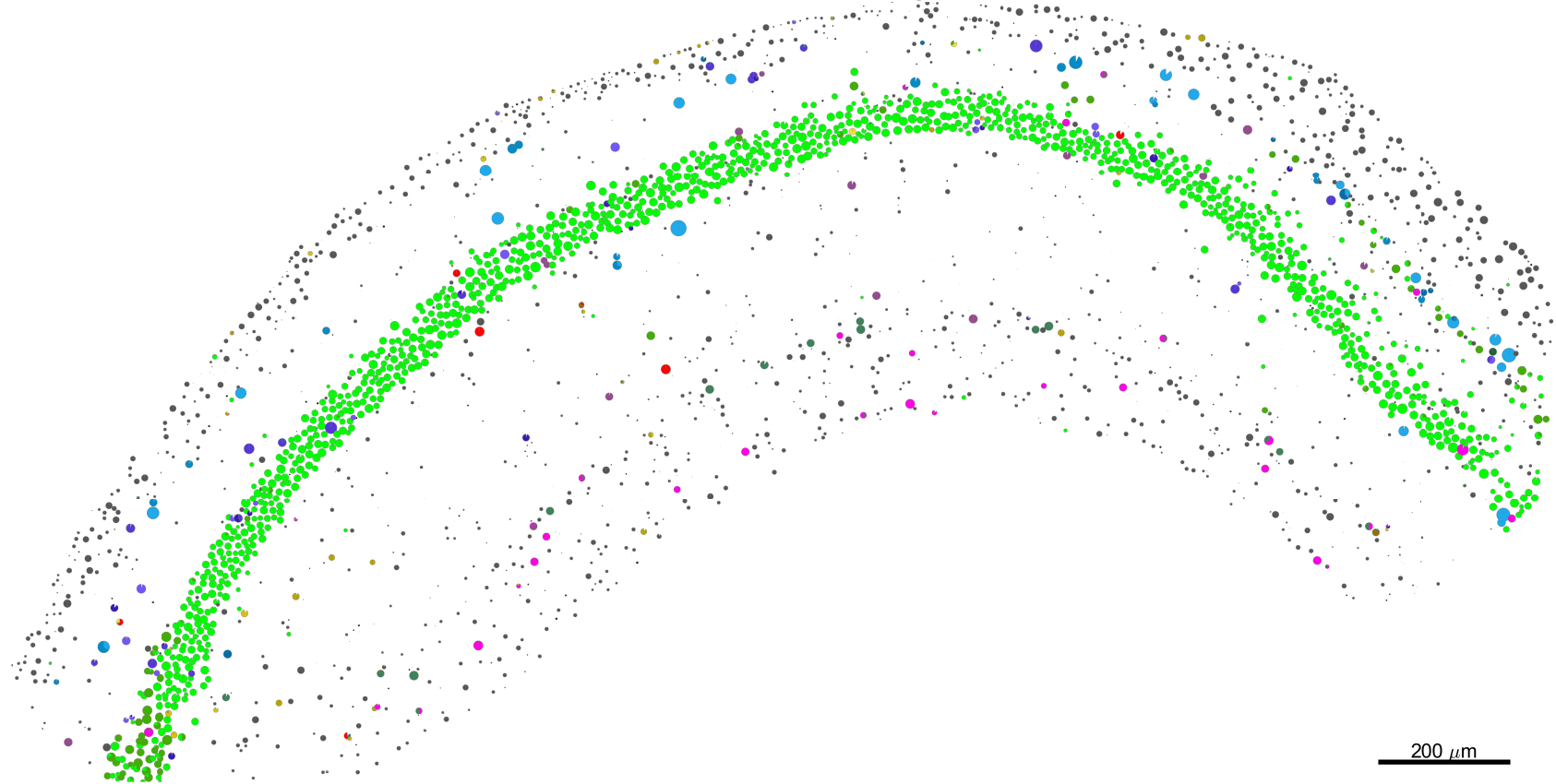
section 5-3 left hemisphere



- O-Bi
- O/LM
- Hippocamposeptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

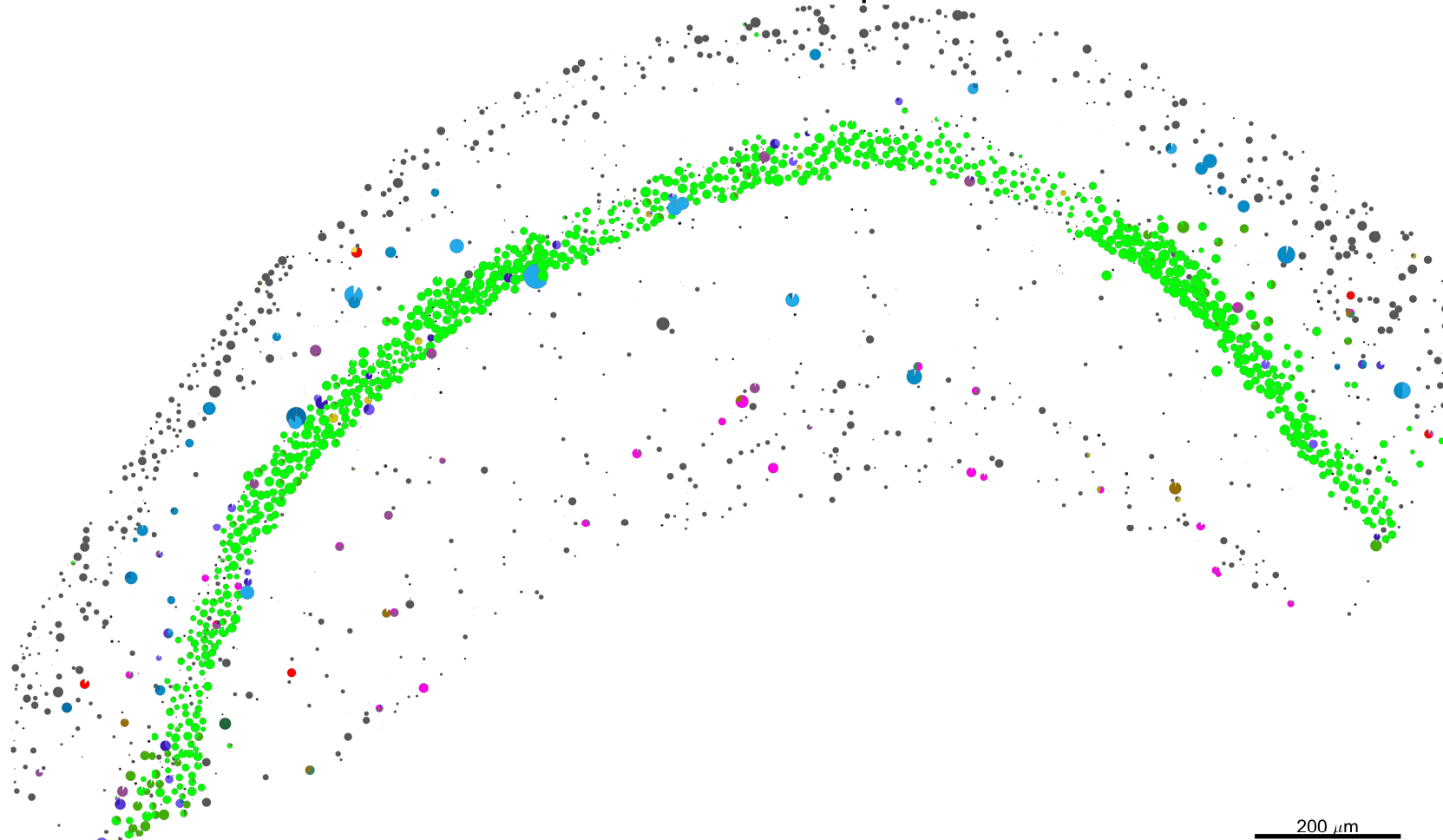
section 4-3 left hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

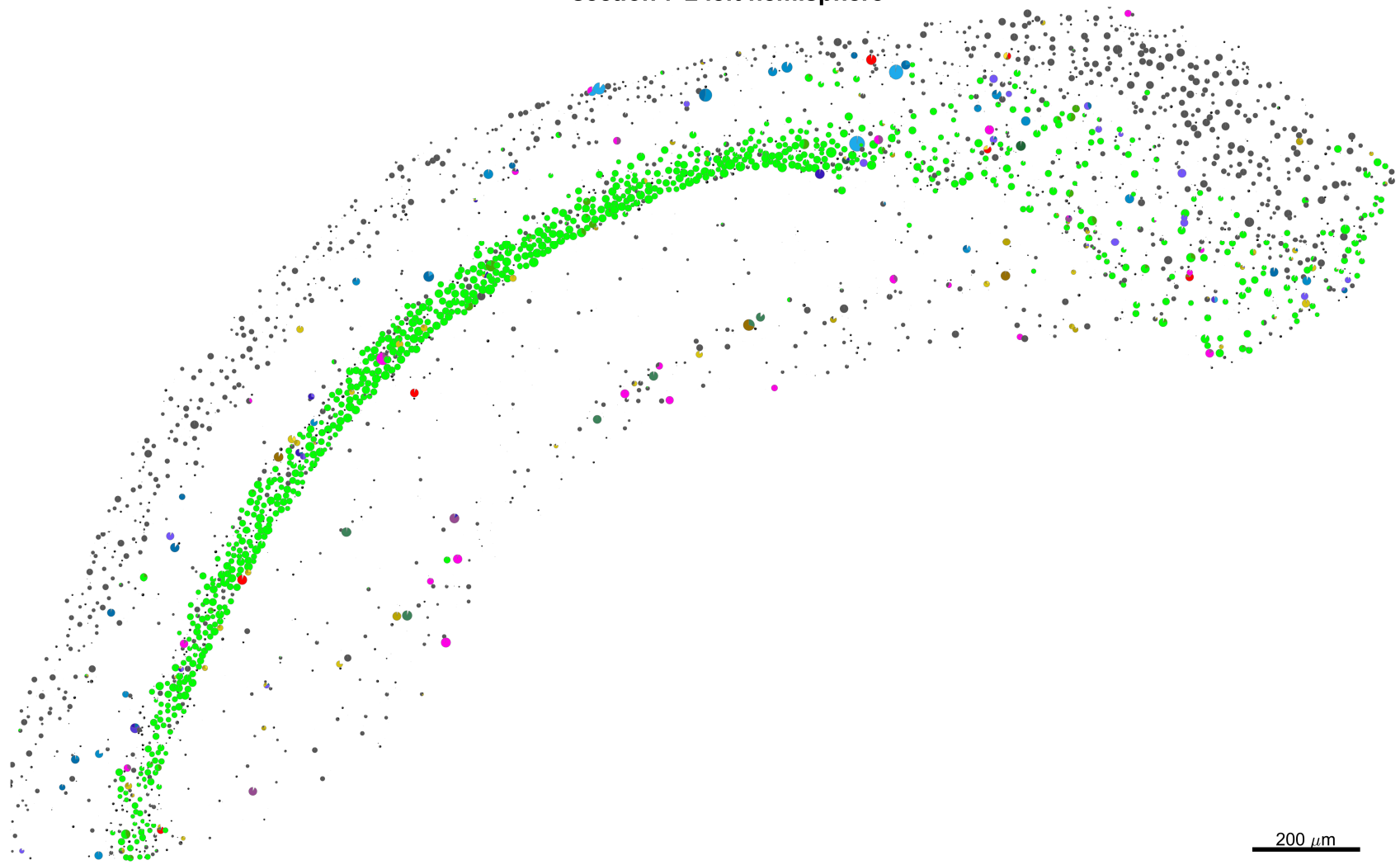
section 3-3 left hemisphere



- O-Bi
- O/LM
- Hippocamposeptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

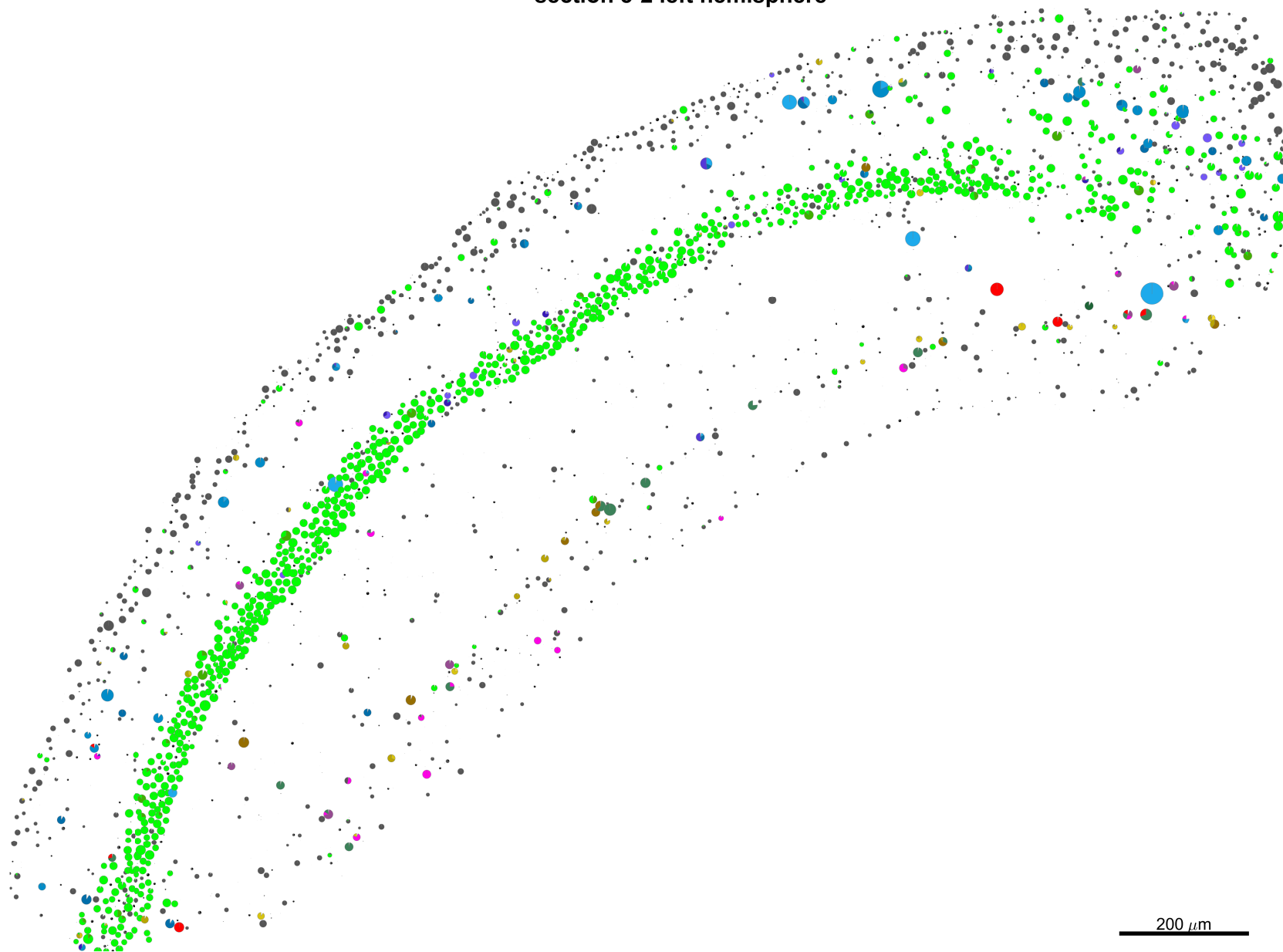
section 7-2 left hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

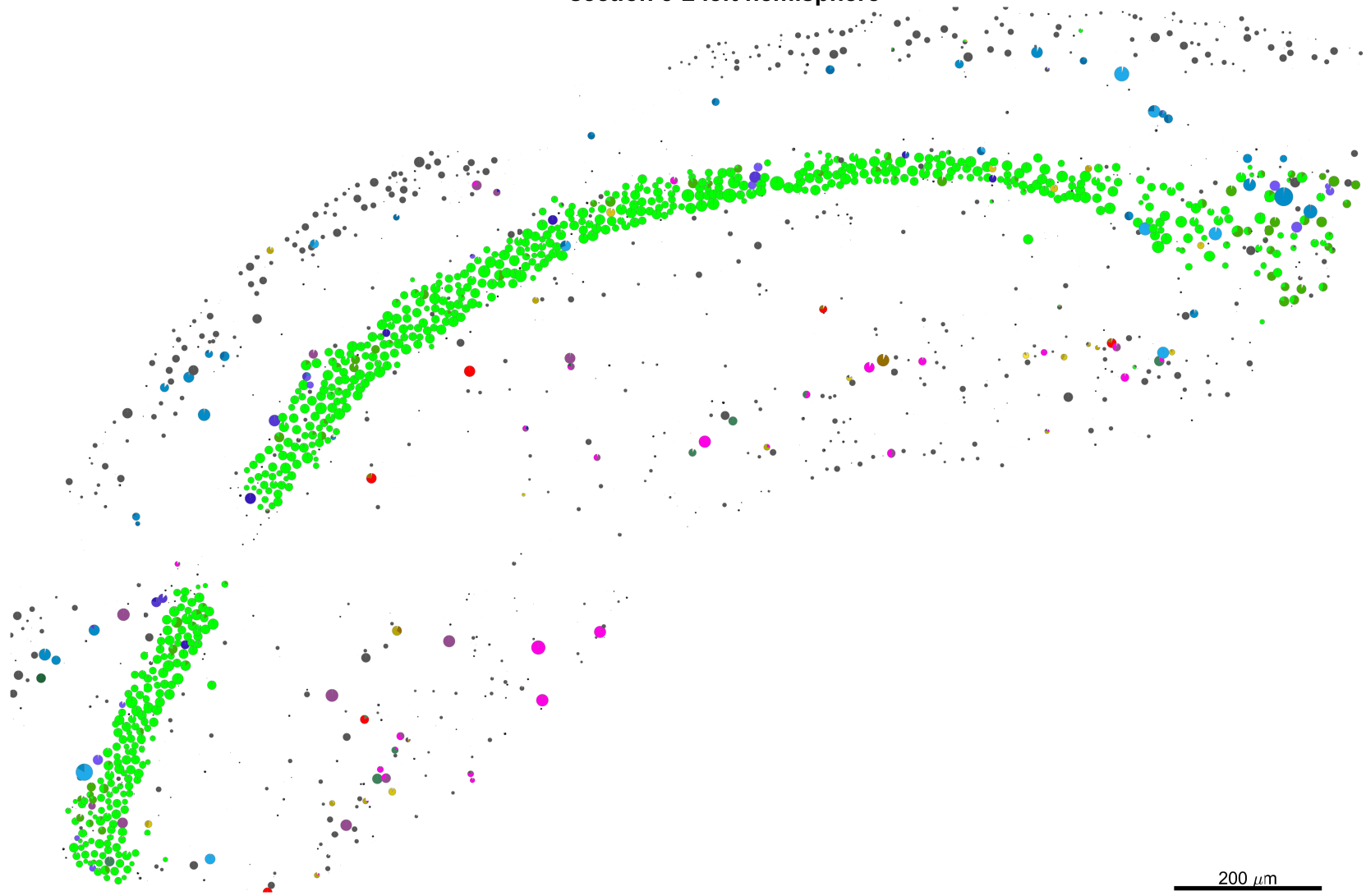
section 5-2 left hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled



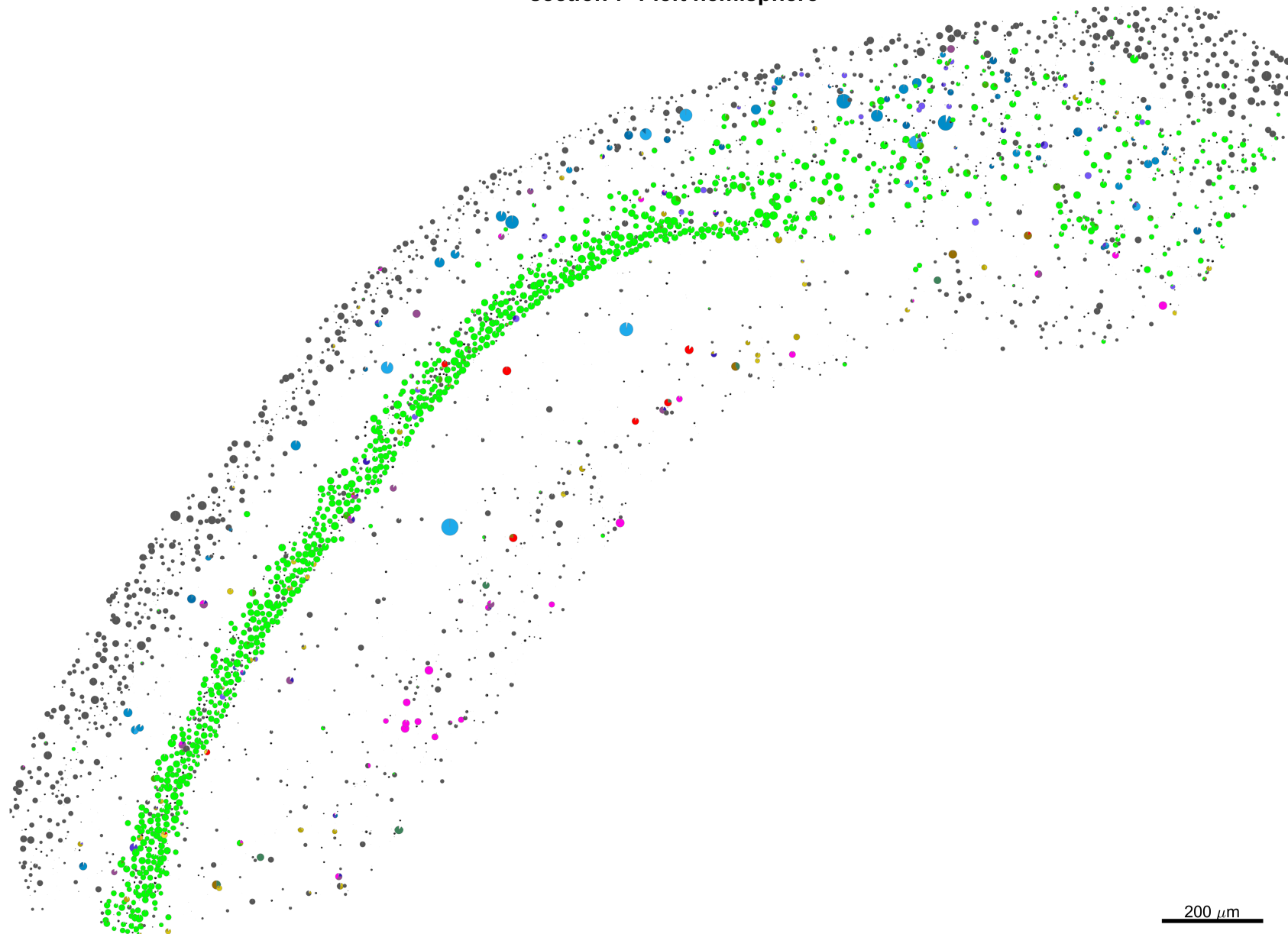
section 3-2 left hemisphere



- O-Bi
- O/LM
- Hippocamoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

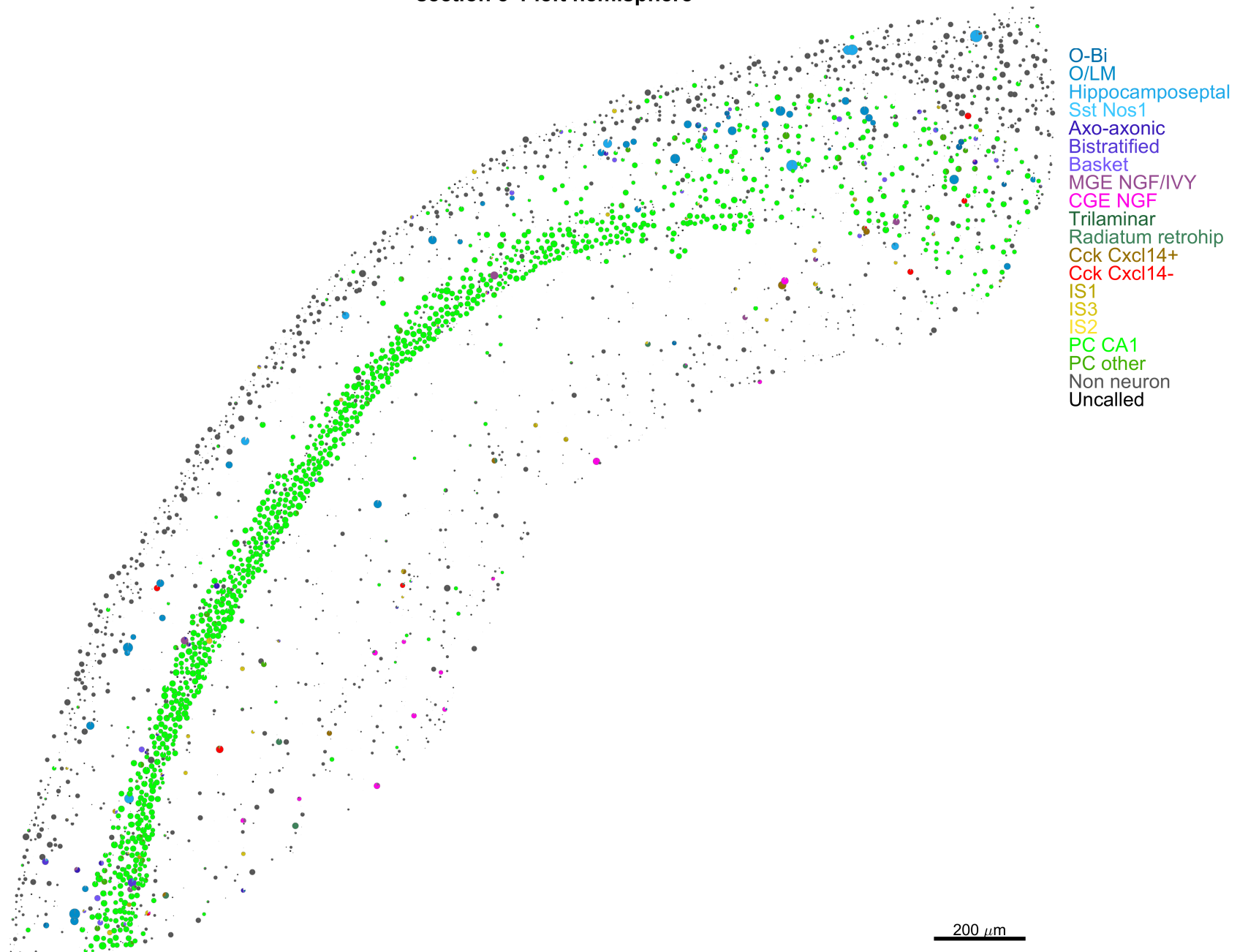
section 7-1 left hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

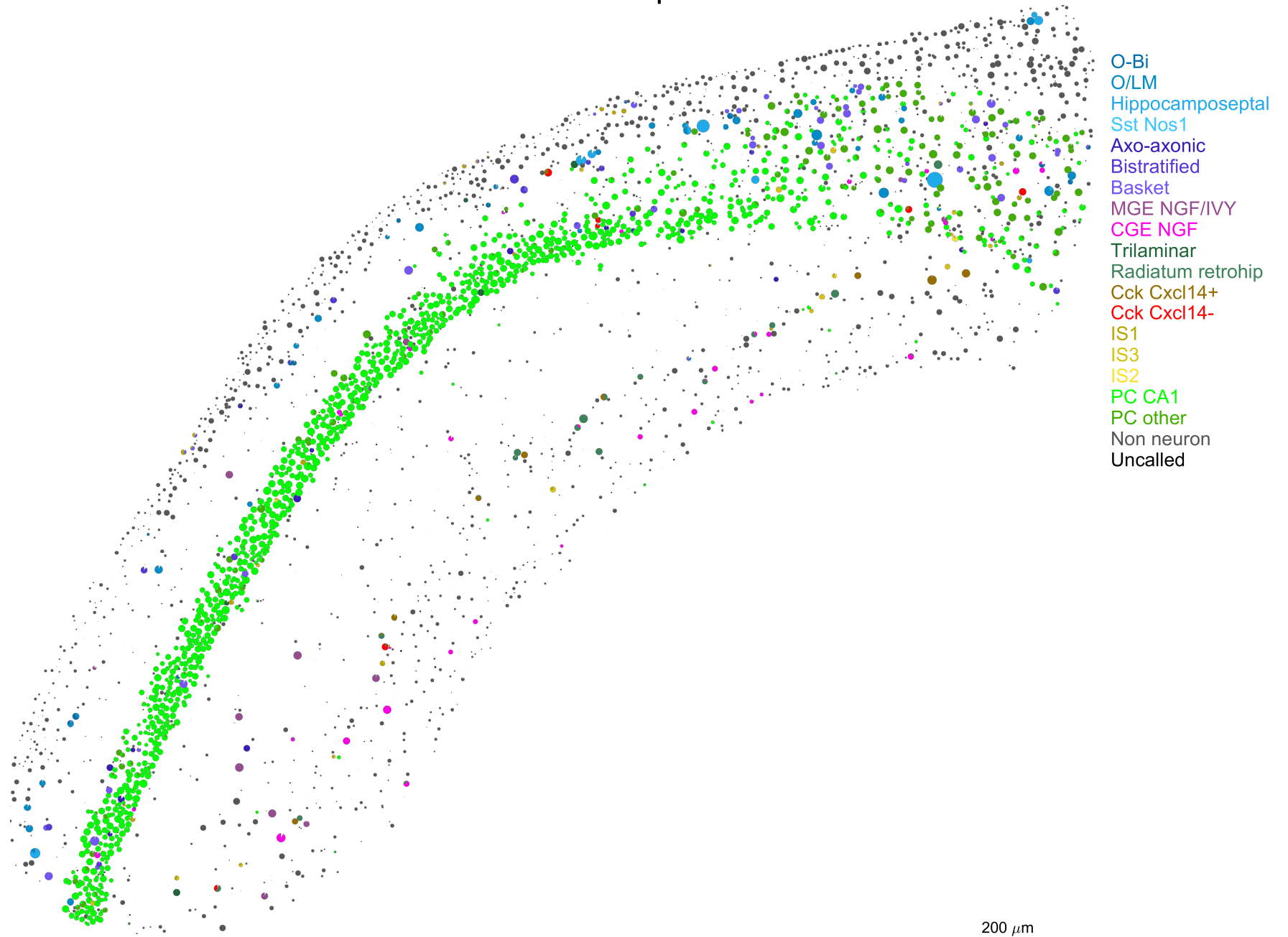
section 5-1 left hemisphere



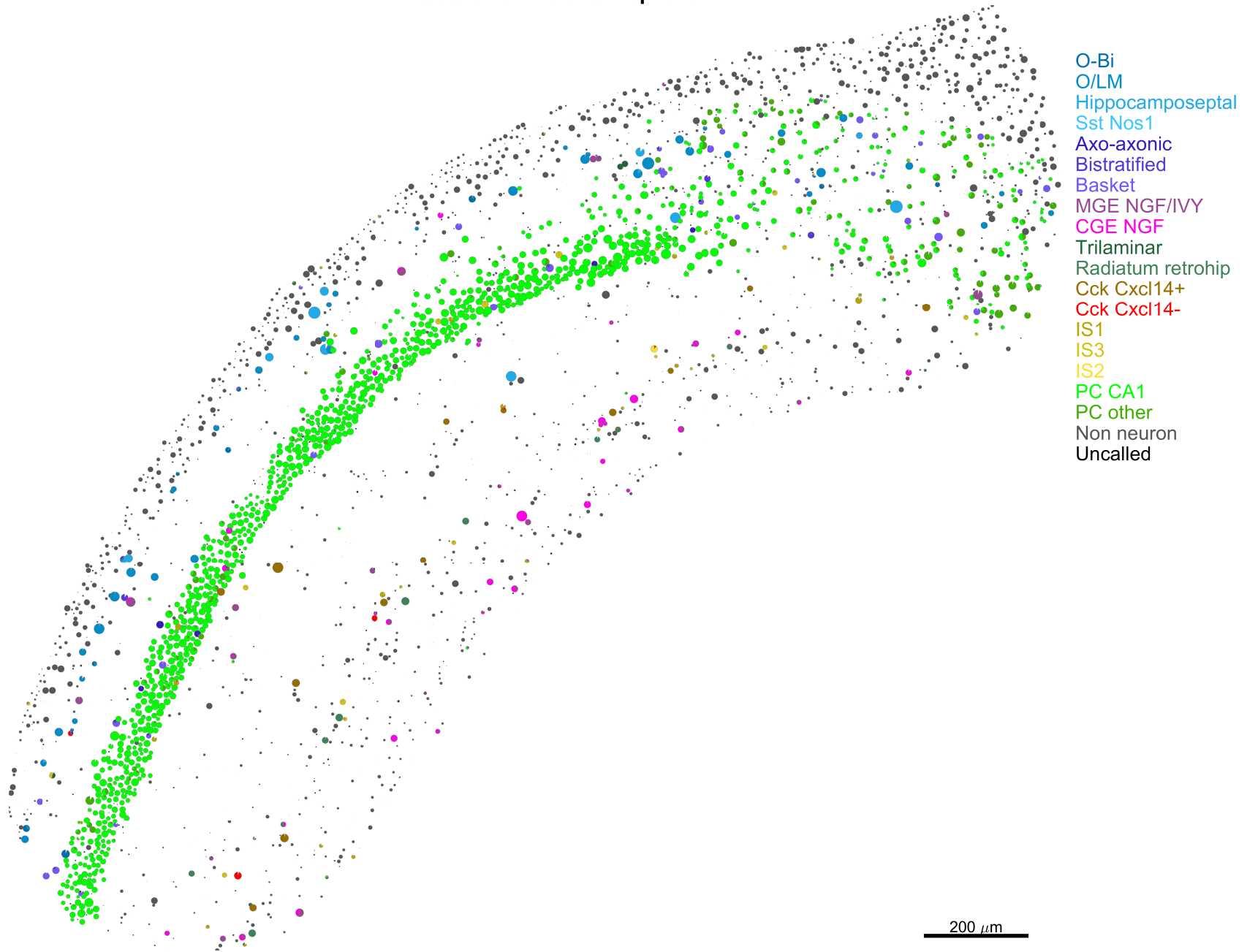
- O-Bi
- O/LM
- Hippamoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

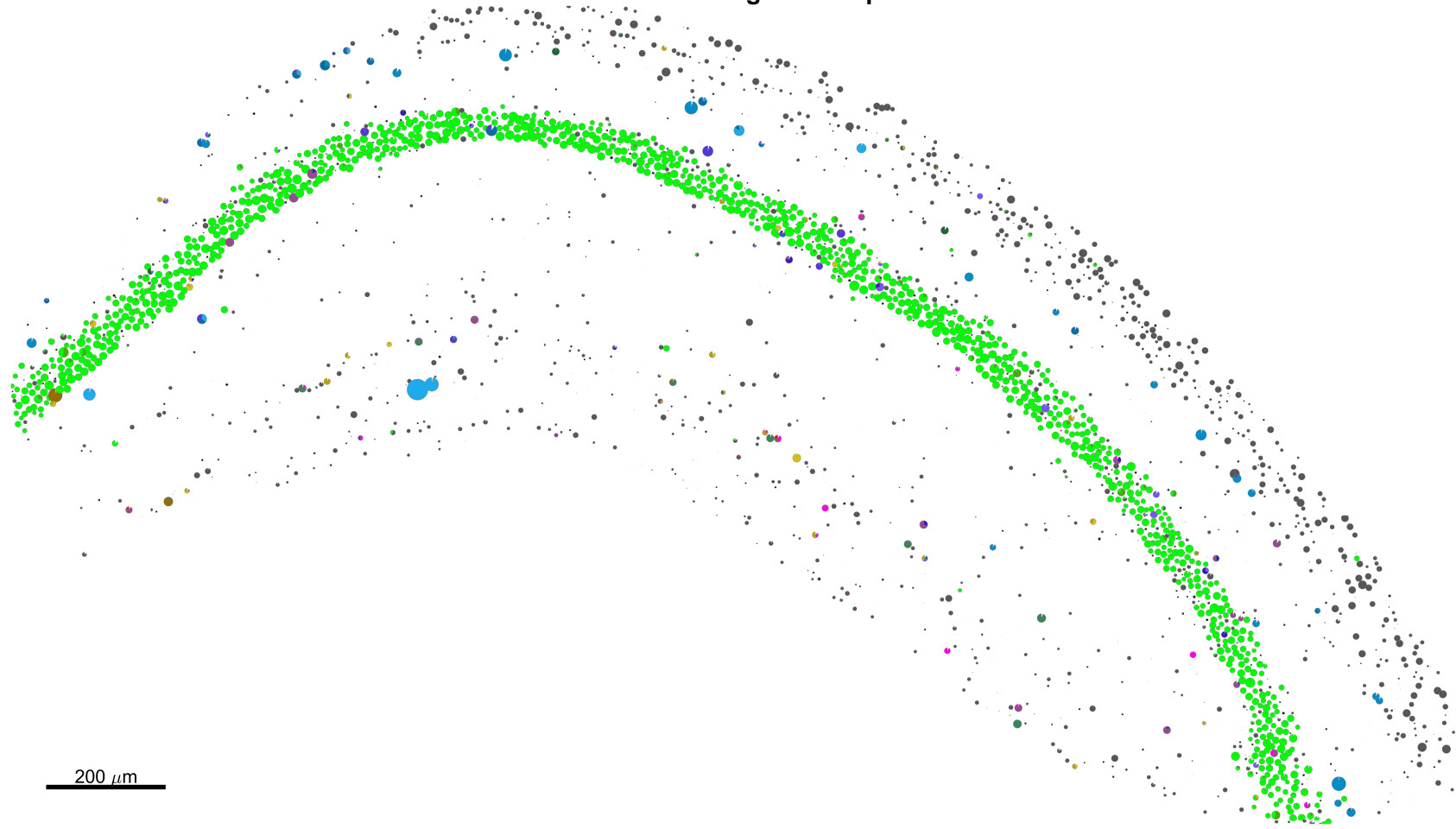
section 4-1 left hemisphere



section 3-1 left hemisphere



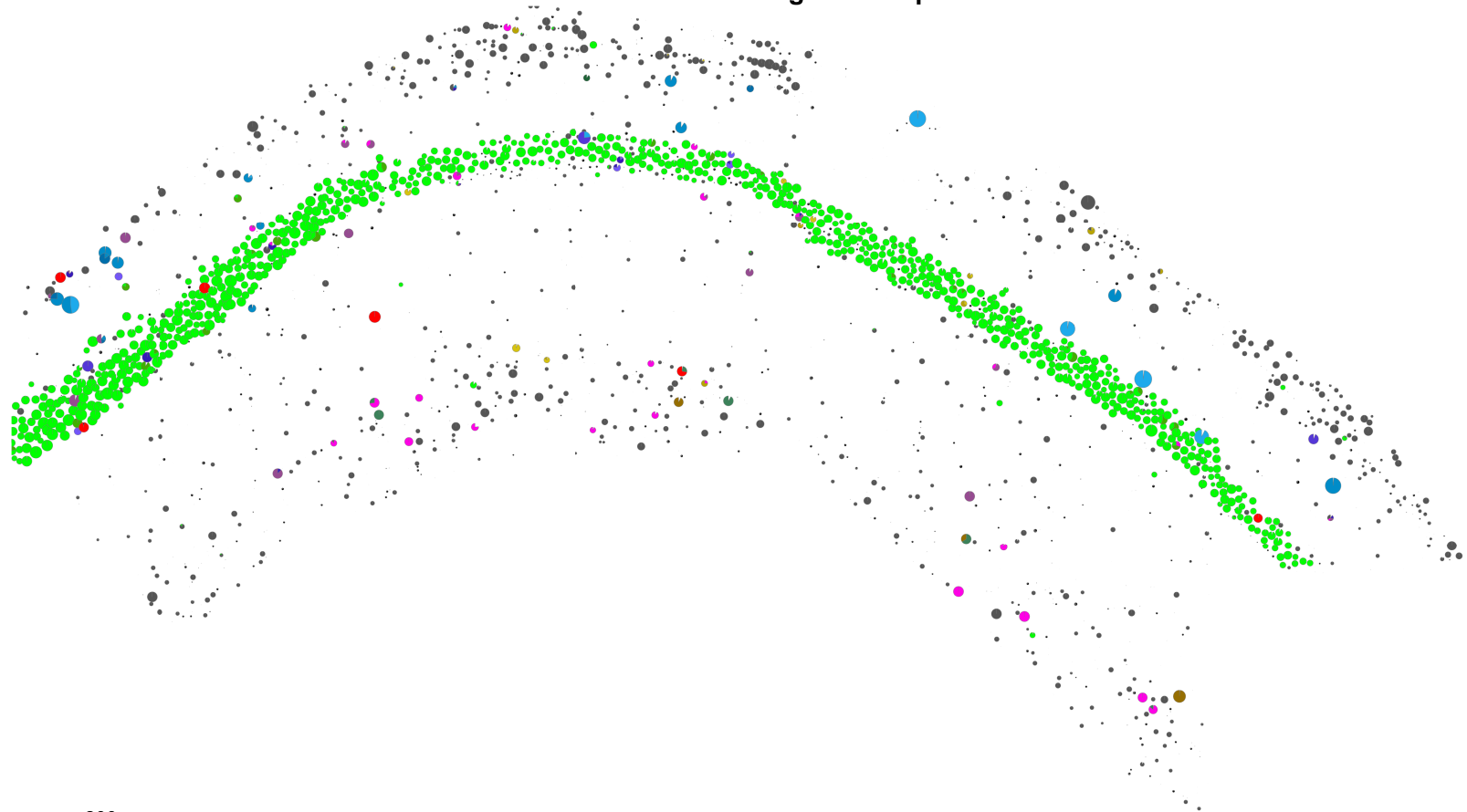
section 7-4 right hemisphere



- O-Bi
- O/LM
- Hippocamposeptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

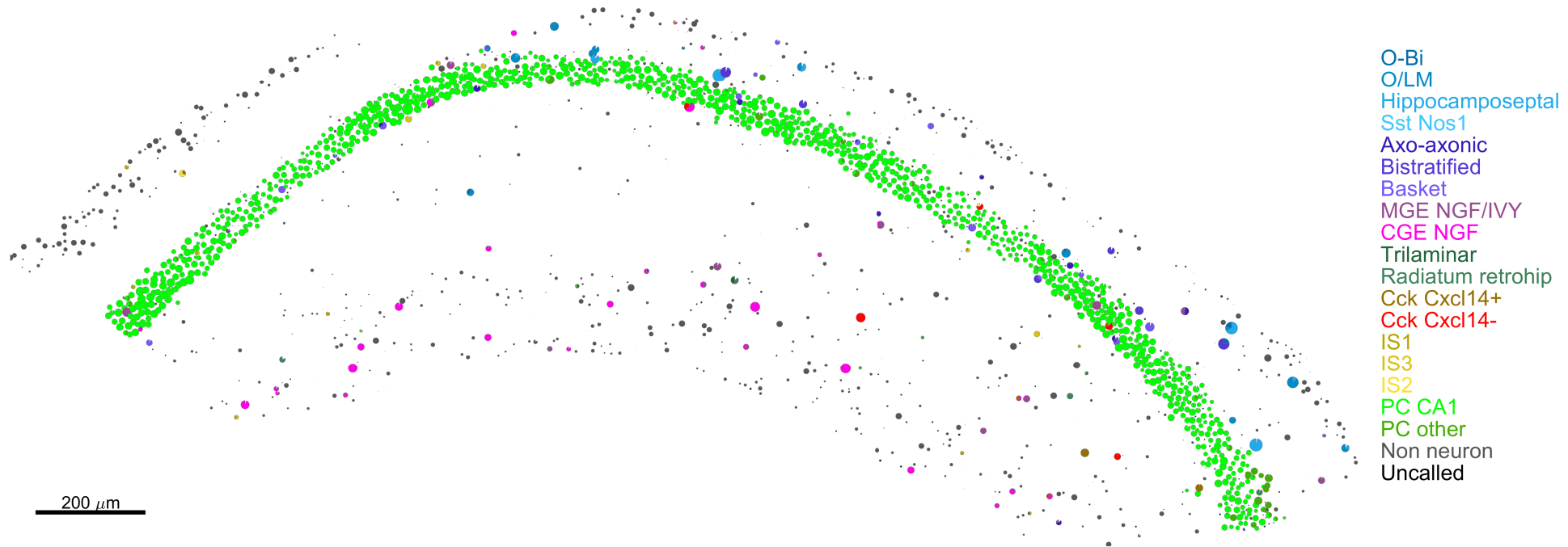
section 5-4 right hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

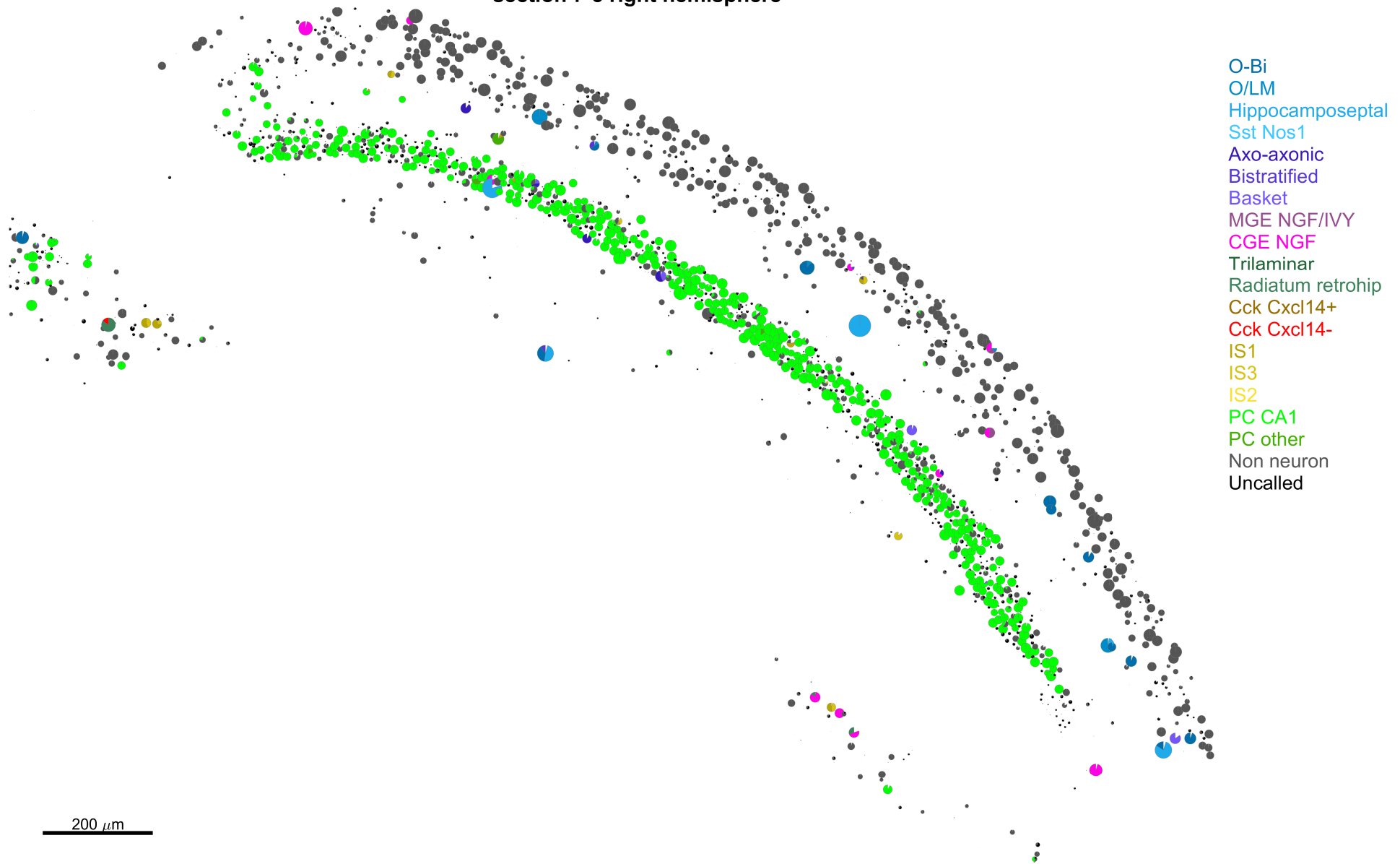
200  $\mu$ m

section 3-4 right hemisphere

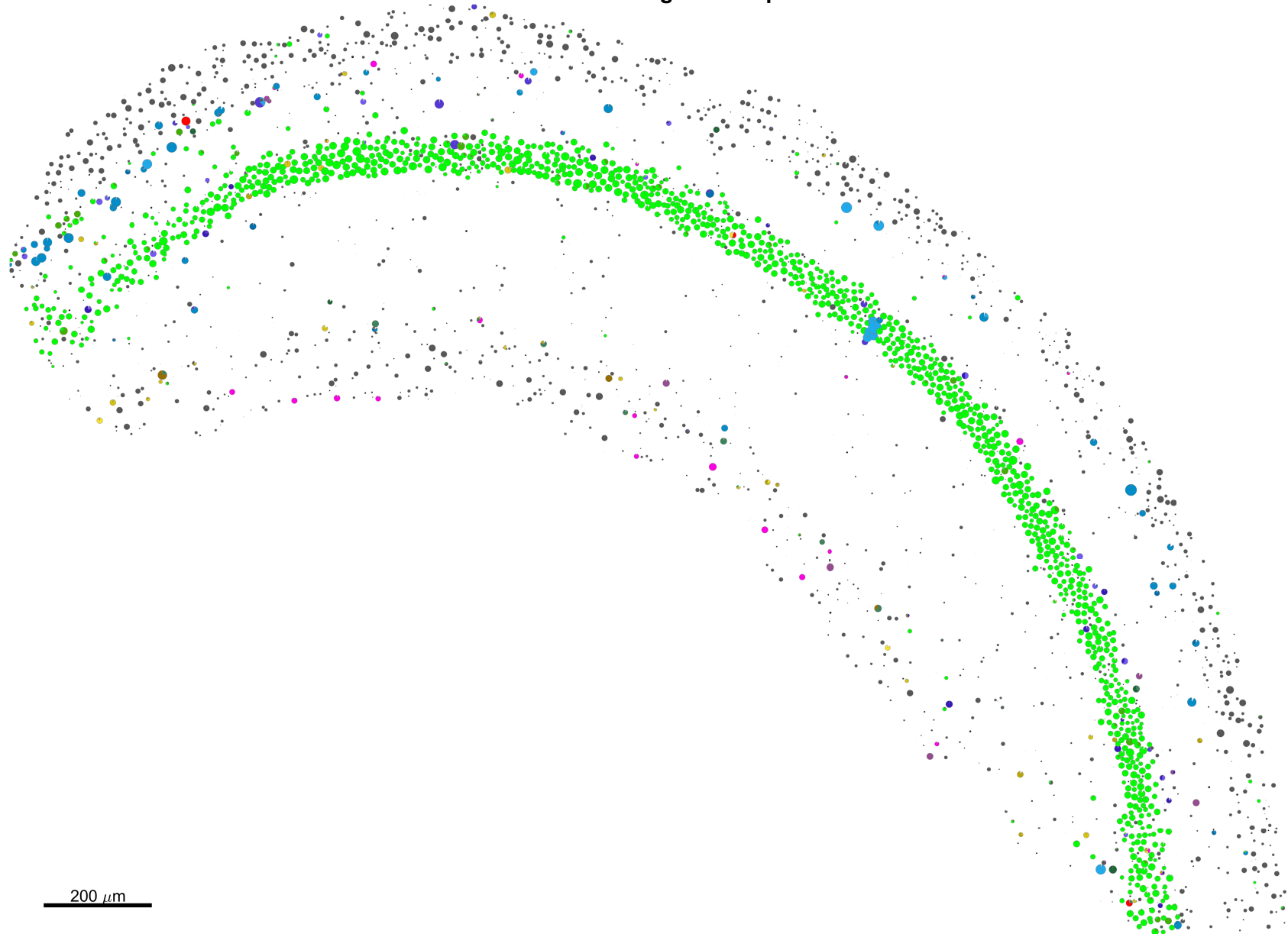




section 7-3 right hemisphere



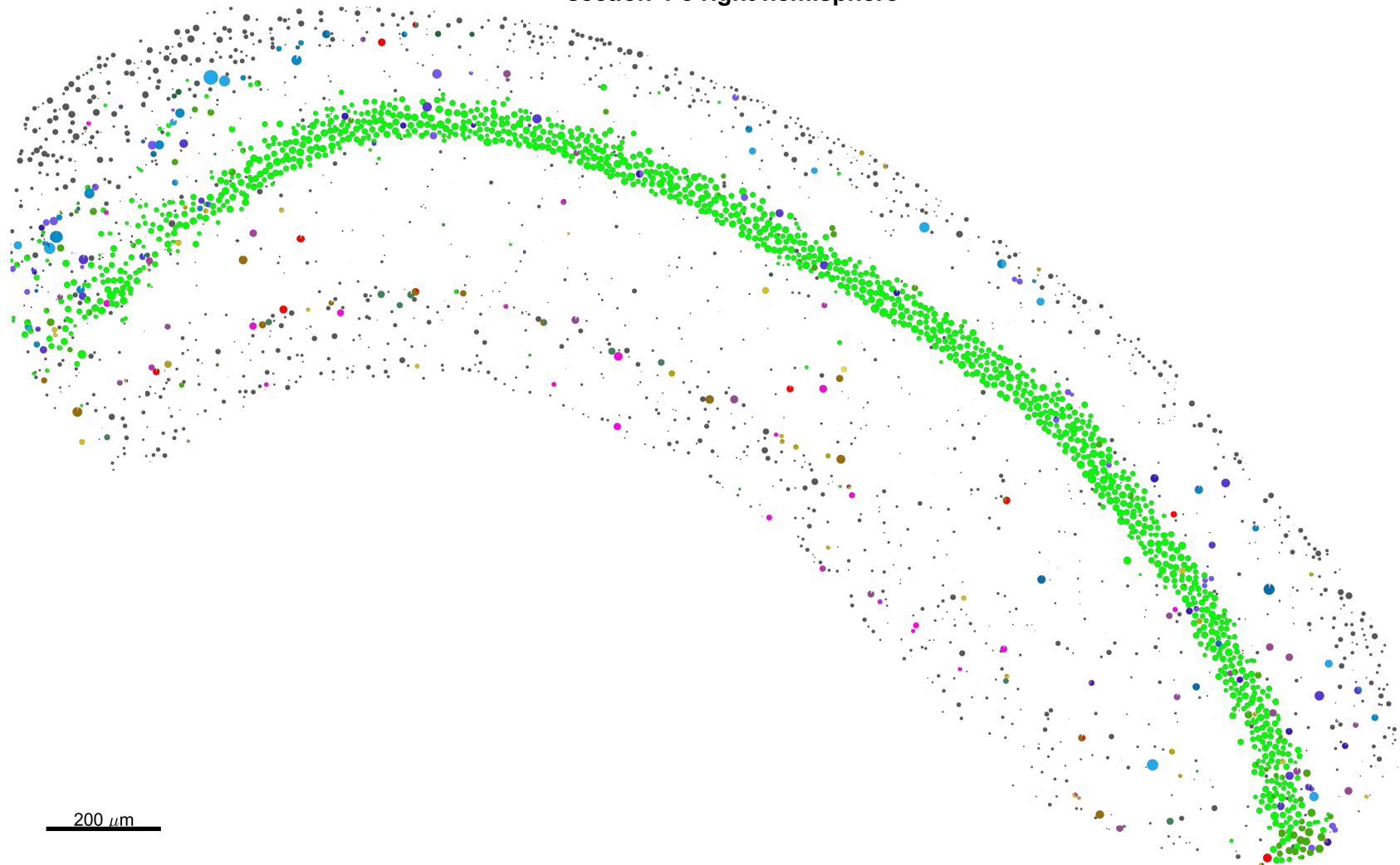
section 5-3 right hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

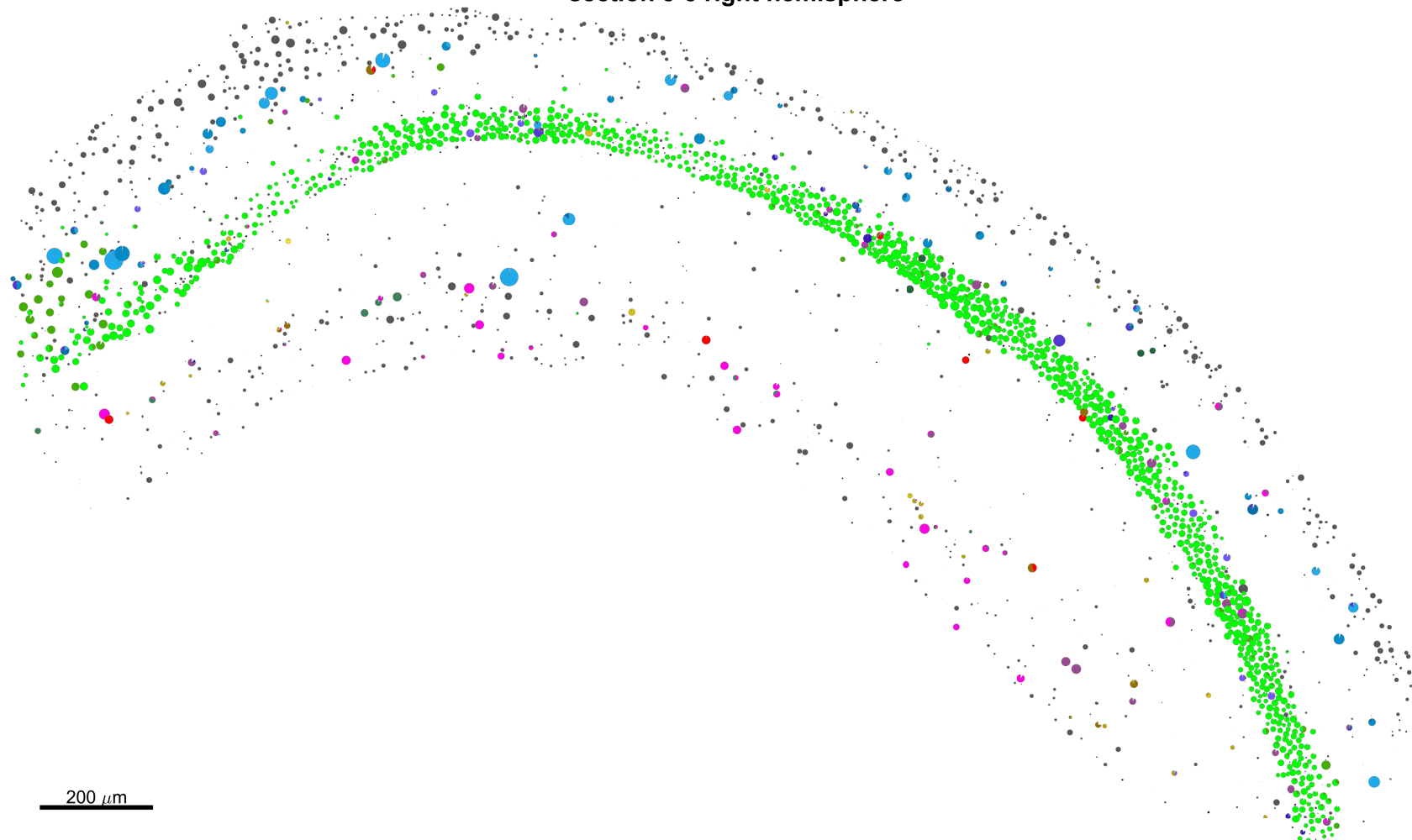
200  $\mu$ m

section 4-3 right hemisphere



- O-Bi
- O/LM
- Hippocamboseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

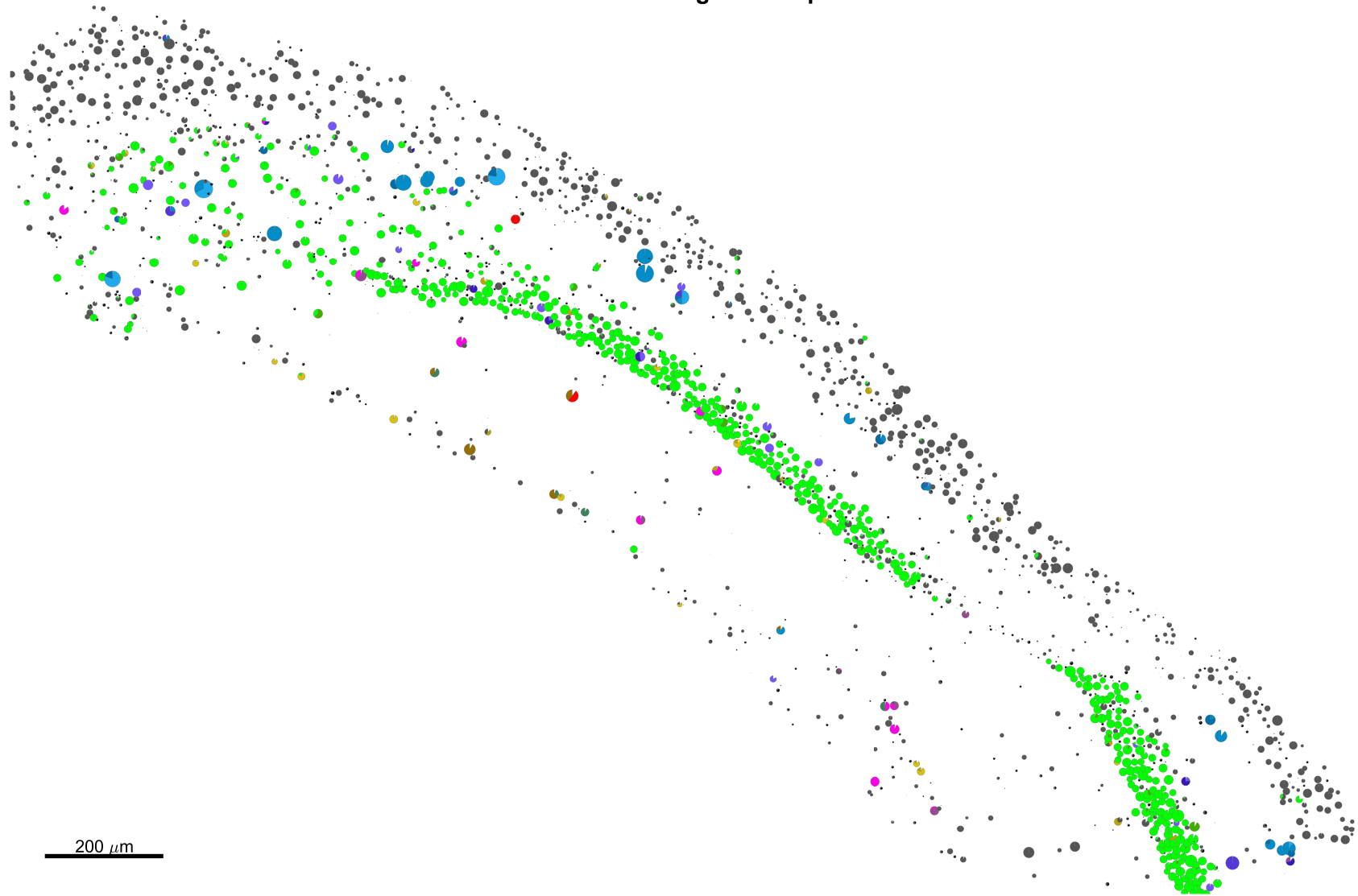
section 3-3 right hemisphere



- O-Bi
- O/LM
- Hippamboseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

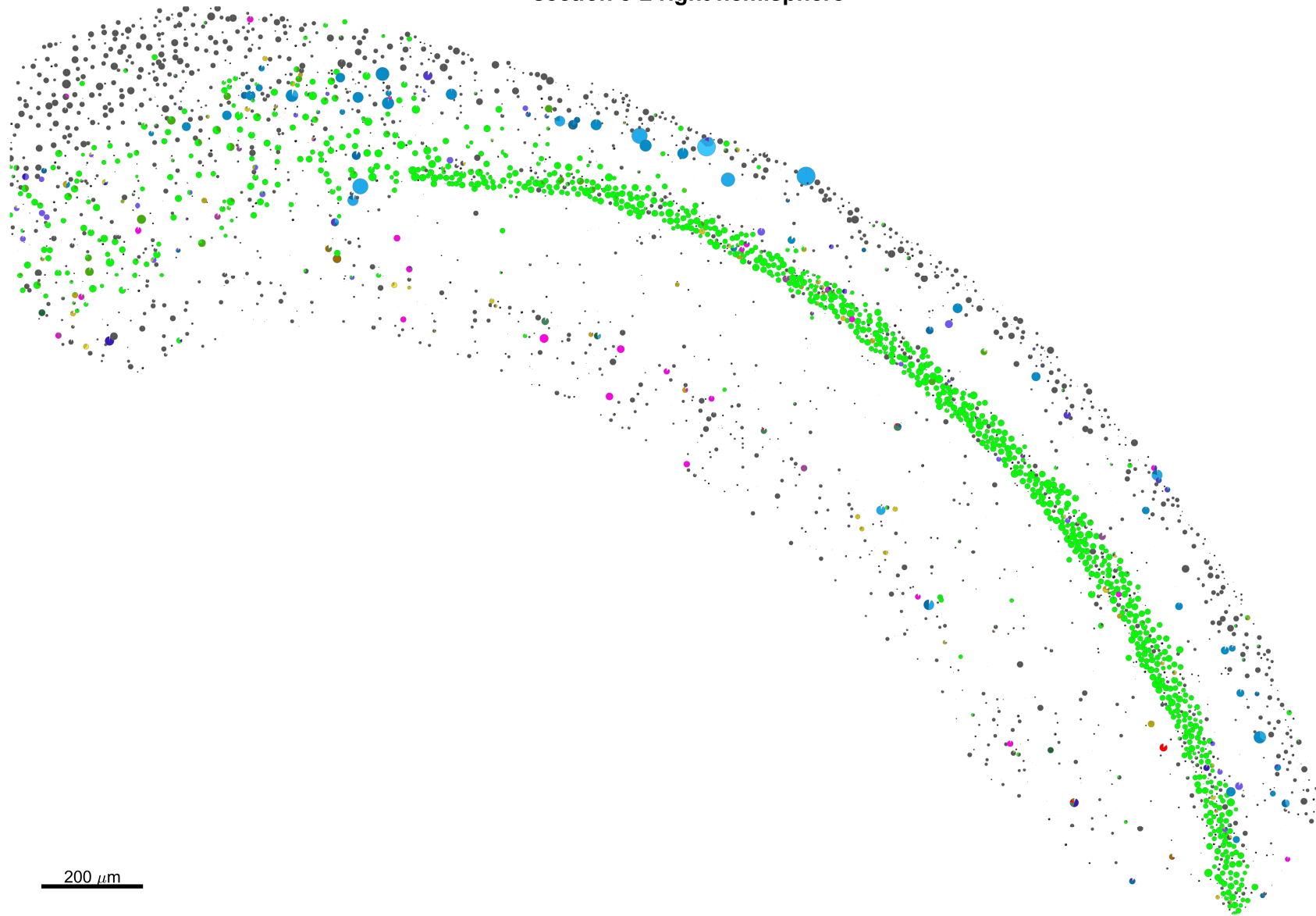
200  $\mu$ m

section 7-2 right hemisphere



- O-Bi
- O/LM
- Hippocamposeptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

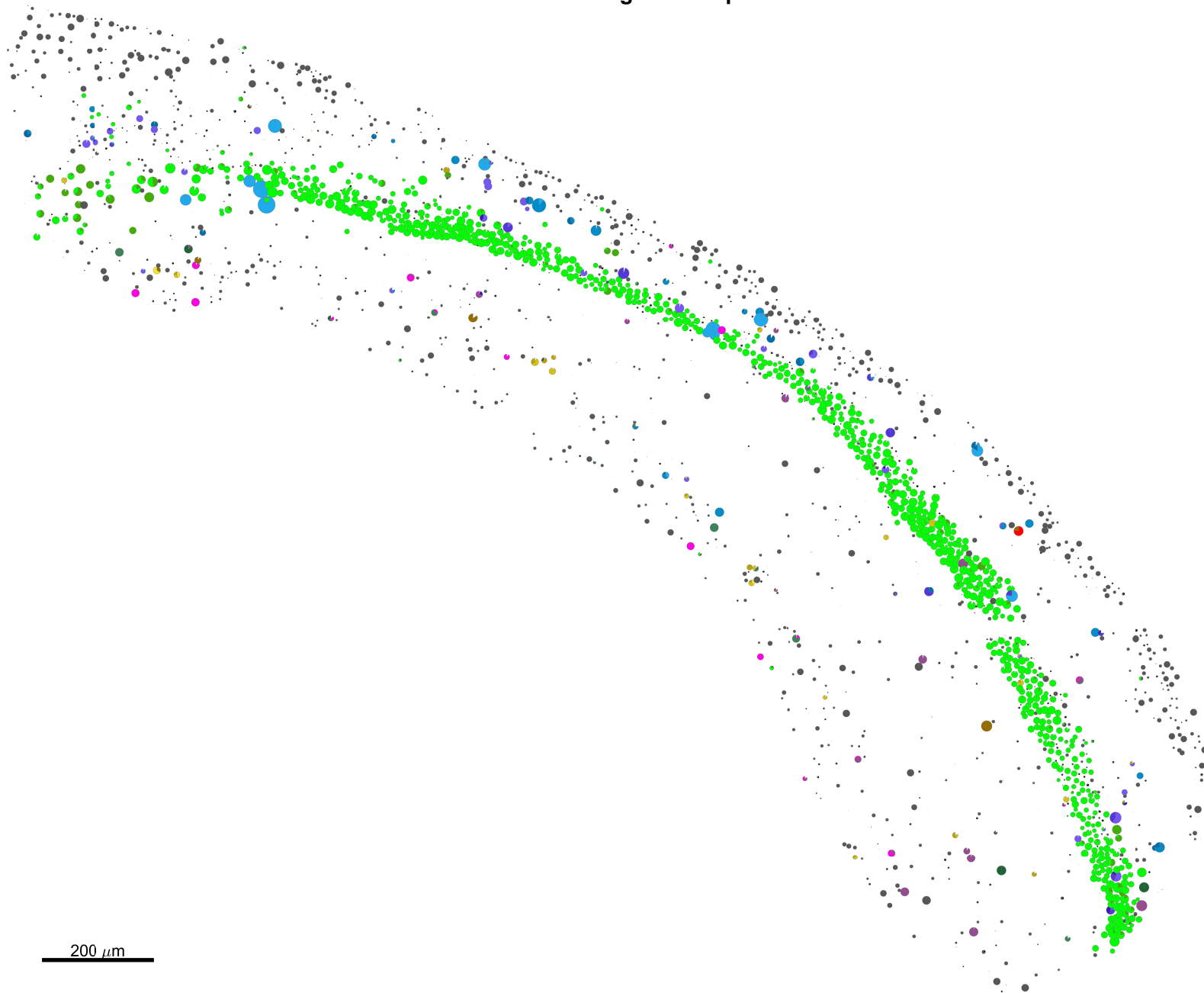
section 5-2 right hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

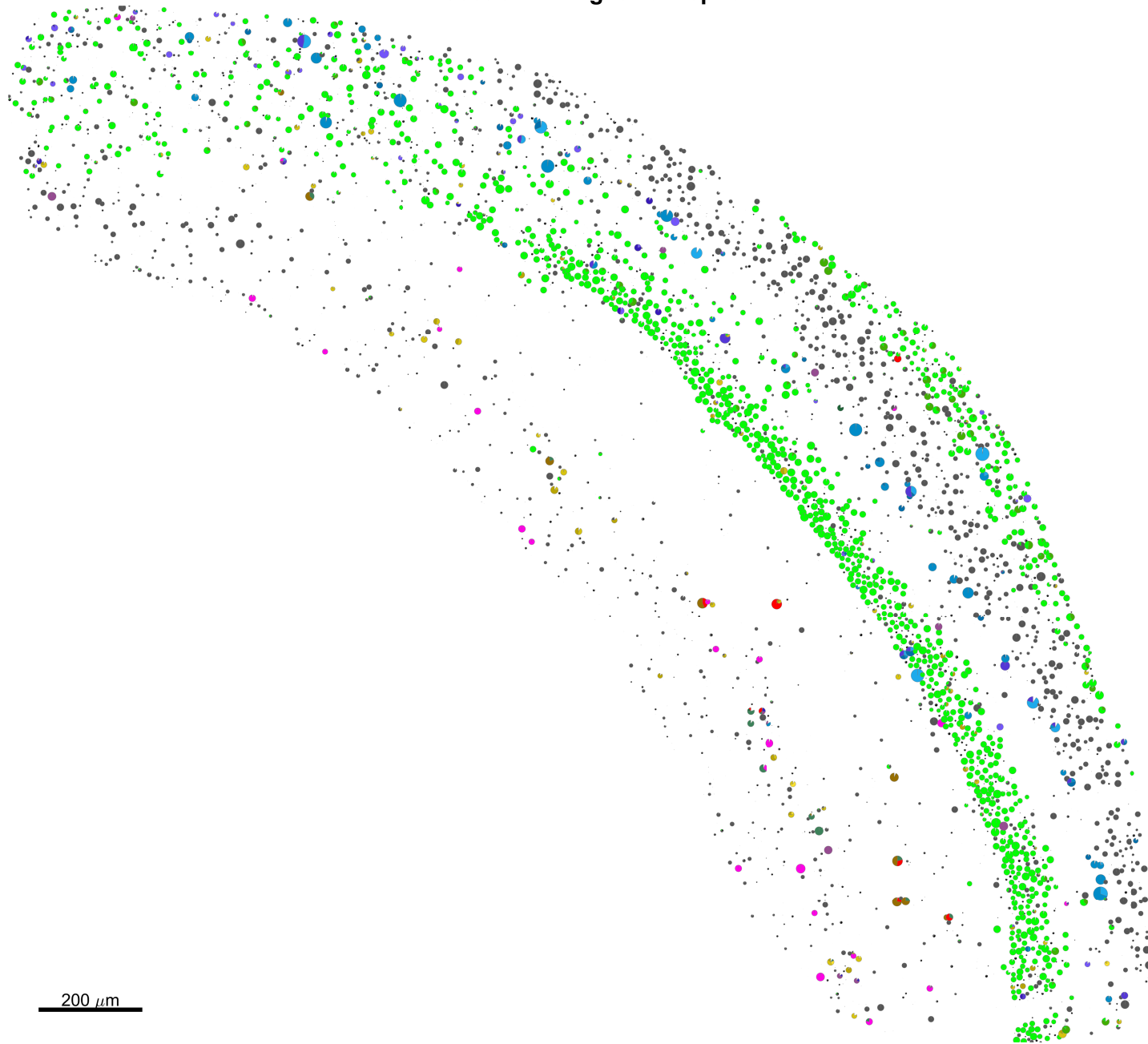
section 3-2 right hemisphere



- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

section 7-1 right hemisphere

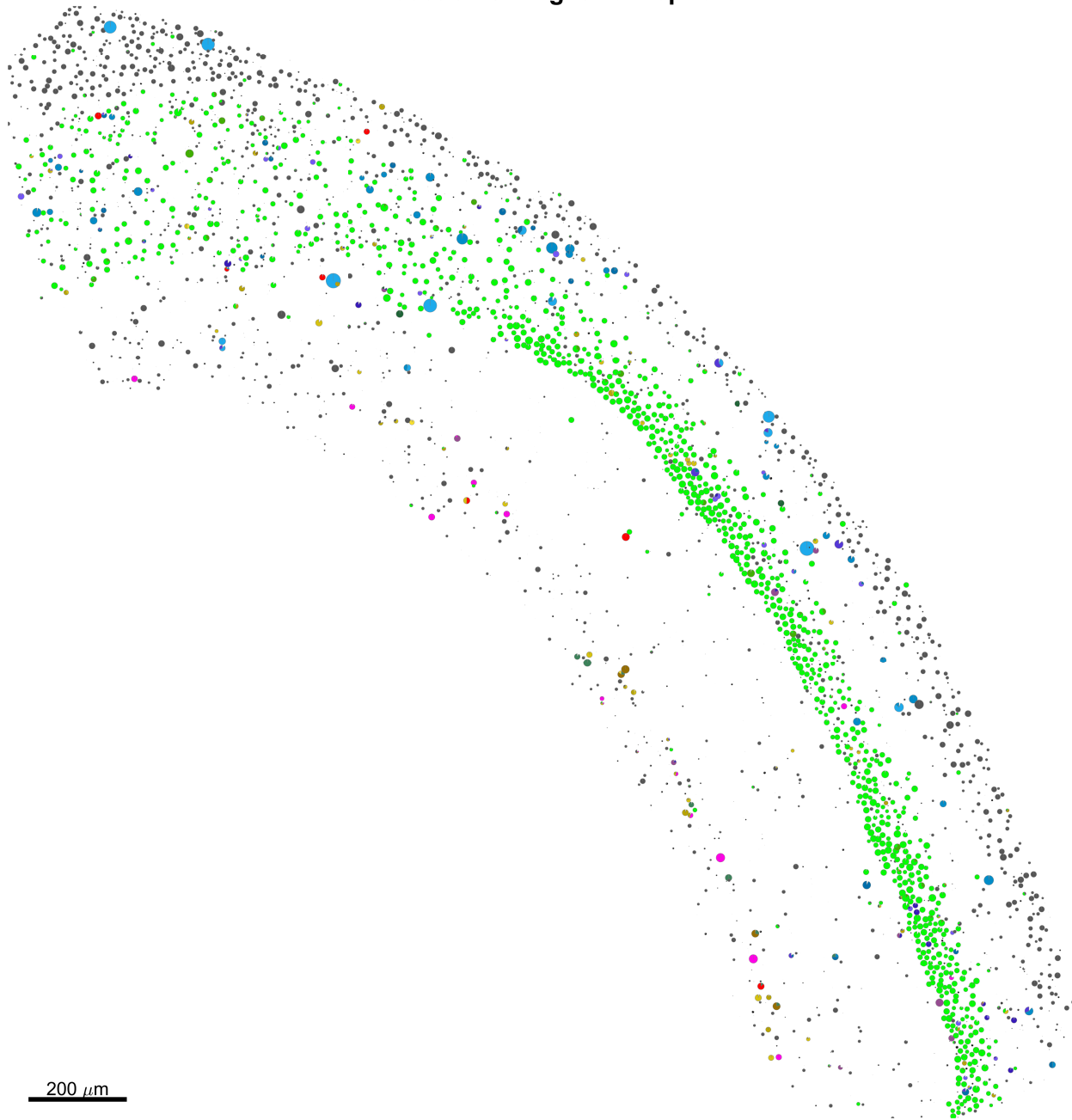


- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m



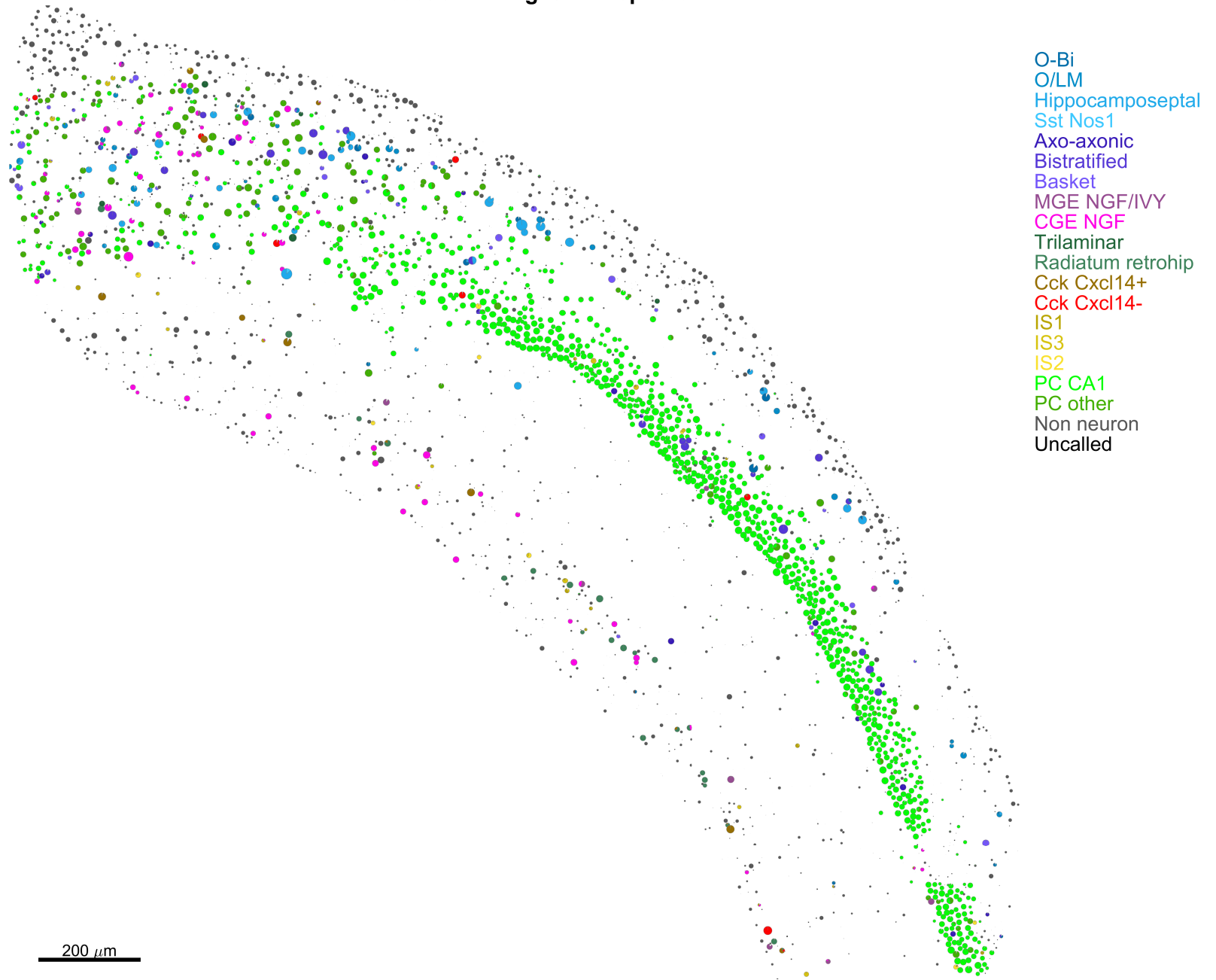
section 5-1 right hemisphere



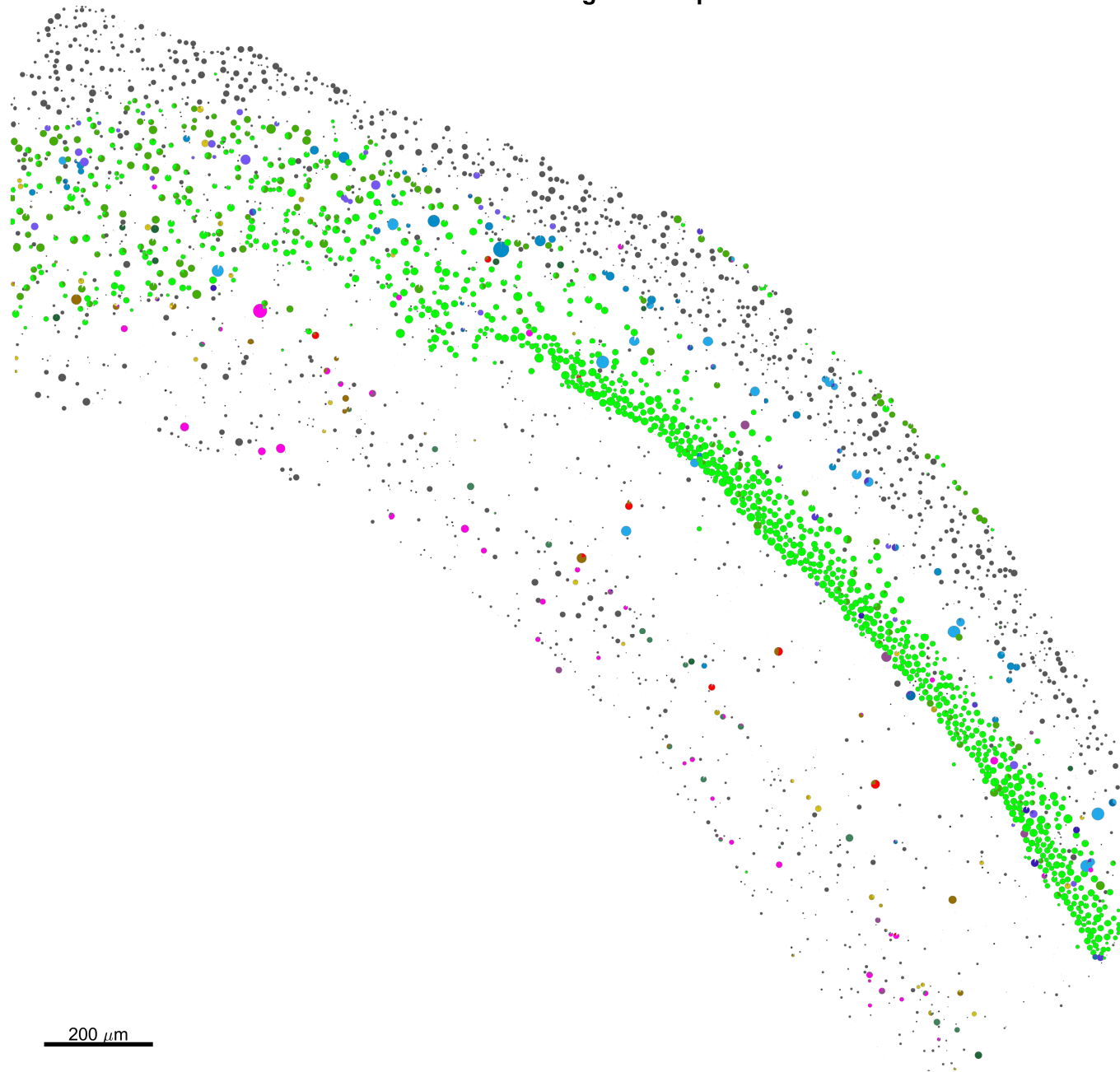
- O-Bi
- O/LM
- Hippocamptoseptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m

section 4-1 right hemisphere



section 3-1 right hemisphere



- O-Bi
- O/LM
- Hippocamposeptal
- Sst Nos1
- Axo-axonic
- Bistratified
- Basket
- MGE NGF/IVY
- CGE NGF
- Trilaminar
- Radiatum retrohip
- Cck Cxcl14+
- Cck Cxcl14-
- IS1
- IS3
- IS2
- PC CA1
- PC other
- Non neuron
- Uncalled

200  $\mu$ m