Massive Multiplexing of Spatially Resolved Single Neuron Projections with Axonal BARseq

Li Yuan[1], Xiaoyin Chen[2], Huiqing Zhan[1], Henry L. Gilbert[1] and Anthony M. Zador[1, †]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

[2]Allen Institute for Brain Science, Seattle, WA, USA

[†]Correspondence: zador@cshl.edu

ABSTRACT: Neurons in the cortex are heterogenous, sending diverse axonal projections to multiple brain regions. Unraveling the logic of these projections requires single-neuron resolution. Although a growing number of techniques have enabled high-throughput reconstruction, these techniques are typically limited to dozens or at most hundreds of neurons per brain, requiring that statistical analyses combine data from different specimens. Here we present axonal BARseq, a high-throughput approach based on reading out nucleic acid barcodes using *in situ* RNA sequencing, which enables analysis of even densely labeled neurons. As a proof of principle, we have mapped the long-range projections of >8000 mouse primary auditory cortex neurons from a single brain. We identified major cell types based on projection targets and axonal trajectory. The large sample size enabled us to systematically quantify the projections of intratelencephalic (IT) neurons, and revealed that individual IT neurons project to different layers in an area-dependent fashion. Axonal BARseq is a powerful technique for studying the heterogeneity of single neuronal projections at high throughput within individual brains.

Introduction

The mouse brain contains over 70 million neurons (Herculano-Houzel, Mota, and Lent 2006), and the combined length of their axonal trees stretches thousands of kilometers; in the human brain, there are orders of magnitude more. These axons form the scaffolding for neural circuits and hence for computation. Tracing these projections represents a formidable challenge. Traditionally, there are two main approaches. At one extreme, the projections of single neurons can be reconstructed at high resolution by labeling neurons one at a time, using e.g. the Golgi method or more modern sparse labeling based on viral delivery of fluorophores such as green fluorescent protein (GFP). Such single-neuron methods have undergone impressive advances in recent years, but even today allow the multiplexing of at most dozens of neurons from a single brain region (Winnubst et al. 2019; Peng et al. 2021; Gao et al. 2022). Alternatively, the projections of major projection pathways can be assessed using bulk tracing methods. For example, a bolus of virus expressing a fluorophore can be injected into one brain area, enabling the major projections of neurons within that area to be visualized by microscopy. These techniques have been used to systematically map the mesoscopic projections of the mouse brain (Oh et al. 2014; Muñoz-Castañeda et al. 2021). Bulk methods reveal the projections of large populations of neurons, but at the cost of single-cell resolution. Thus, there is a tradeoff between throughput and single-cell resolution in traditional methods.

We have recently developed a novel suite of nucleic acid barcode-based tracing techniques, which provide a third alternative. The first-generation method for exploiting barcodes in the context of circuit mapping was Multiplexed Analysis of Projections by Sequencing (MAPseq) (Kebschull et al. 2016). MAPseq can reliably and simultaneously map the projections of hundreds of thousands of individual neurons in a single experiment. MAPseq uniquely labels individual neurons by introducing random RNA sequences ("barcodes") via infection with a barcoded viral library. These random barcodes fill the cells and are co-expressed with a protein that has been engineered to bind to the barcode and drag it to distant axonal terminals. The pool of unique barcode identifiers is effectively infinite; even a 30 nucleotide (nt)-sequence has a potential diversity of $4^{30} \approx 10^{18}$ barcodes, far surpassing the $\sim 10^{8}$ neurons in the mouse brain. This high diversity implies that most neurons are uniquely labeled. The barcode RNA can then be extracted from the axons in an area of interest to determine which neurons project there; the number of molecules with a specific barcode sequence collected from a region is used as a proxy for the strength of the projection (i.e., axonal volume) of that particular barcoded neuron, in much the same way that GFP intensity is used as a proxy for projection strength in conventional bulk injections (Kebschull et al. 2016). Because high-throughput sequencing can quickly and inexpensively distinguish these barcodes, MAPseq can uncover the projections of hundreds of thousands of individual neurons in parallel within a single brain (Sun et al. 2021; Huang et al. 2020). The throughput of MAPseq for assessing single neuron projection patterns in a single brain is currently unmatched by any other approach.

MAPseq was the first approach to exploit barcoding for neuronal mapping. However, because it relies on bulk sequencing of homogenized tissue, its spatial resolution is determined by the precision of dissection. To achieve higher resolution, we developed BARseq (Barcoded Anatomy Resolved by Sequencing), the next generation of sequencing-based tracing (X. Chen et al. 2019; Sun et al. 2021). BARseq relies on *in situ* sequencing. Unlike conventional *in situ* hybridization, which uses a complementary probe to detect a specific RNA molecule in the cell, *in situ* sequencing obtains the exact sequence of each RNA target. This is a key difference, as the RNA barcode in any given cell is unique, unknown and highly diverse, making it very challenging to design probes in sufficient numbers for the desired targets. In contrast, *in situ* sequencing makes it straightforward to discriminate an almost infinite number of sequences. Combining BARseq-based sequencing of somatic barcodes and endogenous gene expression with MAPseq-based dissection and sequencing of barcodes in the axons allows us to associate the projection patterns of individual neurons with soma locations in a highly multiplexed manner (X. Chen et al. 2019; Sun et al. 2021). However, because spatial resolution in MAPseq is limited by the dissection of brain areas prior to

bulk sequencing, axonal projection patterns in this MAPseq/BARseq combined approach can only be crudely resolved. MAPseq and BARseq have been repeatedly validated using multiple methods in a wide range of brain areas (Zeisler et al. 2023; Y. Chen et al. 2022; Hausmann et al. 2022; Tsoi et al. 2022; Webb et al. 2022; Klingler et al. 2021; Mathis et al. 2021; Muñoz-Castañeda et al. 2021; Sun et al. 2021; Gergues et al. 2020; Huang et al. 2020; X. Chen et al. 2019; Han et al. 2018; Kebschull et al. 2016).

We therefore set out to increase the spatial resolution with which highly multiplexed axonal trajectories can be resolved using *in situ* sequencing. To achieve this, we developed a method, "axonal BARseq," for sequencing individual axonal "rolling circle colonies", or "rolonies", *in situ*. Axonal BARseq allows much finer resolution of the spatial organization of axonal projections than can be achieved with MAPseq. Using this approach, we identify the projections of thousands of individual axons projecting from a single localized injection in a single mouse, increasing throughput beyond current methods and eliminating the need to register injections across samples. Axonal BARseq has the potential to scale up to multiple injection sites and reveal projections from multiple sites, raising the possibility of sampling brain-wide projections from multiple neuronal populations at single cell resolution.

Results

Here we describe axonal BARseq, a highly multiplexed method for reconstructing axonal trajectories. We first describe the optimizations necessary to achieve single molecule sequencing of barcodes in axons. Next, we demonstrate its application to determine axonal projections from auditory cortex. We confirm that the single-neuron projection patterns obtained by this method are consistent with previous single-neuron approaches. We then show that the high resolution and multiplexing of axonal BARseq reveals the statistical structure of single neuron projections to different laminae in different areas.

Optimizing BARseq to achieve axonal resolution

We have previously demonstrated *in situ* read-out of barcodes expressed in somata (X. Chen et al. 2019; Sun et al. 2021). This is a much easier problem than the present challenge of reading out single axonal barcodes, because there are several orders of magnitude more barcodes in somata (10^3-10^4; (Kebschull et al. 2016)). We therefore sought to maximize the sensitivity of *in situ* read-out of barcodes to achieve high-efficiency single-barcode readout of barcodes transported millimeters or centimeters from their soma of origin.

To increase the sensitivity of *in situ* sequencing of axonal barcodes, we modified the sequencing protocols originally developed for barcodes in somata (X. Chen et al. 2018, 2019). The basic *in situ* sequencing protocol consists of (1) injection with a Sindbis virus engineered to express a diverse barcode library; (2) tissue preparation 24-48 hours after infection; (3) preparation of rolonies (nanoballs of DNA generated by reverse transcription of the RNA barcode, followed by gap-filling padlock-extension, ligation, and rolling circle amplification) in thin brain slices; (4) *in situ* sequencing by synthesis using standard Illumina reagents; (5) sequential 4-color imaging of each base in the barcode of each rolony (see Methods; Fig. 1A, SupFig. 1A). We optimized the reverse transcriptase used and the gap-filling procedure (SupFig. 1B-D), which increased the sensitivity of barcode detection to an efficiency of 20.2% compared to RNA *in situ* hybridization (SupFig. 1E-F). In addition, we engineered a Sindbis virus using a second-generation carrier protein (VAMP2nl), which carried barcodes more efficiently than our previously described carrier protein (see Methods).

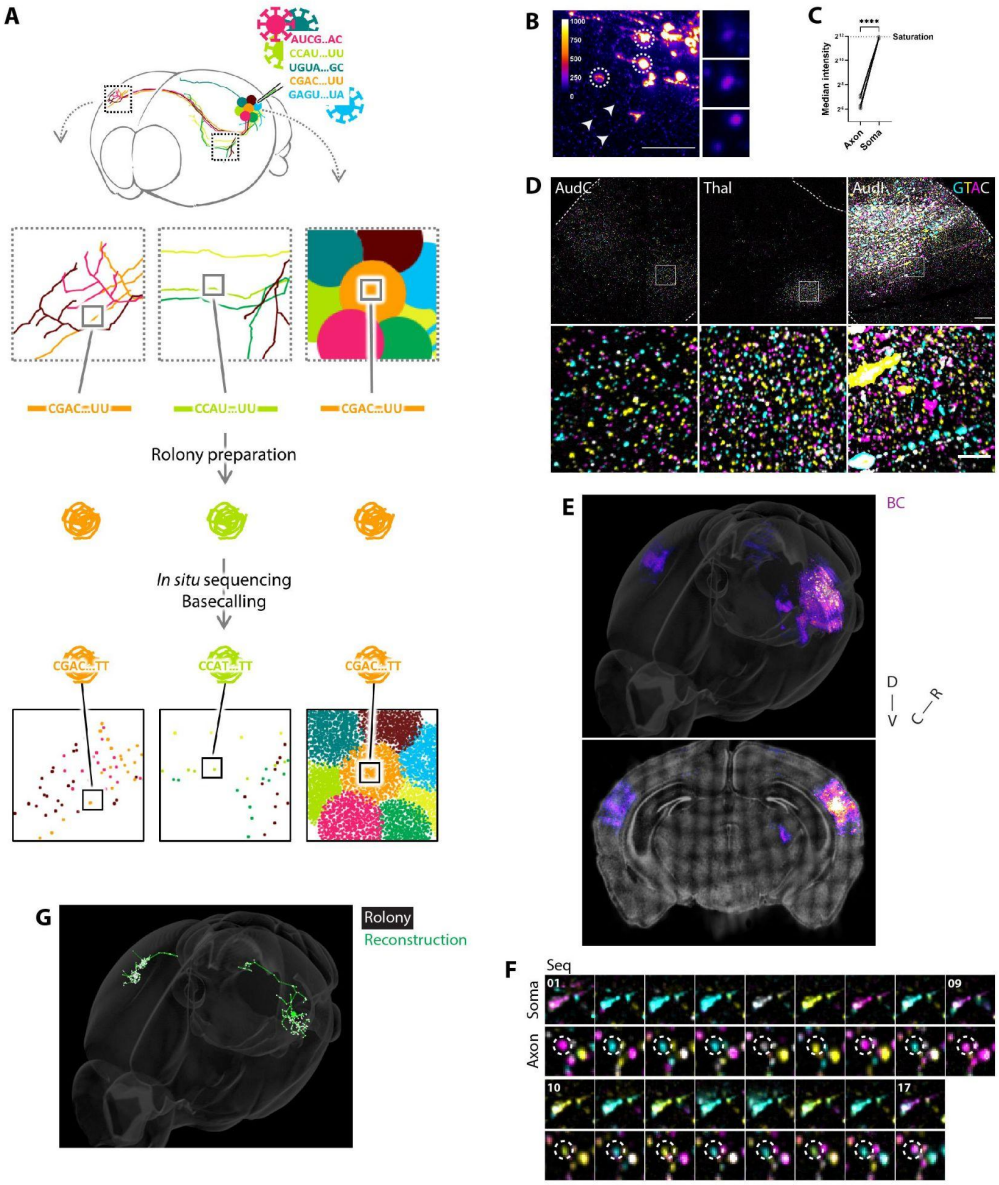


Fig. 1 Overview of axonal BARseq. (**A**) Workflow. Brain is injected with barcoded viral library. After 24-48 hrs of expression, during which RNA barcodes are transported to axon terminals, where they are amplified into rolonies and sequenced. (**B-C**) Single rolonies in axons have significantly weaker signals compared to somatic rolonies. (**B**) Representative image of somatic and axonal rolonies; dotted circle: somatic rolonies; arrow: axonal rolonies, with zoom-in views shown on the right; rolony intensity is color coded. Scale bar: 100 µm. (**C**) Quantification of intensity between axonal and somatic rolonies. Due to the large intensity difference between somatic and axonal rolonies, proper exposure for axonal rolonies often results in saturation of somatic rolonies. Paired *t*-test. (**D**) Representative images of axonal and somatic rolonies in AudC/I and ipsilateral thalamus. Images are from the first cycle of *in situ* sequencing. Dotted line, anatomical boundaries. Scale bar: top, 100 µm; bottom, 25 µm. (**E**) Registered barcode signals in CCFv3. Top, data in 3D model. Gray, brain outline. Bottom, coronal view of 25 µm of the sample. Gray, DAPI. (**F**) Representative images of *in situ* sequencing soma and a single axonal rolony with the same barcode. Soma ROI, 30.25 µm × 30.25 µm from injection site; axonal rolony ROI, 14.85 µm × 14.85 µm from ipsilateral thalamus. In total, 17 sequencing cycles are shown. (**G**) An example of tracing tracks for a single barcoded neuron reconstructed by connecting rolonies. Rolony location is indicated in white; soma location is indicated as a large green dot in ipsilateral cortex.

A further challenge of single-barcode axonal sequencing is to achieve the requisite sensitivity and accuracy during successive rounds of *in situ* sequencing. The signal from a single axonal rolony is not as bright as that from larger somata because somata contain many copies of the same barcode.(Fig. 1B-C). In addition, alignment of single rolonies across successive rounds of imaging poses additional challenges compared with alignment of somata. Overcoming these challenges required considerable modifications and optimization (see Methods, SupFig. 2-4).

Axonal BARseq of projections from auditory cortex

To assess the utility of these optimizations, we used axonal BARseq to reconstruct projections from mouse primary auditory cortex. Two days following unilateral viral injection, we performed 17 cycles of sequencing of coronal sections centered +/- 1 mm around the injection site (108 serial 20 µm sections). These sections contained many of the main projection targets of the auditory cortex, including most of the contra- and ipsilateral auditory cortices (AudC, AudI), contra and ipsilateral visual cortex (VisC, VisI), ipsilateral thalamus (Thal), part of the ipsilateral striatum (Str) and part of the ipsilateral superior colliculus (SupCol) (Fig. 1D-E, SupFig. 5A, SupTable 1). A total of 8620 unique axonal barcodes (obtained from 492950 sequenced rolonies) were used for analysis, with a median of 40 rolonies per barcode. About half (3698/8620) of the reliably detected axonal barcodes could be associated with somata whose position could be confidently determined (Fig. 1F, SupFig. 5B); the remaining barcodes could not be precisely localized due to various experimental and analytical challenges (see Methods). However, for most of the subsequent analyses (with the exception of Fig. 5 and SupFig. 9A-H), we used the entire set of axonal barcodes as the analyses do not require information about soma depth. Rolonies close to somata at the injection site (≤ 20 µm) were excluded (see *Methods*). Barcode statistics are summarized in SupFig. 5C-G; for details of manual validation of sensitivity and accuracy see Methods.

In conventional neuroanatomical single-neuron reconstructions, tracing requires that neuronal processes be filled with markers such as GFP or dyes, enabling visualization of axons as continuous structures. Disruption of this continuity due to errors in sample preparation or imaging can disrupt tracing and lead to catastrophic errors in reconstruction, potentially causing misattribution of an axon to the incorrect soma of origin. By contrast, because BARseq assigns axons to their soma of origin on the basis of their barcode sequences, assignments can be accurate even when barcodes are sparse. Errors in appropriate attribution of a barcode (e.g. due to sequencing errors) are rare and, importantly, are not catastrophic because they are independent, i.e. a given error affects only a single rolony.

We identified 8620 barcoded neurons with axonal projections outside the injection site, including ipsi- and contralateral cortex (CtxI and CtxC), thalamus (Thal), caudal striatum (Str), and superior colliculus (SupCol) (SupFig. 6). For visualization purposes it can be convenient to connect barcode rolonies to generate images that are similar to conventional neuroanatomical reconstructions. An example of such a connect-the-dots visualization, with straight lines linking nearby barcodes with the same sequence (see Methods), is shown in (Fig. 1G). In cases where the inter-barcode distance is large, this reconstruction is only an approximation, since axons can sometimes take tortuous paths, the details of which may not be captured by this approximation. However, these reconstructions are used for display purposes only; all quantifications rely directly on the rolonies themselves rather than the reconstructions. Fig. 2A-B shows the trajectories of 100 neurons (out of 8620), color-coded for display purposes (see 3D rotation animation, SupVideo).

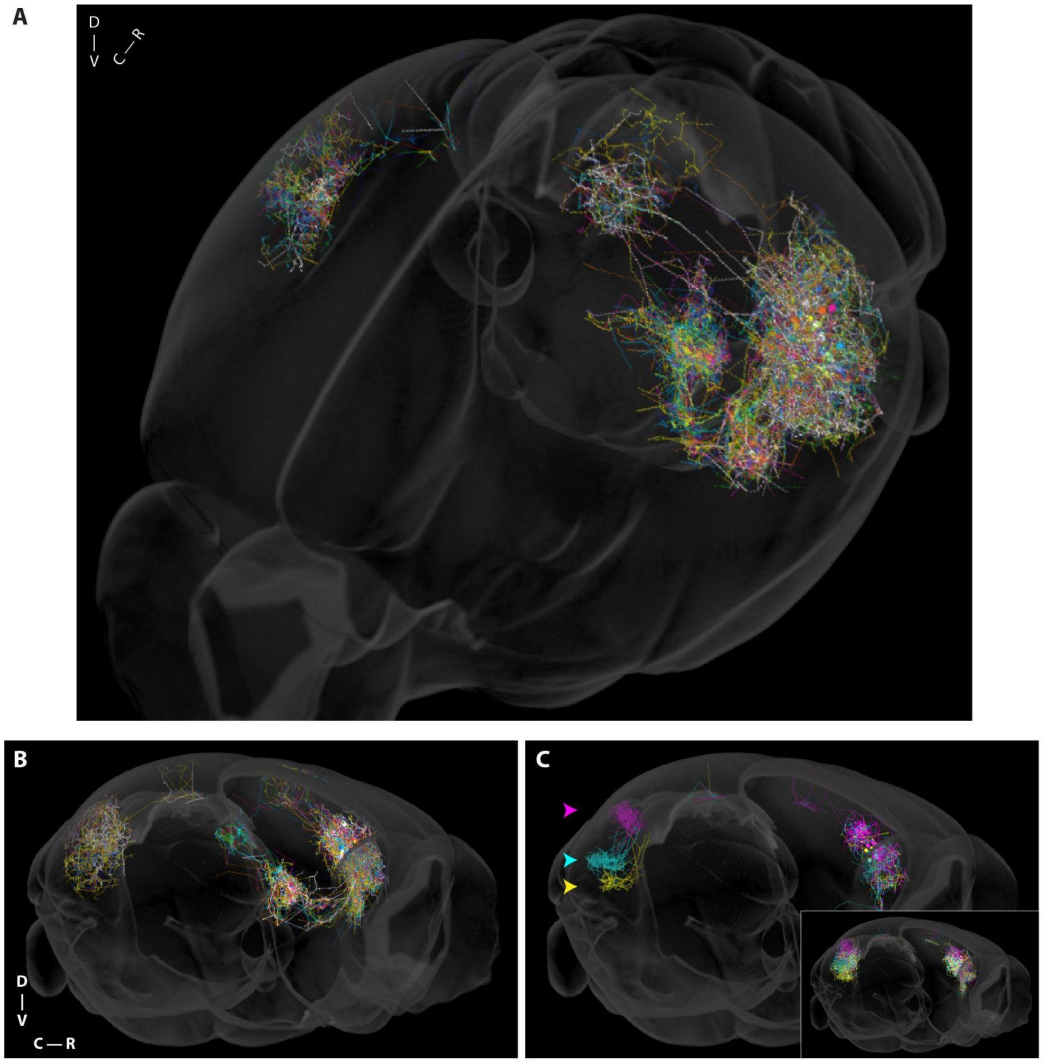


Fig. 2 Axonal barcodes can be used to reconstruct axonal projection in anterograde and simulated retrograde tracing. (**A-B**) Single-cell reconstruction from 100 barcoded neurons, 25 from each major cell type as described below. Somas (colors randomly assigned) are indicated by large dots in the left hemisphere. Not all tracts were imaged. (**C**) Identifying cells with focal projections to the same locations by simulated retrograde tracing. Cells from three simulated retrograde injections are indicated by arrows in right hemisphere. (*inset)* For each simulated injection, 25 neurons were randomly selected and plotted, regardless of whether their projections were focal or broad.

The large number of barcoded single neurons allowed us to identify subpopulations of neurons with distinct projection patterns. Fig. 2C (*inset*) shows a simulated contralateral retrograde injection of three colors. Among these neurons, subsets could be identified that projected very narrowly to specific patches (Fig. 2C). The identification of such subpopulations is facilitated by the high density of labeling obtained with axonal BARseq and would have been difficult to identify using conventional anterograde or retrograde methods. Analyses such as these highlight how the high degree of multiplexing (within a single sample) inherent in axonal BARseq enables identification of potentially interesting subpopulations.

Axonal BARseq can identify cell types based on projection trajectory

Following previous analyses of auditory cortex (X. Chen et al. 2019) and other cortical structures (K. D. Harris and Shepherd 2015), we manually clustered barcoded neurons into major cell types (Fig. 3A, SupFig. 6A-B). The top-level partition, between corticofugal (CF) and intratelencephalic (IT) classes, was based on the presence of subcortical projections descending below striatum, including the ipsilateral thalamus and the superior colliculus. IT cells were further divided into ITi and ITc, based on whether they had projections to the contralateral cortex. Barcoded somata were distributed across laminae and particularly enriched in mid layers (SupFig. 6C). Consistent with previous studies, CF somata were found predominantly in layer 5 (L5) and layer 6 (L6), whereas ITi and ITc somata are distributed across layers consistent (Fig. 3B-C, (X. Chen et al. 2019; J. A. Harris et al. 2019; Winnubst et al. 2019; Muñoz-Castañeda et al. 2021)). Thus, the projection patterns observed with axonal BARseq recapitulate those observed with conventional methods and with previous studies using BARseq.

CF neurons are divided into two major types (K. D. Harris and Shepherd 2015): extratelencephalic (ET, also known as pyramidal tract/PT neurons) and corticothalamic (CT). ET and CT neurons are distinct in the laminar positions of their somata (K. D. Harris and Shepherd 2015), axonal trajectory (Oh et al. 2014), gene expression (Tasic et al. 2016, 2018), and projection targets. ET neurons from the auditory cortex projects to both the tectum (X. Chen et al. 2019) and higher order thalamic nuclei, including lateral posterior nucleus (LP) and posterior limiting nucleus (POL) (J. A. Harris et al. 2019; Llano and Sherman 2008). In contrast, CT neurons do not project to the tectum and mainly project to medial geniculate body (MGB) in the thalamus (J. A. Harris et al. 2019; Llano and Sherman 2008). In previous work (X. Chen et al. 2019) we distinguished ET neurons from CT neurons by the fact that only ET neurons project to the tectum. In the current experiment, however, we did not sample the entire tectum, and thus could not distinguish these two populations of neurons by the presence or absence of tectal projections. Instead, we exploited the high spatial resolution of axonal BARseq to partition neurons based on axonal trajectory.

Fig. 3D shows representative axonal trajectories to the thalamus of CF neurons. One group follows a dorsal route and travels through the reticular nucleus, whereas the second follows a ventral route (SupFig. 7A-C). These two routes are consistent with the two axonal trajectories of CT and ET neurons, respectively (Mitrofanis and Baker 1993; Oh et al. 2014). We further combined axonal BARseq with immunohistochemistry to distinguish projections to different thalamic nuclei (SupFig. 7D-E). Consistent with the hypothesis that these two trajectories distinguish CT and ET neurons, neurons taking the dorsal route are concentrated in L5 (peak around 60% depth; Fig. 3B-C) and project to LP and POL (Fig. 3D, SupFig. 7D-E). By contrast, neurons taking the ventral route are concentrated in L6 (peak around 90% depth; Fig. 3B, C) and project to the MGB (Fig. 3D, SupFig. 7D-E). These results indicate that axonal BARseq can distinguish populations of projection neurons based on axonal trajectory.

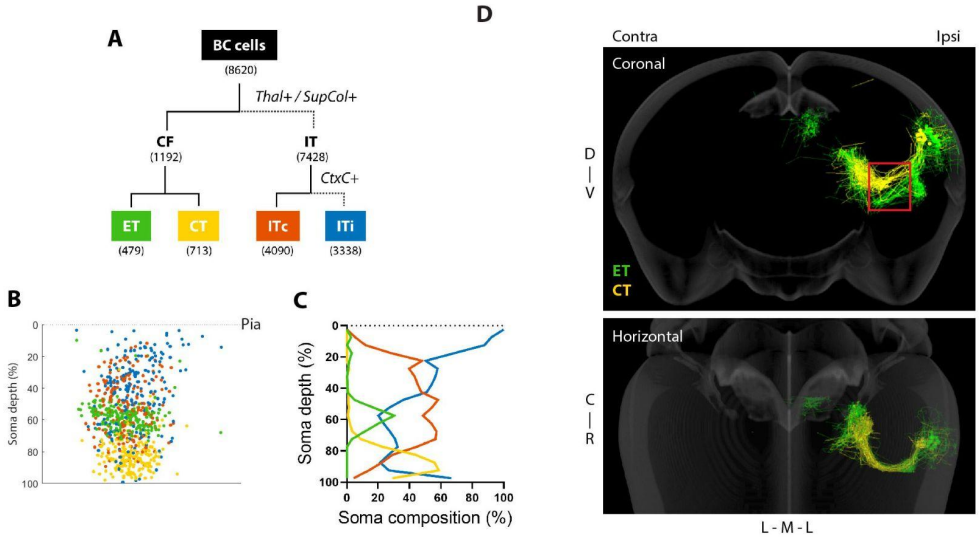


Fig. 3 Cells can be divided into major cell types using projection targets and axonal trajectory. (**A**). Barcoded cells were divided into four cell types: ET, CT, ITc and ITi. Cells were divided into CF/IT using Thal/SupCol projection; IT cells were divided into ITi/ITc using contralateral cortical projection. SupCol was identified using midbrain area in CCFv3. Solid line, with the specific projection; dotted line, without the projection. Numbers next to the group names indicate cell counts of the groups. (**B-C**) Soma depth of four cell types in injection site. (**B**) Scatter plot of 200 randomly selected cells per group. (**C**) The distribution of the four cell types along depth. Cell types were color coded as in Fig. 3A. (**D**) Projection tracts of CT and ET neurons consistent with spatial features of both cell types, 60 randomly selected neurons per type. Red box, region of interest for CT/ET grouping. Coronal view, top; horizontal view, bottom. CT/ET Rolonies with region boundaries were shown in SupFig. 7D-E.

Diversity of laminar terminations across areas

Neurons within a major cell type are often heterogeneous. For example, subtypes of IT cells differ according to dendritic morphology, projection patterns and gene expression (K. D. Harris and Shepherd 2015). One approach to studying such heterogenous characteristics is to define subpopulations by exploiting differential gene expression using selective expression of Cre-recombinase (J. A. Harris et al. 2019; Taniguchi et al. 2011). Although in some cases neuronal populations marked by Cre lines show remarkable precision in their projections, with e.g. terminations restricted to just a narrow band in a single lamina, in other cases neurons within a single population identified by Cre recombinase can project to different lamina in different areas (J. A. Harris et al. 2019). However, because most previous studies relied on mapping bulk (non-single cell) projections, it is not clear whether a given neuron withinsupthe population projects to multiple laminae, or if instead the population is itself heterogenous, with some neurons projecting to one lamina and others to another. We therefore studied the characteristics of laminar projections across a large population of single neurons.

The projections of single IT neurons can terminate in different layers in different cortical targets (Winnubst et al. 2019; Peng et al. 2021; Gao et al. 2022). For example, the neuron depicted in Fig. 4A mainly had superficial projections in the ipsilateral medial cortical region, but most of its projections in the two lateral cortical regions terminated in deeper layers. To assess how well a neuron's projection pattern to one area could predict its projection pattern to another, we compared the laminar projections of those IT cells that projected to a minimum of two cortical targets. For the purposes of this analysis we divided the cortex into four targets: LatI, MedI, MedC and LatC (ipsilateral-lateral, ipsilateral-medial, contralateral-medial, contralateral-lateral, Fig. 4B). LatI/C mainly consists of auditory areas, whereas MedI/C mainly consists of higher-order visual areas. The bulk-level laminar termination patterns of corticocortical projections are consistent with those observed in the Allen Connectivity atlas (Oh et al. 2014): Projections to medial areas were largely localized to superficial layers, whereas those to lateral areas were distributed across layers (SupFig. 8E-G). At the single-neuron level, neurons tended to have similar laminar patterns in similar targets across hemispheres: 132 cells with both ipsi- and contralateral medial projections tended to terminate in superficial layers in both medial targets, whereas 3473 cells with both ipsi- and contralateral lateral projections tended to terminate in both upper and lower layers (SupFig. 8H). However, 709 cells that projected to both a medial and a contralateral-lateral area often terminated in different layers, with a tendency to project more superficially in medial areas (Fig. 4C-E, SupFig. 8J-L). Thus, these results indicate that the laminar termination patterns of the axons of IT neurons are largely symmetric across the two hemispheres, but distinct across medial and lateral areas.

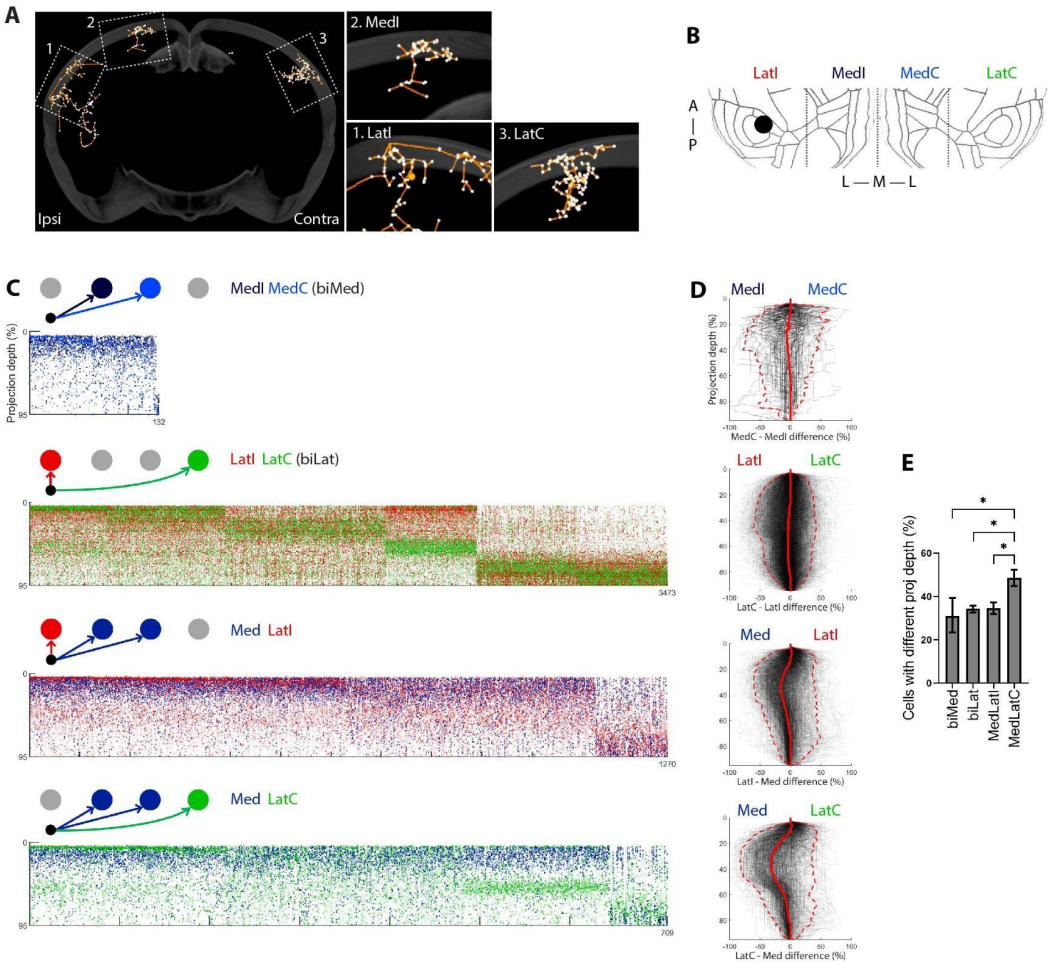


Fig. 4 Projections of single IT neurons to different areas often terminate in different laminae. (**A**) Coronal view of representative single-cell reconstruction (orange) and axonal rolonies (white) of an IT neuron with multiple cortical targets. Soma, big orange dot in AudI. Zoom-in view: 1, LatI; 2, MedI; 3, LatC. (**B**) Division of cortex into medial and lateral targets in flatmap. *Black disk*, injection site. (**C**) Rolony depths for pairs of cortical areas for each neuron. Each column represents rolonies from a single cell. Graphs were scaled on x-axis for display purposes. Bin size, 1% depth. (**D**) Pairwise comparison of cumulative distribution of rolony depths. *Black line*: individual barcoded cells; *solid red line*: median; dotted red line: 2.5 and 97.5 percentile (**E**) Cells with different laminar patterns were found in all groups with the highest rate in MedLatC. Bar graph, median with confidence intervals. *, significant difference with no overlap CI.

Diversity of laminar terminations within an area

We next examined how the laminar position of a neuron's soma was correlated with the structure of its axonal projections to different cortical targets. We divided Lat-projecting cells into three groups based on soma depth (Fig. 5A): upper (roughly layer 2-4; ≤ 35% depth, in red), middle (roughly layer 4 to upper layer 5; 35-60% depth, in orange) and deep (roughly lower layer 5 to layer 6; > 60% depth, in blue). The laminar positions of these three groups corresponded to annotations or internal markers in LatI (SupFig. 9A-B). At the bulk level, we found that deep (*blue*) somata mainly projected to deep layers, whereas the upper and middle somata preferentially projected to upper-middle layers of their targets (Fig. 5B, SupFig. 9C).

We then compared the organization of these projections to four targets at the single-cell level. Three example neurons are illustrated in Fig. 5C-D. In these examples, and across the population, projections from upper and middle layers to all target areas (LatI, LatC, MedI, and MedC) were largely restricted to upper-middle layers (Fig 5E). However, we observed a marked correlation between the depth of a deep layer soma and the proportion of deep projections (SupFig. 9D). Thus, the most superficial of the "deep" somata had more upper layer projections, and the proportion of upper layer projection gradually decreased with somatic depth. This result is consistent with previous observations that deep layer IT neurons project mostly to deep layers (X. Chen et al. 2019), but further reveals heterogeneity of termination patterns within a layer-defined population of neurons.

Our data also recapitulated known differences in the areas of projections across different IT neuron populations (Fig 5F-G, SupFig. 9E, I-K). Upper layer cells were less likely to project to ipsilateral striatum, and middle layer cells were more likely to project to Med compared to the other two groups. These results are consistent with previous observations that IT neurons in the superficial layer have more dedicated projections, whereas neurons in the deeper layers project to more target regions per neuron (J. A. Harris et al. 2019; Muñoz-Castañeda et al. 2021).

Because axonal BARseq has both single-cell resolution and high spatial resolution, we could further distinguish the fanout pattern of axonal projections within a target area. Upper neurons tended to focus their axonal terminations in a small area (Fig. 5C-D, *red neuron*), whereas deep neurons tended to project more diffusely (Fig. 5C-D, *blue neuron*), with middle neurons somewhere intermediate (eg two foci; see Fig. 5C-D, *orange neuron*). These observations are summarized across the population in Fig. 5H (*see also* SupFig. 9G-H, L-M). Taken together, axonal BARseq resolved systematic differences in the projections of layer-defined subpopulations of IT neurons in the laminar patterns of axon termination, their projection targets, and the projection patterns within a target.

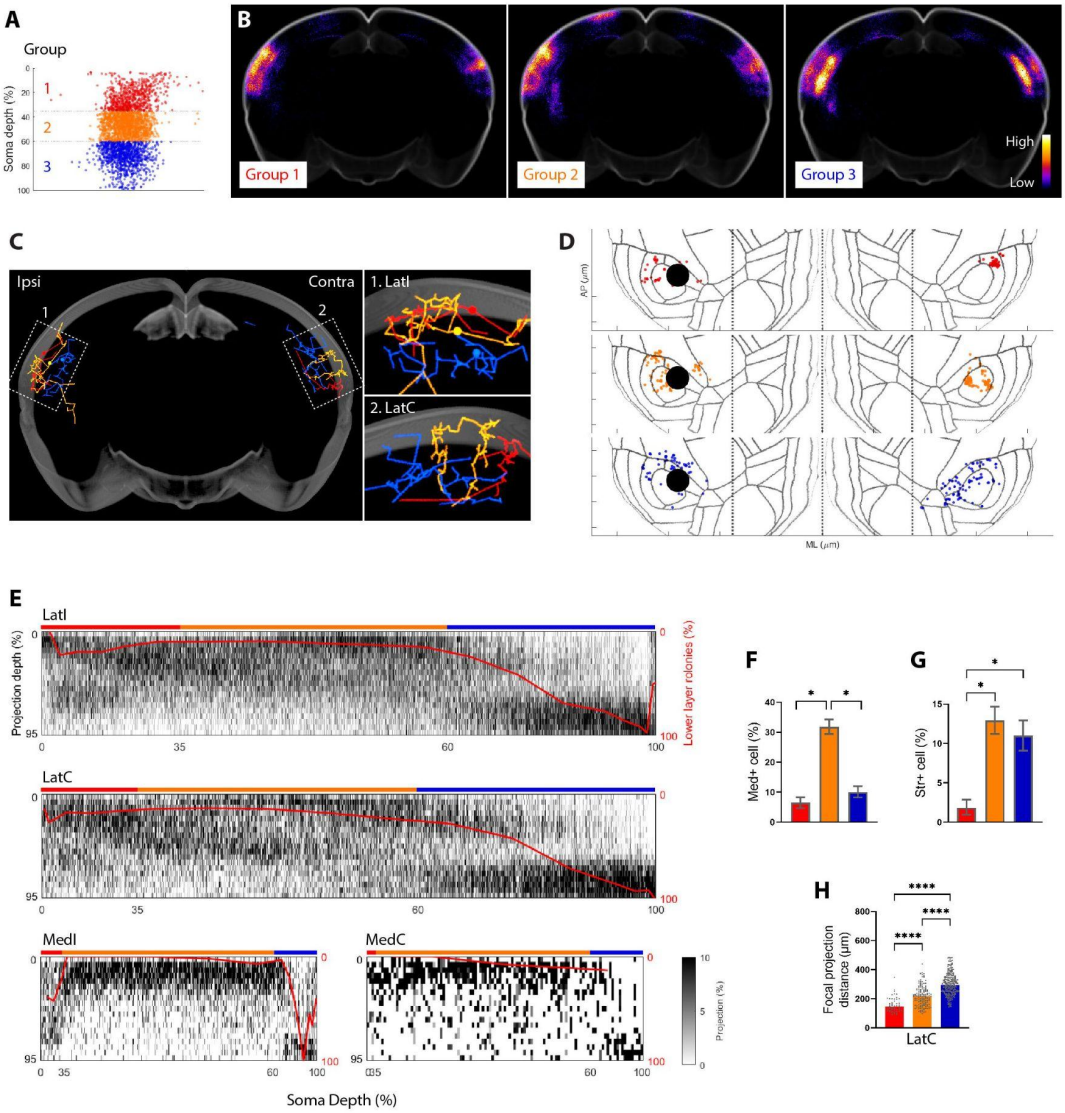


Fig. 5 Simple grouping of IT cells using soma depth captures the cell type difference in laminar termination, projection targets and axonal morphology. (**A**) Lat-projecting IT cells were divided into three groups using soma depth: group1, 0-35%; group2, 35-60%; group3, > 60%. Dotted line: 35 and 60% depth. Cell counts: group1, 710; group2, 1355; group3, 1044. (**B**) Three groups have different rolony distribution patterns at bulk level. 710 barcodes per group were rancomly selected for display purposes. Images shown are sum projections of coronal views. (**C**) Representative single-cell reconstruction of a biLat cell. Zoom-in panel: 1, LatI; 2, LatC. Soma, big dot in AudI. (**D**) Rolony location of individual neurons in Fig. 5A on cortical flatmap. Dotted lines, Med-Lat boundaries. (**E**) At single-cell level, lower layer Lat-projecting cells mainly project to lower layers while upper layer cells preferentially project to upper layers. Heatmap, frequency distribution of rolony along depth; bin size, 5% depth; one cell per column. Cells were sorted by soma depth (x-axis). Red line: percentage of lower projection (> 60% depth). Cells were binned into 5% bin using soma depth, and median percentage of lower projection were calculated per bin. Cell counts: LatI, 3012; MedI, 525; MedC, 113; LatC, 1612. (**F-G**) Group 2 has higher percentage of cells projects to median cortical region (F) while group 1 has lower percentage of cells project to ipsilateral striatum (G). Bar graphs show medians and confidence intervals. *, significant difference with no overlap CI. (**H**). Upper-layer-projecting cells have more focal projection in LatC compared to lower-layer projecting cells. Bar graph, median; dots, individual cells. Cell counts, LatC group1 51, group2 134, group3 233. Kruskal-Wallis test, Dunn's test for multiple comparisons.

Discussion

We have described axonal BARseq, a highly multiplexed method for mapping neuronal projections with single-cell resolution. A key advantage of axonal BARseq over conventional optical methods is the large number of projections that can be mapped in a single brain. As a proof-of-principle, we used axonal BARseq to map the projections of more than 8000 neurons from primary auditory cortex of a single mouse. Axonal BARseq represents an advance in spatial resolution over first-generation BARseq, which relied on bulk sequencing to read out projection barcodes. We used this large data set to systematically quantify the heterogeneity of auditory cortical projections to multiple targets. Additionally, we showed that axonal BARseq can be combined with routine immunohistology (SupFig. 7E).

The central challenge in multiplexed mapping of axonal projections is that the axons are densely packed and tangled together. When the distance between two axons approaches or falls below the limit of optical microscopy, the fidelity with which they can be distinguished using classical methods decreases. The greater the number of labeled axons, the greater the probability that two axons will be indistinguishable, and thus the greater the probability of error. Tracing errors are catastrophic because the error implies that an axon will be misattributed to the incorrect soma of origin. These challenges are particularly acute when tracing axons over long distances, because axons often travel in bundles. Such considerations limit the number of labeled axons that can be optically reconstructed within a single specimen.

Axonal BARseq circumvents these challenges by eliminating the need to trace axons. Instead, barcodes provide a direct means of associating the axon with its parent soma. Errors in appropriate attribution of a barcode (e.g. due to sequencing errors) are rare and, importantly, are not catastrophic because they are limited to a single rolony. Moreover, projections to distant targets can be assessed even without the need to trace the entire axonal path from soma to target. This enables efficient mapping of projections to multiple target areas, even if the targets are widely separated in space.

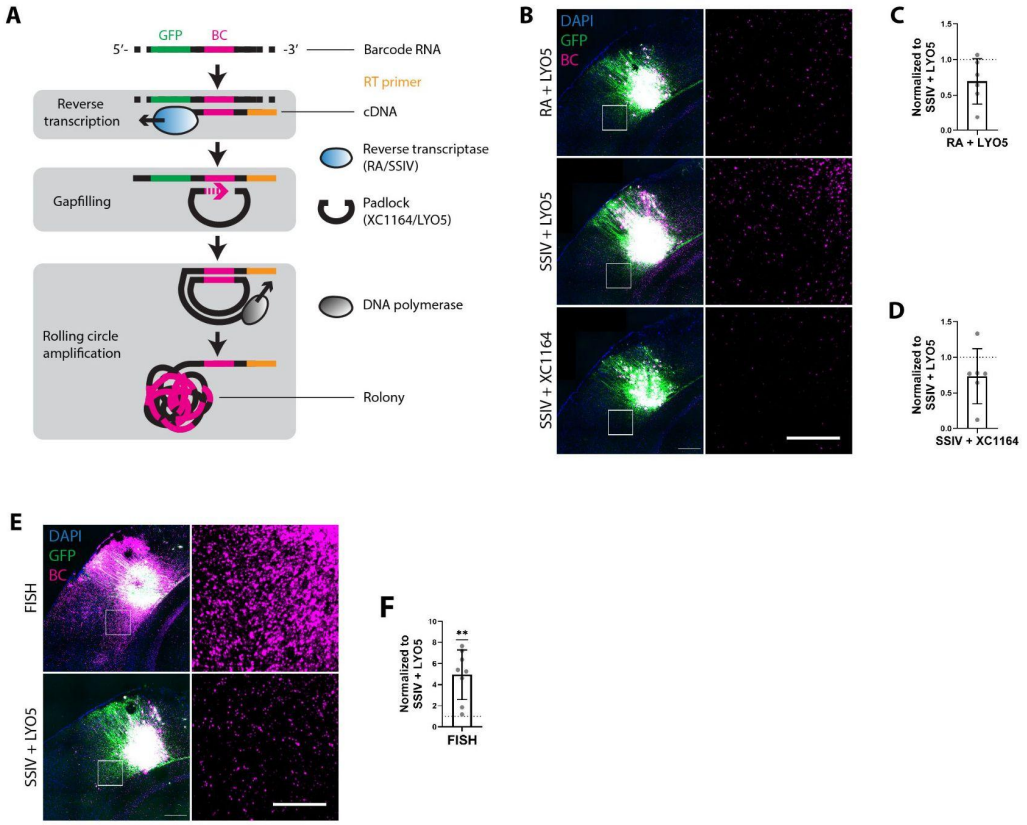
The high throughput of axonal BARseq is useful for three reasons. First, high througput allows for statistical analyses using large numbers of single neurons, which has the potential to reveal statistical structure that is not evident with smaller sample sizes. Second, the fact that the samples come from a single animal is useful when individual animals are rare or valuable, such as for nonhuman primates, non-canonical model systems, and transgenic animals. Finally, axonal BARseq allows for dense mapping of projections within a single brain, obviating the need to register all results to a single reference atlas. Avoiding registration eliminates the errors that arise from comparing across brains. Moreover, registration implicitly assumes that all brains are the same, whereas in some cases idiosyncratic differences between brains may be important. These advantages make axonal BARseq uniquely useful for certain applications, such as studying the relative topography of projections.

Here, we used axonal BARseq to simultaneously trace different cell types within a single wild-type animal. By tracing the subcortical projecting cell types, ET and CT, we directly observed the differences between them, including the laminar distributions of their somata and their projection patterns in the thalamus (Fig. 2B-C, SupFig. 7D-E). We identified hundreds of IT cells projecting to multiple cortical targets, and were able to quantify the extent to which single neurons projections to different brain areas terminated in different laminae (Fig. 4E). We also found that IT cells can have focal or sparse patterns in their contralateral projections, with sparse projections originating from lower-layer cells and targeting lower layers (Fig. 5H). Focal projections originated from upper-middle layer cells with a different laminar distribution. Our results demonstrate the effectiveness of axonal BARseq in recapitulating previously observed differences between cell types and making novel discoveries in heterogeneous cell populations.

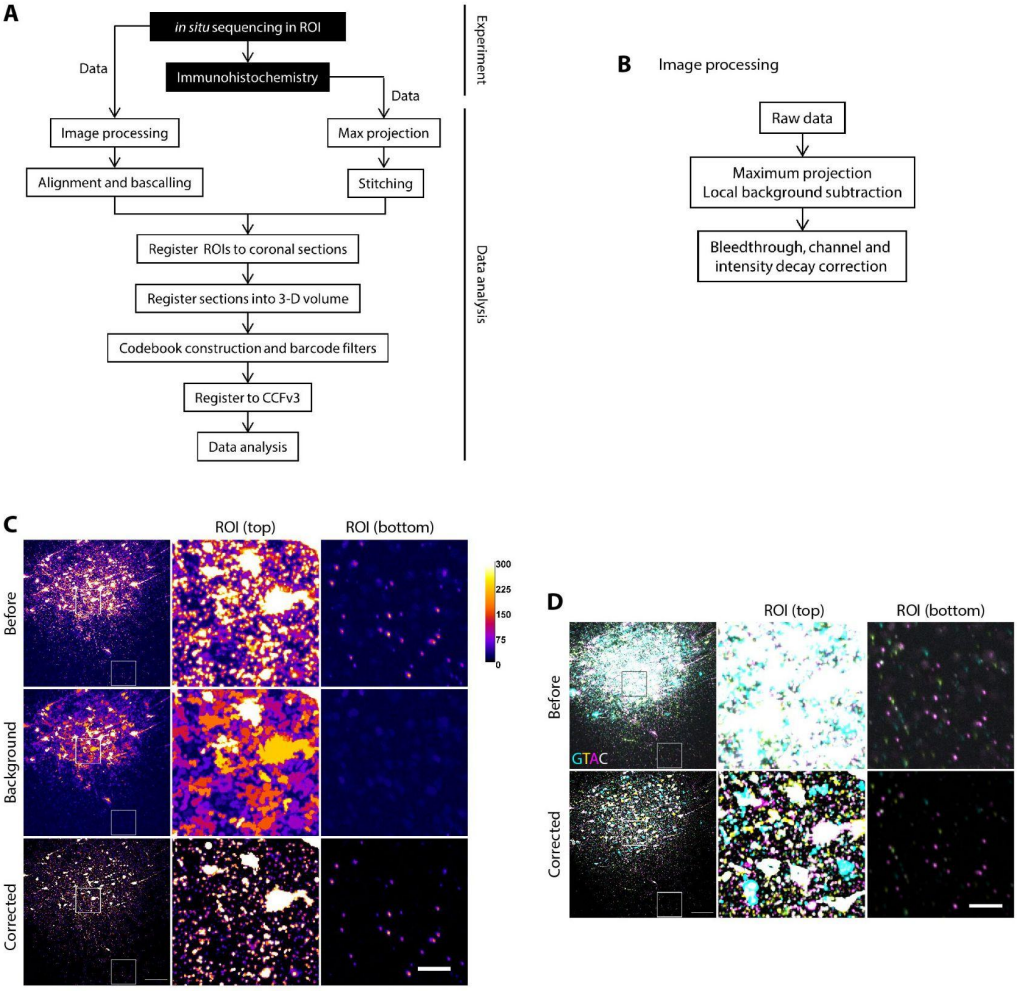
Limitations and future developments

Axonal BARseq has several limitations compared with other methods. First, like conventional GFP-based tracing approaches, axonal BARseq reveals only axonal projections but not synaptic connections. To achieve synaptic resolution requires electron microscopy or visualizing synaptic markers markers using super-resolution microscopy or expansion microscopy. Second, the reconstructed axons may not have high fidelity, as the axonal rolonies can be spaced as far as tens of microns apart. This means that branch points or even entire branches may be missed, which can affect the accuracy of the reconstructed neuronal projections. Although the neuroanatomical literature has traditionally placed a high premium on reconstructing the finest processes with high fidelity, for many applications the increased throughput—thousands of neurons per sample–may represent a reasonable tradeoff. For example, if the main interest is in the laminar distribution of axonal innervation (Fig. 4C and Fig. 5E), the fact that not all fine axonal processes are recovered may represent an acceptable compromise. It may also be possible to increase the density of axonal rolonies and thus the fidelity of reconstruction by improving the delivery of barcodes to axons (e.g. with a better carrier protein) and by refining the protocols for rolony recovery. Alternatively and additionally, it might be possible to combine axonal BARseq with either classic fluorophore-based tracing or brainbow (Livet et al. 2007). Finally, in the current work we did not attempt to resolve local axons near the injection site because of limitations of the current algorithms for automated basecalling of rolonies. However, newer algorithms may make it possible to resolve rolonies at high density.

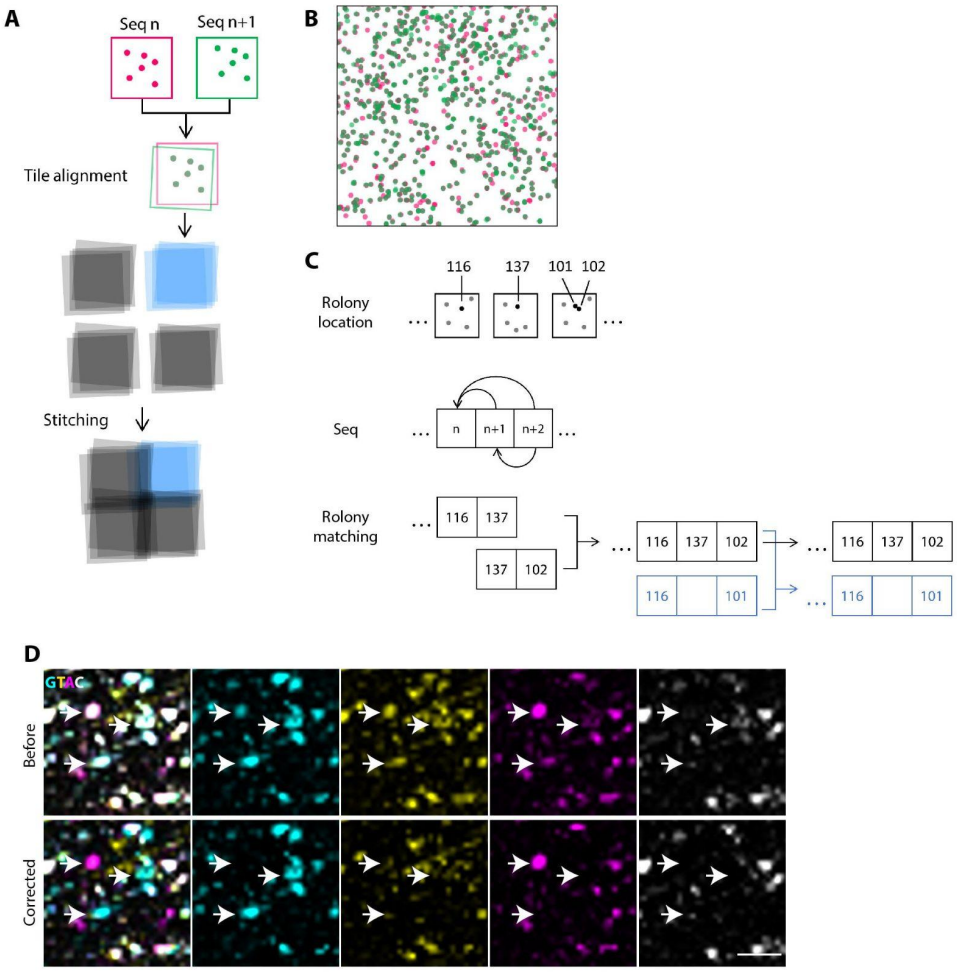
There are several potential avenues for improving upon the current axonal BARseq method. First, axonal BARseq could be combined with conventional GFP-based tracing techniques. By combining the higher resolution of conventional single neuron tracing–the ability to resolve even the finest axonal branches–with the higher throughput of axonal BARseq. Second, axonal BARseq can be combined with the expression of endogenous genes, which would enable us to correlate projection patterns with transcriptomically defined cell types, allowing a better understanding of the differences in projection patterns both among and within cell types. Finally, we expect that it will be possible to increase the number of cells that can be analyzed using axonal BARseq. In general, the number of cells recovered by BARseq is determined by the size of the injection. In this study we restricted our injection to a single site, labeling a relatively small number of neurons. However, in previous work (Huang et al. 2020) we have barcoded more than 100,000 neurons in a single brain, and there is no technical barrier to labeling comparable numbers of neurons for axonal BARseeq in future studies. Furthermore, Sindbis virus can infect diverse species including primates (Zeisler et al. 2023), so axonal BARseq could potentially be modified to map projections in many model systems, especially those in which conventional tracing-based approaches are impractical. Axonal BARseq thus has the potential to emerge as a powerful tool for massively multiplexed mapping of single neuron projections in diverse model systems.



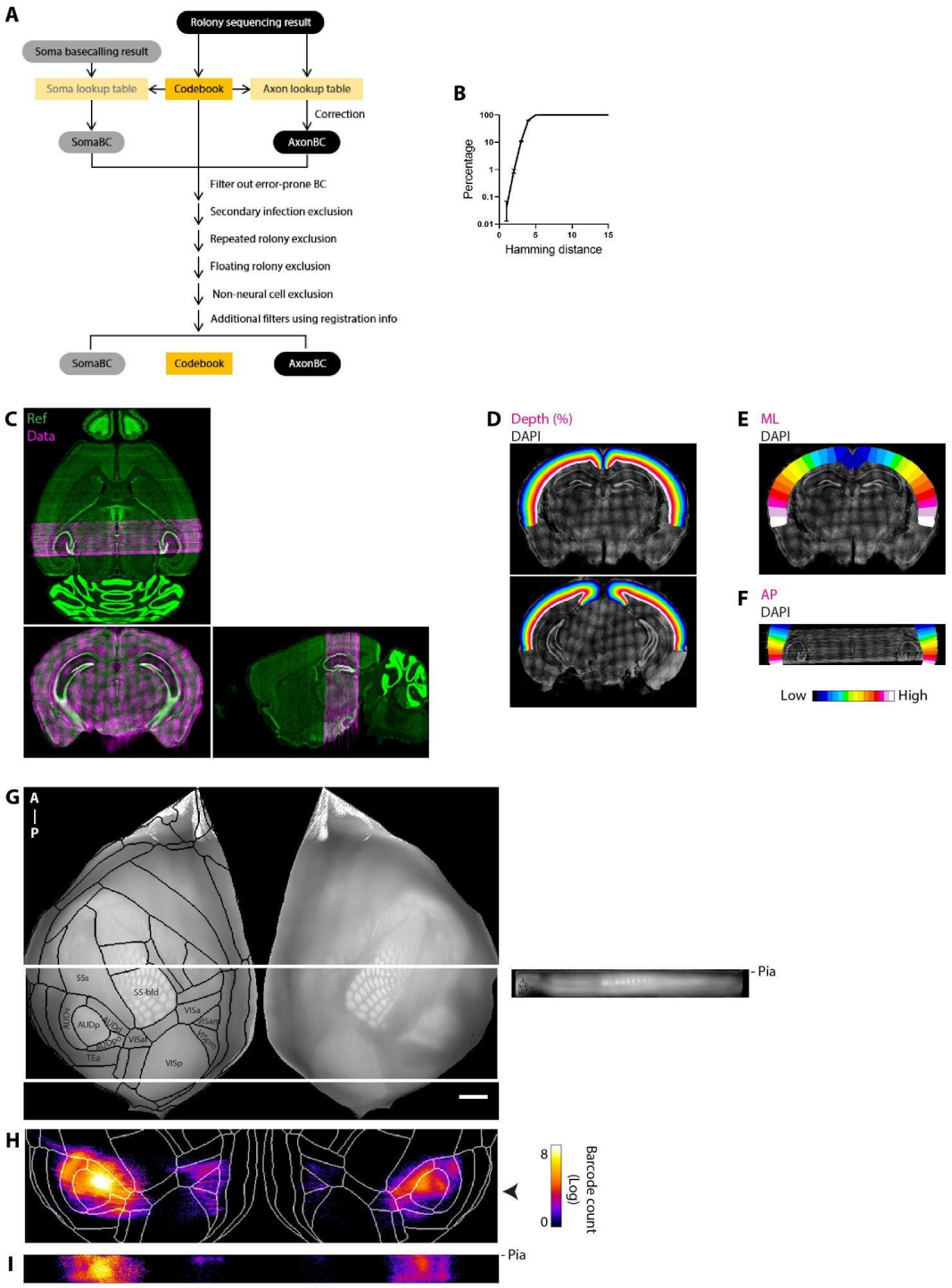
SupFig. 1 Optimization and efficiency estimation of axonal rolony preparation. (**A**) A brief summary of rolony preparation. Barcoded RNAs in axons and somata are first reverse transcribed into cDNA using reverse transcriptase. A padlock probe is then used to bind to the cDNA through adjacent regions of the barcode. The gap in the padlock is filled in using DNA polymerase, forming a circular single-stranded DNA molecule (ssDNA) through ligation by ligase. The circular ssDNA serves as a template for rolling circle amplification, resulting in the production of multiple copies of the barcode in a cluster of ssDNA called rolony. (**B-D**) Improvement of axonal rolony density using modified rolony preparation protocol. RA (RevertAid H Minus reverse transcriptase) and padlock probe XC1164 were used for reverse transcription and gap-filling for BaristaSeq (X. Chen et al. 2018). In the modified protocol, SSIV (SuperScript IV) and padlock probe LYO5 were used for reverse transcription and gap-filling. (**B**) Representative images of axonal rolonies using different rolony preparation conditions. Images were taken around the injection site in primary auditory cortex (AudI). Barcoded cells expressed both barcode RNA and GFP protein. (**C-D**) Both modifications yielded higher axonal rolonies density (one sample *t*-test). Each data point is the median of normalized rolony density from one brain section, two sections per mouse, total 3 mice. (**E-F**) The density of axonal barcodes was compared using two methods: FISH (RNAscope) and rolony preparation. Each barcoded RNA molecule has one GFP and one barcode. FISH detects the GFP component in the barcode RNA, while rolony preparation detects the barcode component. (**E**) Representative images of axonal barcode RNA around the injection site in AudI. (**F**) Comparison of the RNA density using FISH and the modified rolony preparation protocol (one sample *t*-test). Each data point is the median of normalized rolony density from one brain section, two sections per mouse, total 4 mice. Scale bar: top, 250 µm; bottom, 100 µm.



SupFig. 2 Data processing workflow and imaging processing. (**A**) The workflow of data acquisition and analysis for axonal BARseq. Additional information on image processing, alignment, basecalling, codebook construction, and barcode filtering can be found in SupFig 2-4. (**B**) The workflow of image processing. Raw data were acquired as z-stacks and then converted into maximum projections with background subtraction. Rolony intensity was then detected and used for further correction including bleedthrough, unequal channel intensity, and intensity decay during *in situ* sequencing. (**C-D**) Background subtraction effectively reduced the impact of uneven illumination, tissue and nucleus background on the images. The pixel intensity in the corrected images represented the signal beyond the background and was used for downstream analysis. Representative single tile from AudI, comparison of (**C**) single channel and (**D**) all four channels before and after background subtraction. Scale bar, 25 µm; color bar, single channel intensity.

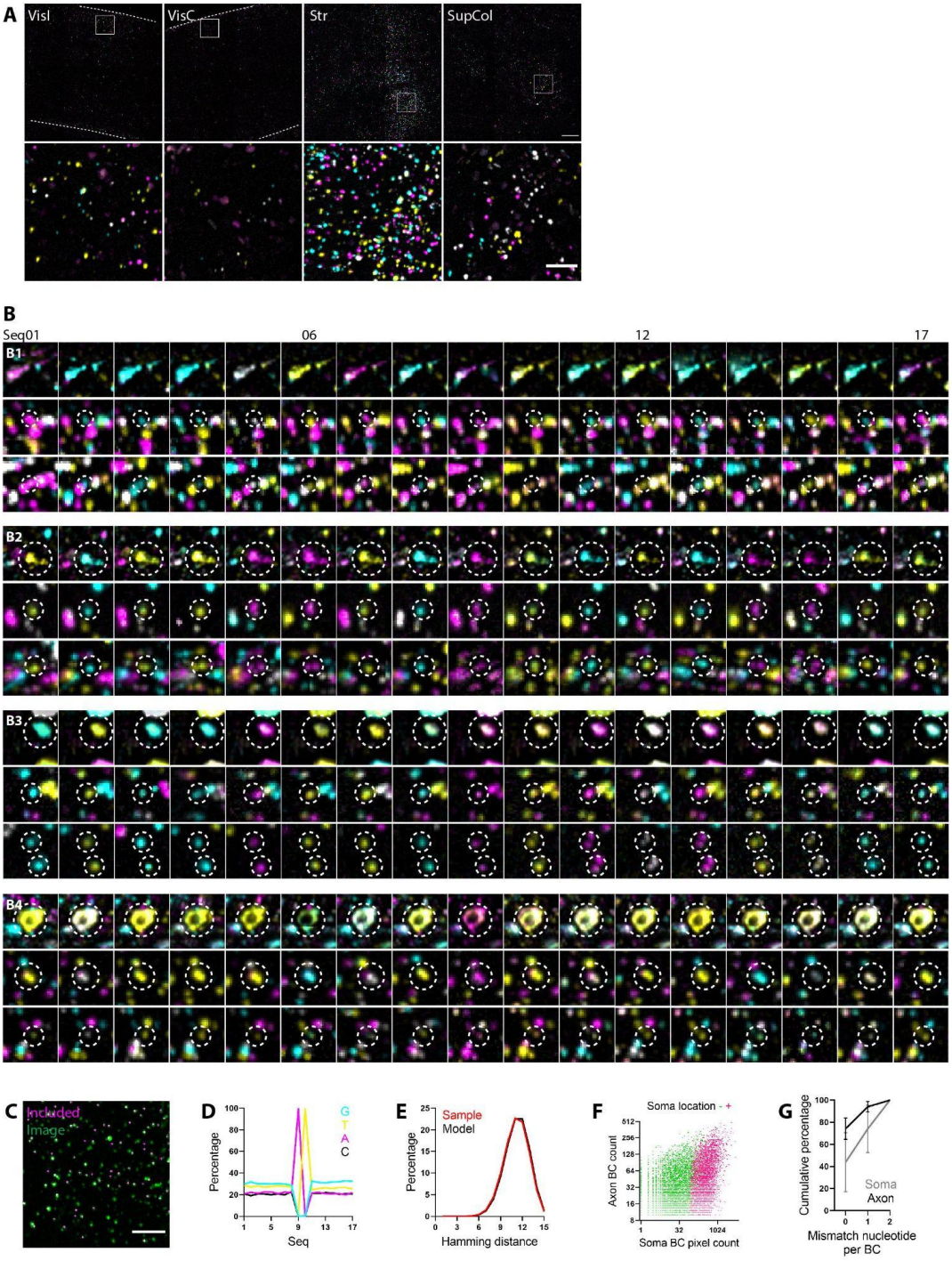


SupFig. 3 The processing steps for alignment, stitching and basecalling. (**A**) The workflow of alignment and stitching for axonal BARseq. Rolonies were identified as local maxima in individual tiles for each sequencing cycle. Tiles were first aligned across cycles using the locations of the rolonies, and then the aligned tile stacks were stitched together using the locations of the rolonies in the overlap regions. (**B**) Example of point cloud registration between neighboring sequencing cycles (magenta and green). 400 × 400-pixel area from AudI. (**C**) The workflow of rolony basecalling. During each sequencing cycle, unique IDs were assigned to individual rolonies (top row). Rolonies were then matched to rolonies in multiple previous sequencing cycles (middle row). For each match, a pair of IDs was identified, and pairs from different matches were merged or branched into an additional sequence (blue). Barcode sequences were assembled using the nucleotide associated with each ID. If no ID was found in a particular cycle, no nucleotide was assigned and the grid remains empty. (**D**) Representative images for non-specific signal correction in somata. Images were selected from Seq11 in AudI. Most of the somata had T in the previous sequencing cycle due to fixed barcode digits, and the signal was carried over to this cycle (top row). The correction effectively reduces the non-specific signals (bottom row). Scale bar, 50 µm.



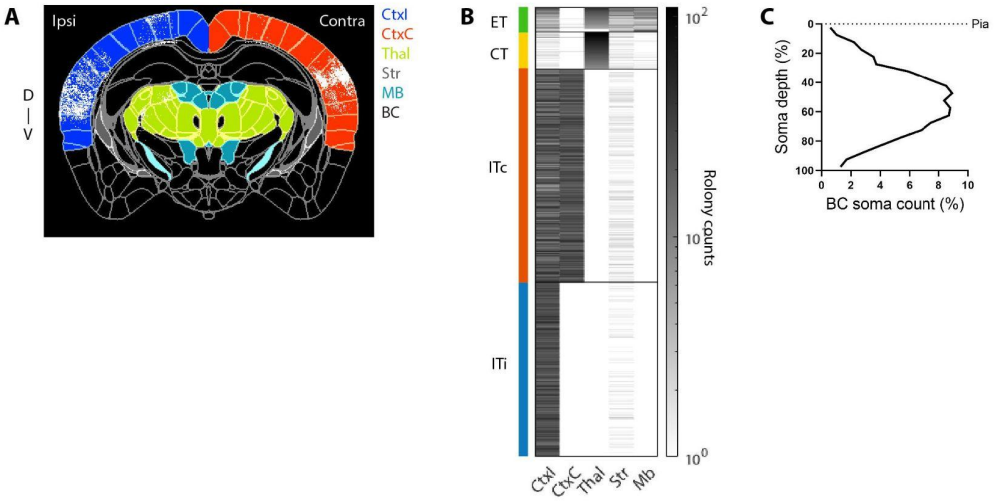
SupFig. 4 Barcode analysis and data registration. (**A**) The workflow of constructing a codebook and lookup tables for identifying somatic and axonal barcodes. (**B**) The possibility of barcodes having counterparts at various hamming distances. 10000 15-nt random generated barcode, mean ± SD from 100 iterations. (**C**) Representative plates showing the registration of the data to the Allen Brain Atlas common coordinate framework (CCFv3). Magenta, DAPI staining from registered data; green, Nissl staining of the reference atlas. (**D-F**) Representative images of cortical depth/ML/AP-value of the registered brain. Colormaps encode continuous gradients for depth/ML/AP-axis. E-F, coronal view; G, horizontal view. Gray, DAPI staining from the data. (**G**) Cortical flatmap

generated from CCFv3 average template. Grayscale: maximum intensity of the average template on ML-AP plate. White box: region of the axonal BARseq dataset. Area boundaries and annotation are shown in the left hemisphere. Area boundaries were drawn according to the most frequent area per pixel across cortical depth. Scale bar: 1 mm. Right panel, representative ML-depth view 125 µm slide of the right hemisphere. Top, cortical surface; bottom, the inner surface of the cortex. Depth percentage was converted to µm by assuming 1000 µm cortical thickness. (**H-I**) Distribution of barcodes on the flatmap. (H) ML-AP view. White, area boundaries; arrow, AP location of I. (I) Representative ML-depth view of barcode locations in cortex. Image included signals from 250 µm volume along the AP-axis.

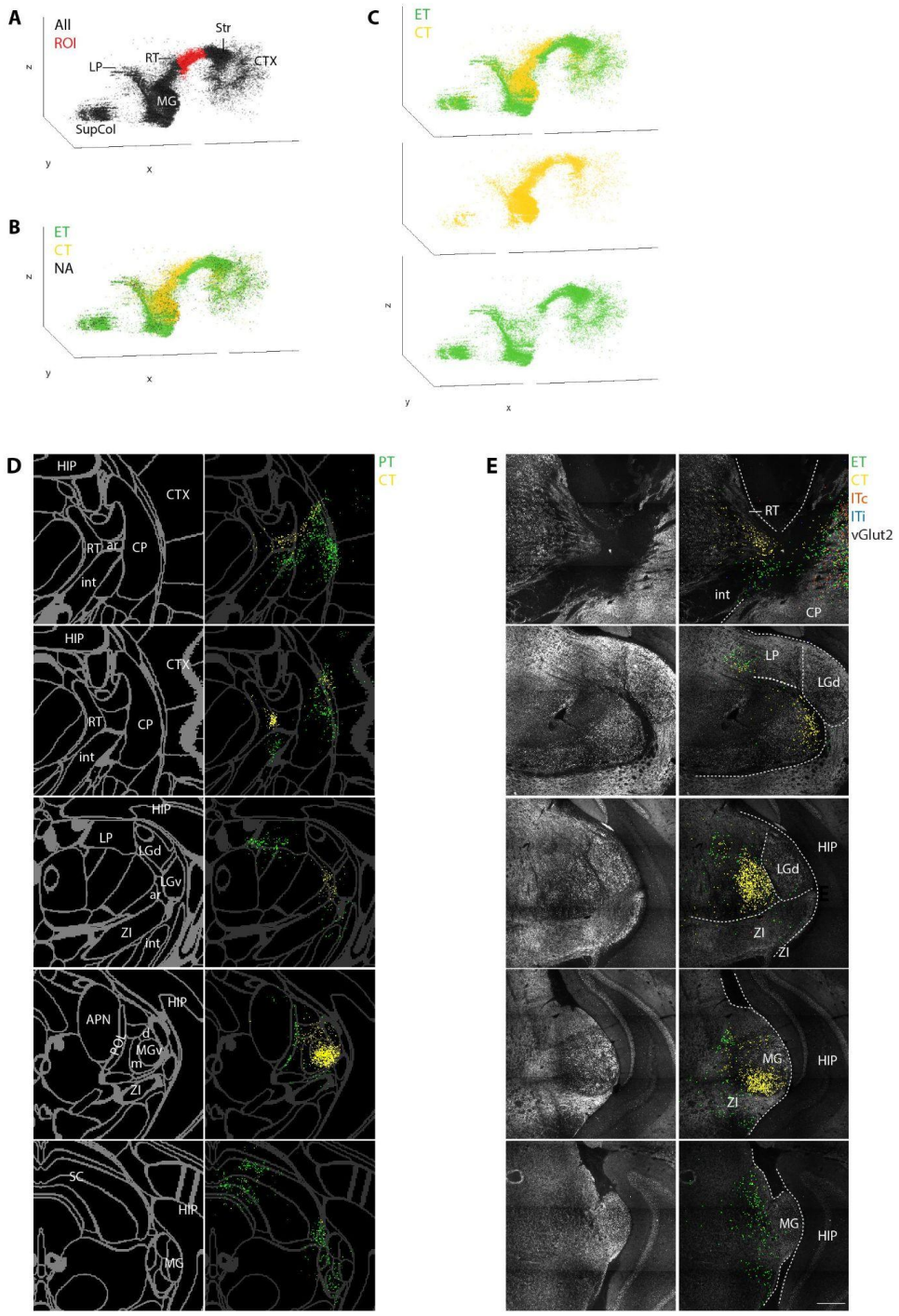


SupFig. 5 Example images and statistics of the dataset, related to Fig 1. (**A**) Representative images of axonal rolonies in VisI, VisC, Str and SupCol. Images are from Seq 01. Dotted line, cortical boundaries. Scale bar: top, 100 µm; bottom, 25 µm. Vis was identified according to CCFv3, it included posterior parietal cortex or parietal associative area (Lyamzin and Benucci 2019). (**B**) Additional examples of somatic and axonal rolonies from the same barcode are shown to illustrate the variability of sequencing signals in this dataset. Four examples are presented, with the top row showing images from somata and the middle and bottom rows showing images from axonal rolonies. The examples display differences in the size, shape, and intensity of axonal and somatic rolonies.

Some somata exhibit phasing (B3-4; phasing correction is described in SupFig. 3D), while others have clean signals (B2). In B2 and B3, duplications of the same rolonies were found in some sequencing cycles due to imperfect stitching, which was corrected in a later processing step (see Methods). (**C**) Example of the proportion of barcoded rolonies included in this dataset. Rolonies included in this dataset are indicated in magenta. Scale bar, 25 µm. (**D**) Nucleotide proportion of barcodes across sequencing cycles. The 9-10th nucleotides were fixed for this barcode library, 99.1% of the barcodes contained both fixed nucleotides. (**E**) The distribution of the all-to-all Hamming distance of the sample was similar to that of the model. Sample, 15-nt barcode without the two fixed nucleotides. Model, 10000 random generated 15-nt barcodes, mean ± SD of 100 iterations. (**F**) The relationship between the counts of somatic and axonal barcodes in individual barcoded cells. Cells were divided into two groups based on whether or not their soma location was identified. Cells that were somatic location-negative may still have barcodes detected in AudI, but their somatic locations were excluded from the analysis due to potential errors (see Methods). (**G**) Cumulative percentage of mismatch nucleotides per barcode in axon and soma. Barcodes with soma location were used to compute the soma barcode percentage; mean ± SD. Because the current codebook was constructed using axonal barcodes, it resulted in higher accuracy in identifying barcodes in axon compared to soma.

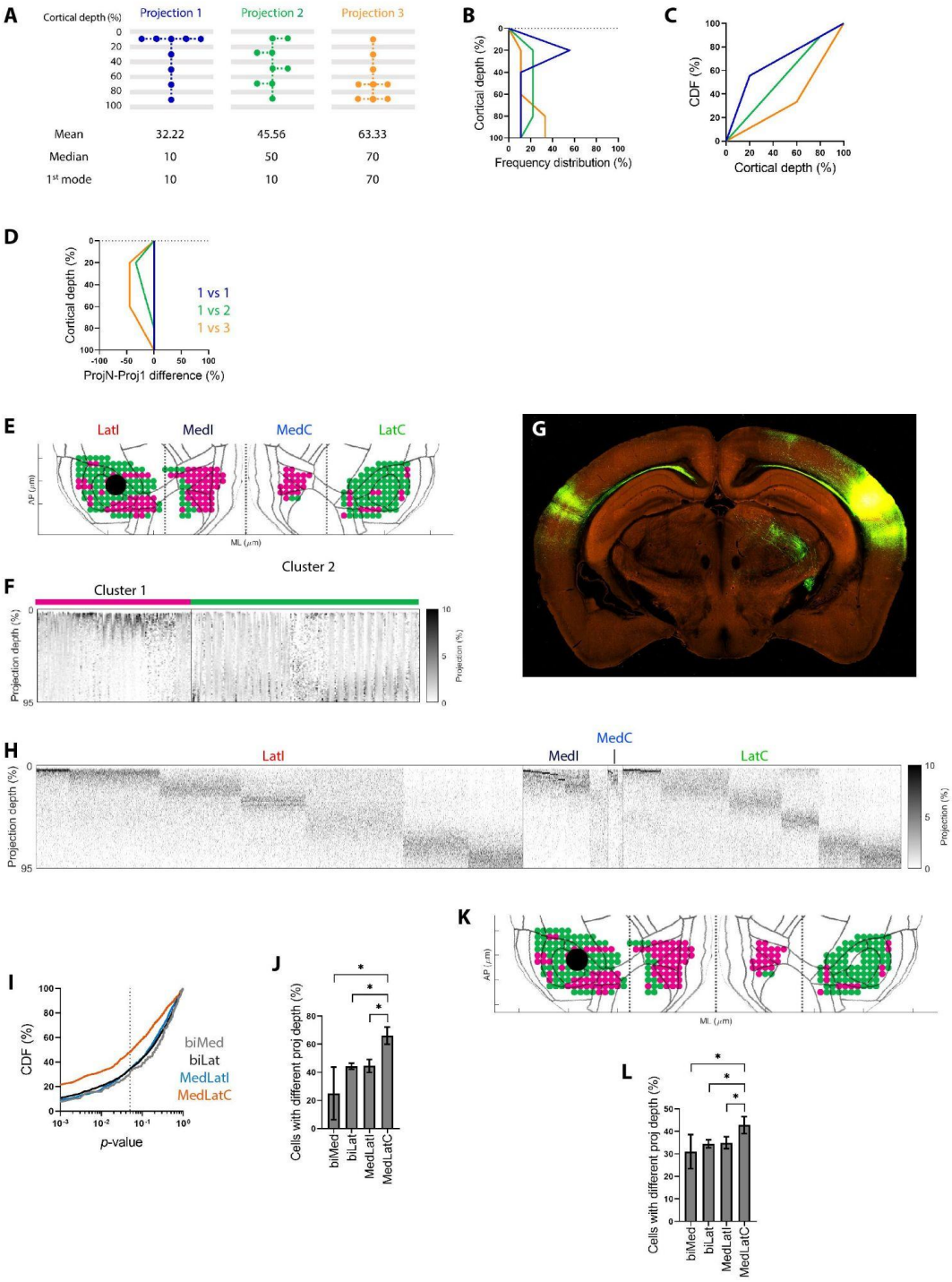


SupFig. 6 Projection targets of major cell types, related to Fig 3. (**A**) Representative coronal view of rolonies and five targeted brain regions. Rolonies located within the fiber bundles were not included in the five regions, 468026 of 492950 rolony were included. Target regions are color-coded; gray line: anatomical boundaries. CtxI, ipsilateral cortex; CtxC, contralateral cortex; Thal, thalamus; Str, striatum; MB, midbrain. (**B**) Heatmap of rolony counts in different targets from four cell types, ET, CT, ITc and ITi. (**C**) Distribution of barcoded somata along the cortical depth. Total 3698 barcoded somata. Bin size: 5% cortical depth.



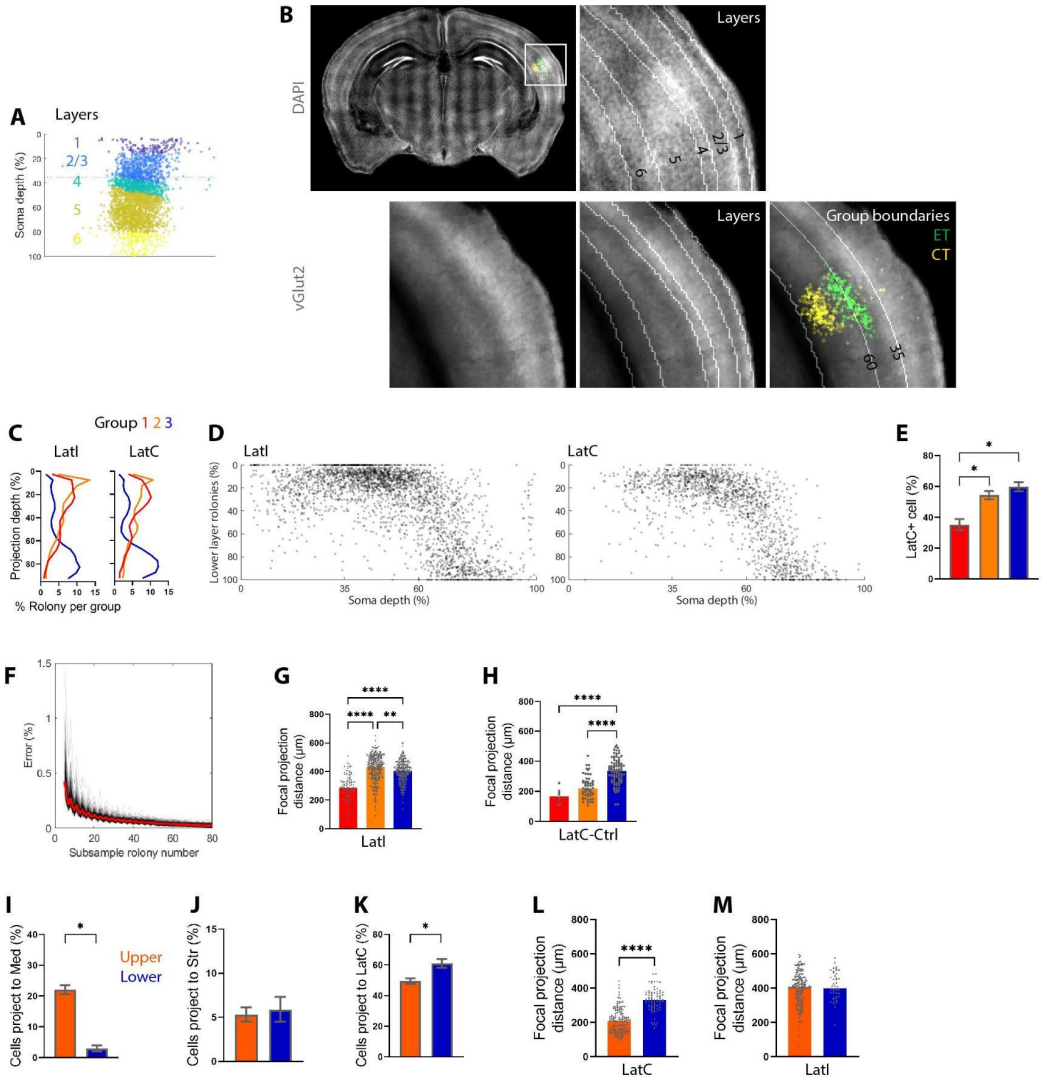
SupFig. 7 Division of CT and ET cells using projection targets and axonal trajectory, related to Fig 3. (**A-C**) The region of interest (ROI) and results from CT/ET grouping of Thal+ cells. (**A**) Side view of rolonies from Thal+ cells (black). Red, rolonies within the ROI (around RT and striatal-thalamic fiber tract), used for grouping CT and ET for Thal+ cells. (**B**) Initial grouping result using rolony within the ROI. Group identities were assigned to barcoded neurons according to their rolony locations within the ROI; individual rolonies were color-coded based on their cell identity. Black dots, rolonies from cells without rolony in ROI. (**C**) Final grouping result. Cells without rolony in the ROI were assigned to the group as their closest neighbors (see Methods). All Thal+ cells were assigned to either ET or CT. (**D**) Localization of registered CT and ET rolonies in striatum, thalamus and superior colliculus. Left column,

region boundaries with labels; right column, rolonies in the same region with tune-down boundaries for visualization. Each row is one coronal section, arranged from anterior (top) to posterior (bottom), from 6925, 7050, 7500, 8150 and 8775 µm in CCFv3, 25 µm/section. (**E**) Localization of CT and ET rolony with vGlut2 as an internal marker for thalamic structures. Left column, vGlut2 staining; right column, rolonies in the same region with tune-down vGlut2 signal for visualization. MG contains large but lower density of vGlut2 puncta (Hackett et al. 2016). To preserve the original resolution of vGlut2 staining, images and rolony locations were not registered. Each row is one coronal section, arranged from anterior (top) to posterior (bottom), from 220, 720, 1120, 1560 and 2020 µm in this dataset, 20 µm/section. Scale bar: 250 µm. CTX, cortex; CP, caudoputamen; ar, auditory radiation; RT, reticular nucleus; int, internal capsule/cerebral peduncle; HIP, hippocampus; LP, lateral posterior nucleus; LG(d/v), lateral geniculate complex (dorsal/ventral part); APN, anterior pretectal nucleus; POL, posterior limiting nucleus; MG(v/d/m), medial geniculate complex (ventral/dorsal/medial part); ZI, zona incerta; SC, superior colliculus.



SupFig. 8 Description and comparison of laminar patterns, related to Fig 4. (**A-D**) Example for representing and comparing cortical laminar patterns. (**A**) Diagram of three projections with distinct laminar patterns across the cortical depth. Dots represent rolonies and gray horizontal lines indicate cortical depth with 20% intervals. Projection 1 primarily projects to the superficial layer, projection 2 spreads across layers, and projection 3 primarily projects to the lower layers. The mean, median, and mode depths for each projection are listed below. Using a single measurement to describe laminar patterns can create bias in some cases. For example, median may be a better measurement than mean for a Poisson distribution, but it may not be as accurate for a multinomial distribution. Here, we described projection laminar patterns using frequency distribution (**B**) and cumulative distribution (**C**). (**D**)

Pairwise comparison of cumulative distribution of rolony depth. Each line represents the distribution difference (x-axis) at the cortical depth (y-axis). For projection 1 vs. projection 1 comparison, all the data points are 0 across depth due to both projections being completely the same. (**E-F**) Different laminar patterns of bulk projections in medial and lateral cortical targets. Data were computed and plotted using cortical flatmap. Cortical rolonies were binned into 250 × 250 μm bins on ML-AP plate; frequency distributions of rolony depth were calculated for each bin (1% depth per bin). Bins were split into two clusters using k-means and color-coded as magenta and green. Minimum 50 rolonies per bin, total 301 bins. (**E**) Spatial location of these bins on flatmap. Each bin is indicated by a dot and color-coded by cluster number. (**F**) Frequency distribution of rolony depth for all bins in each cluster, one bin per column. (**G**) Representative AUD bulk tracing result from C57BL/6J (experiment 158314278) from Allen Connectivity Atlas (Oh et al. 2014). Additional AUD tracing results from C57BL/6J (115958825, 146858006) and Emx1-IRES-Cre (576332845, 554421791, 562671482). (**H**) Different laminar patterns of single-cell projections in medial and lateral cortical targets. One cell per column; cell counts for each area: LatI, 6878; MedI, 1196; MedC, 212; LatC, 3931. F-G, bin size, 1% depth. (**I**) The possibility of similar laminar projection patterns in two different areas originating from a single cell. The possibilities were represented by *p*-values. The distributions were plotted as cumulative distribution functions. Dotted line: 0.05. (**J**) Control analysis for Fig. 4E, with minimum 15 rolonies per cortical target. Cell counts: biMed, 16; biLat, 1834; MedLatI, 484; MedLatC, 245. (**K-L**) The laminar pattern differences in the dataset were shown, with rolonies in the mirror somatic region of LatC excluded (LatC local-exclusion control). (**K**) Spatial locations of the two types of cortical bins as SupFig. 8E, the corresponding soma region in LatC is empty due to data exclusion. Total 297 bins. (**L**) The proportion of cells with different laminar patterns as Fig. 4E. *, significant difference with no overlap CI.



SupFig. 9 Projection differences across IT cell groups, related to Fig 5. (**A**) Layer annotations of Lat-projecting IT cells in Fig. 5A. Flatmap view. Dotted line: 35 and 60% depth. Layer annotations are from CCFv3. (**B**) Coronal view of the laminar boundaries of the three groups with annotations and internal markers. Zoom-in view: injection site. Top row: DAPI; bottom row: vGlut2. The 35% boundary approximately corresponds to the upper boundary of layer 4, which has relatively strong vGlut2 and DAPI signals (Nahmani and Erisir 2005; Coleman et al. 2010; Sato et al. 2022). The 60% boundary corresponds to the peak of the ET neuron distribution (green, also see Fig. 3C), and the weak vGlut2 band in upper layer 5 (Hirai et al. 2012; Oswald et al. 2013). The Images were generated from 312.5 µm volume of the registered data. Soma positions are sum projections and markers are median projections of the volume. (**C**) At bulk level, group 3 Lat-projecting cells mainly projected to lower layers while group 1/2 cells projected to all layers with preference to upper-middle layers. Lines represent the frequency distribution of rolony depth per group. Bin size, 5% depth. (**D**) Gradual increase of lower layer projection at single-cell level in Lat. One cell is indicated by a dot. (**E**) Group 1 has a lower percentage of cells projected to contralateral Lat cortex. Bar graph, median and confidence intervals. *, significant difference with no overlap CI. (**F**) Effect of rolony number on computing focal projection distance. LatC-projecting cells with minimum 80 rolonies were used for testing, and 5-80 rolonies were randomly sampled per cell. The error caused by downsampling was calculated using the distance difference between downsampling and ground truth (see Methods). Ground truth distance was represented by distance computed with all rolonies. Black line, median error from each barcoded cell, total 314 cells; solid red line, median of black lines. The median error < 5% when sample rolony number ≥ 55, therefore, the minimum rolony

number for this test was set to 55. (**G-H**) Projections from group 1 Lat-projecting cells were more focal compared to group 3 in LatI (G) and in LatC local-exclusion control (H). Cell counts: LatI group1 117, group2 286, group3 256; LatC-Ctrl group1 8, group2 56, group3 93. Kruskal-Wallis test, Dunn's test for multiple comparisons. (I-M) Morphological difference between lower and upper layer-projecting cells without soma location information. Lat-projecting cells were split into upper/lower layer-projecting cells using median depth in LatI/C: upper, ≤ 60%, lower, > 60%. In 1958 biLat cells, 209 (10.7%) were assigned to different groups in two Lat regions and they were excluded. Cell counts: upper, 2993; lower, 1025. Consistently, upper layer-projecting cells had a higher percentage of cells projected to Med (I), a lower percentage of cells projected to LatC (K), and more focal projections in LatC (L). Due to upper layer-projecting cells consisting of group 1 and 2 neurons, no difference was found in (J) and (M). I and K, *, significant difference with no overlap CI. L-M, Mann-Whitney test.

Methods

Animal processing and tissue preparation

All animal procedures were approved by the Institutional Animal Care and Use Committee at Cold Spring Harbor Laboratory (protocol 19-16-10-07-03-00-4). Experiments were performed on 7-10 week old male C57BL/6 mice (Charles River). The VAMP2nl SINV library (~4 million total barcodes) was injected into the primary auditory cortex using the NanojectIII (Drummond) at the following coordinates: -2.5 mm AP, ±4.2 mm ML, 0.9, 0.6, 0.3 mm depth, with a volume of 150 nL per depth. The mouse used for axonal BARseq was injected at -4.2 mm ML. After 2 days of expression, animals were anesthetized and perfused with 4% PFA in 1XPBS. The samples were post-fixed at 4°C for a day and then transferred to sucrose gradients (10-15%, 20-22%, 30% sucrose in 1XPBS at 4°C) and frozen in OCT. The brains were cryosectioned at 20 µm thickness, mounted onto glass slides using UV-solidified glue (Norland Optical Adhesive NOA81, 8-10 s UV) to minimize section distortion or detachment during high temperature and chemical treatments.

Rolony preparation

Before starting sample preparation, the sections were thawed and a hybridization chamber was placed on top. In the axonal rolony experiments, one section per chamber was utilized, while in the axonal BARseq experiment, two adjacent sections were used per chamber. To eliminate any residual fluids, chambers and samples were rinsed with water or reaction buffer before crucial reactions. For extended reactions or overnight reactions, humidified chambers were employed to prevent section dehydration. The catalog numbers and oligos utilized are listed in SupTable 2 respectively.

Sample pretreatment

Samples were washed twice with 1% PBSTE (1XPBS with 1% Tween-20 and 5 mM EDTA) and incubated in 1% PBSTE at 65 °C for 8-9 min. Next, they were placed on ice for 2 min and washed twice with 1% PBSTE. The samples were then dehydrated in 50%, 70%, and 85% ethanol and incubated in 100% ethanol overnight at 4 °C. After two washes with 100% ethanol, the samples were washed twice with water and 1% PBSTE to smooth the chamber. They were briefly washed in 4 mM HCl to adjust pH for pepsin digestion and then digested with 0.1-0.2% pepsin (w/v) in 4 mM HCl with 1 µM XC1215 at pH ~3 at 33 ℃ for 30-40 min. It is important to note that the activity of pepsin solution varies batch-to-batch and the activity of each batch was tested. Similarly, pH of the solution was monitored as low pH results in high nuclear background during *in situ* sequencing, while more neutral pH leads to low pepsin activity. Finally, the duration of pepsin digestion was closely monitored as over-digestion can cause tissue/rolony to degrade/tear fall off, while insufficient digestion can lead to low permeability and low rolony density.

Reverse transcription

After digestion, samples were washed twice with 1% PBST and then washed in 1X SSIV (SuperScript IV) buffer containing 0.4 µg/µL BSA and 5 mM DTT for 5-15 min at room temperature. Reverse transcription was performed on the samples using 1 µM XC1215, 20 U/µL SSIV, 500 µM dNTP, 0.2 µg/µL BSA, 1 U/µL RiboLock RNase Inhibitor, and 5 mM DTT in 1X SSIV buffer at 45 ℃ for approximately 4 hours. Samples were then transferred to a new reaction mix and incubated overnight at 45 ℃.

For the reagent comparison experiment (SupFig. 1B-C), samples treated with RA (RevertAid H minus reverse transcriptase) were washed in 1X RA buffers containing 0.4 µg/µL BSA for 5-15 min at room temperature. Reverse transcription was performed on the samples using 1 µM XC1125, 20 U/µL RA, 500 µM dNTP, 0.2 µg/µL BSA, and 1 U/µL RiboLock RNase Inhibitor in 1X RA buffer at 37 ℃ for approximately 4 hours. The samples were then transferred to a new RA reaction mix and incubated overnight at 37 ℃.

After reverse transcription, the samples were washed with 1X PBS and crosslinked with 25 mM BS(PEG)9 in 0.2% PBST for 30 min at room temperature. They were then washed with 0.2% PBST (0.2% Tween) and incubated in 2 mM lysine in 1X PBS for 30 min.

Gapfilling

After crosslinking, the samples were washed with 0.2% PBST twice and water twice, and then gapfilled with 100 nM padlock probe LYO5, 0.5 U/µL Ampligase, 50 µM dNTP, 0.4 U/µL RNaseH, 50 mM KCl, 20% formamide, and 12.5 mU/µL Phusion in 1X Ampligase buffer at 37 ℃ for 30-40 min, and 45 ℃ for 45 min. To prevent Phusion from reacting with the padlock, the reaction mix was kept cold and Phusion was added immediately before the reaction. In reagent comparison (SupFig. 1B and D), padlock probe LYO5 was replaced by XC1164.

Rolling circle amplification (RCA)

After gapfilling, the samples were washed thoroughly with 0.2% PBST and rinsed with water. They were then incubated with RCA mix (1 U/µl EquiPhi29 polymerase, 0.25 mM dNTP, 120 µM aadUTP, 0.2 µg/µL BSA, and 1 mM DTT in 1X EquiPhi29 buffer) at 37 °C overnight. After incubation, the samples were washed with PBS once and crosslinked with 25 mM BS(PEG)9 in 0.2% PBST for 15 min at room temperature. They were then washed with 0.2% PBST twice and quenched with 1M Tris pH 8.0 for 30 min. This RCA-crosslinking step was repeated two more times. After three rounds of RCA, the samples were crosslinked with 25 mM BS(PEG)9 in 0.2% PBST for 30 min at room temperature. They were then washed with 0.2% PBST twice and quenched with 1M Tris pH 8.0 for 30 min.

Axonal barcode detection comparison

To measure the sensitivity of rolony preparation, we compared it to FISH, a standard method with high single-molecule sensitivity. In these experiments, rolonies were hybridized with fluorescence-conjugated probes. After rolony preparation, the samples were hybridized with 0.25 µM probe XC92 in 2X SSC, 10% formamide for 15-30 min at room temperature. Any excess probes were washed away with 2X SSC, 10% formamide three times and 0.2% PBST three times.

In the experiments used to compare rolony preparation and FISH, the FISH samples were pretreated in the same way as the rolony preparation samples. After digestion, they were washed, and FISH was performed using GFP probes and the RNAscope kit according to the manufacturer's protocol.

Quantification was performed using max-projected and stitched images. Similar regions of interest were manually selected in the AudI/AudC/Thal/VisI for each brain section, avoiding somatic rolonies. Rolony counts were measured using 'Find maxima' with fixed prominence in Fiji, and density was calculated by dividing the area size. Rolony densities were normalized to the density of the same region in neighboring SSIV + LYO5 samples. The median of the normalized density was calculated from 2-4 regions per section.

Interestingly, we found that a 2 nt length difference between padlock probes LYO5 and XC1164 significantly affected rolony signals. This may be because template length affects Phi29 efficiency during RCA(Joffroy et al. 2018). While the modified protocol generated more axonal rolonies, it was less cost-effective for producing somatic rolonies. Therefore, for somatic barcode sequencing, the original BaristaSeq protocol (X. Chen et al. 2018) is sufficient due to the high abundance of somatic barcodes.

Axonal and somatic rolony comparison

Probe-hybridized samples in SupFig. 1E-F were used for comparison of axonal and somatic rolonies. For each brain section, 10-20 somatic areas were manually selected in AudI, and somatic rolony intensities were represented by the maximum intensity of each somatic area. In the same stitched images, 6-7 300x300 pixel ROIs were manually selected for axonal rolonies, avoiding somatic rolonies. Within each ROI, axonal rolonies were identified using 'Find maxima' in Fiji, and axonal rolony intensities were represented by the intensity of the maxima. The median intensities of axonal and somatic rolonies were calculated for each section, with background subtraction.

***In situ* sequencing**

Axonal BARseq samples were split into seven rounds of rolony preparation (SupTable 1). For each round, the samples were divided into two batches for *in situ* sequencing. After rolony preparation, the samples were incubated in 2X SSC, 80% formamide at 65 ℃ for 15 min. 2.5 μM sequencing primer LYO23 was hybridized to the rolonies in 2X SSC, 10% formamide for 15-30 min at room temperature. Any excess primers were washed away with 2X SSC, 10% formamide three times and 0.2% PBST three times. *In situ* sequencing was performed using the HiSeq Rapid SBS Kit v2. The reagents used in this process included the Universal Sequencing Buffer (USB), Cleavage Reagent Mix (CRM), Cleavage Wash Mix (CWM), Incorporation Master Mix (IMT), and Universal Scan Mix (USM). Before the first cycle, the samples were washed with USB at 60 ℃ for 4-5 min twice. Then, they were incubated with CRM at 60 ℃ for 5 min. The samples were washed with CWM, 1% TT (20 mM Tris pH 8.0, 1% Tween-20) three times and PBS twice. Next, they were blocked with iodoacetamide (9.3 mg tablet in 2 mL 1XPBS) at 60 ℃ for 4-5 min and washed with 0.2% PBST three times. For each sequencing cycle, the samples were washed with USB at room temperature twice and incubated with IMT at 60 ℃ for 4 min. They were then washed with 1% TT with 5 mM EDTA once, and 1% TT at 60 ℃ for 4 min 3-5 times. The samples were incubated in USM and were ready for imaging. After imaging, the samples were washed with 1% TT three times and USB twice, incubated with CRM at 60 ℃ for 4 min, and washed with CWM. In the later sequencing cycles, the C-channel often had a high level of nonspecific background, additional 1% TT washes were included to decrease this background. In round 1 of this dataset, the samples did not receive CRM treatment before the iodoacetamide incubation prior to Seq01. Additionally, an additional iodoacetamide treatment was applied after the first CRM step following Seq01 imaging.

Immunohistochemistry

After the final sequencing cycle (Seq17), the samples were treated with CRM and CWM to remove any remaining sequencing signals. They were then blocked with 5% BSA in 1XPBS and incubated with a vGlut2 antibody (1:500) in 2% BSA in 1XPBS at 4 ℃ for 2 days. Following washes with 0.2% PBST, the

samples were incubated with a secondary antibody (1:1000) in 2% BSA for 2-4 hours at room temperature. After additional washing, the samples were stained with DAPI and imaged using USM.

Microscopy

Images were obtained using a Nikon TE2000-E microscope equipped with a X-Light V2 spinning disk (Crest Optics), Prime 95B camera (Teledyne Photometrics), and LDI-7 laser diode illuminator (89North). A 20X Plan Apo objective (Nikon) was used for all experiments. It is important to note that factors such as optical distortion and uneven illumination in the microscope system can affect the sensitivity and accuracy of axonal BARseq. All images were taken as z-stacks with the following settings: 0.55 µm per pixel, 12-bit depth, a total of 17 stacks with 3 µm intervals, and 15% overlap for tiling. The lasers and filters used for each channel were listed in SupTable 2. Briefly, each nucleotide was imaged in a separate channel during sequencing. We found that maximum intensity projection preserved most of the signal while resulting in smaller file sizes and reduced computational demands during analysis. As a result, we converted the z-stacks to max projections.

Image processing and rolony identification

The general workflow for data acquisition and analysis is described in SupFig. 2A. The imaging processing workflow for *in situ* sequencing is described in SupFig. 2B. To reduce fixed pattern noise during *in situ* sequencing, we subtracted the 3rd lowest intensity plane of the z-stack from the maximum projection image. Local background subtraction was performed by taking advantage of the fact that pixels without a barcode have discontinuous intensity profile along the z-axis. This process effectively removed local signal distortions and backgrounds such as uneven illumination, nuclear, and tissue background (SupFig. 2C-D). However, while this method was effective for axonal rolonies in target areas, errors were encountered in pixels around the somata due to the high signal density and the aberrant intensity distribution along the z-axis compared to single rolonies. To correct for bleed-through, uneven channel intensity, and intensity decay across sequencing cycles, we based intensity corrections on local maxima for each experiment. To decrease variability between individual batch, we used z-scores for intensity correction. Rolonies were typically between 3-7 pixels in diameter on the maximum projection images. Therefore, we identified local maxima within a 5-pixel diameter range for each sequencing cycle. A local maximum was considered a rolony location if it met the following criteria: (1) in the z-stack, the slides with the highest intensity were neighboring slides (e.g. the 1st max intensity slide was next to the 2nd/3rd max intensity slides); (2) the channel with the local maximum had the highest intensity before and after image corrections; (3) the channel intensity passed the intensity ratio filter (i.e. 2nd max/1st max < 0.95); (4) the max channel intensity passed a threshold. To improve the accuracy of matching rolonies during base-calling, we calculated the subpixel locations of local maxima using interpolation.

For immunohistochemical experiments, maximum projections of image tiles were stitched into whole coronal sections using phase correlation, with max projection in the overlapping region.

Tile alignment and stitching

The workflow for alignment and stitching is described in SupFig. 3A. The alignment process consists of two steps: (1) pre-alignment using stitched images; (2) point cloud registration for individual tiles.

During pre-alignment, tiles from the same image were initially stitched using imaging positions and then aligned across sequencing cycles using phase correlation. We used imaging position-based stitching to avoid errors from intensity-based algorithms. Additionally, stitched images were aligned to 1-3 sequencing cycles to minimize errors. After pre-alignment, rolony coordinates were aligned to the reference or neighboring sequencing cycles using a projective/affine transformation (SupFig. 3B). The transformation matrix between point clouds was calculated using a frequency-based algorithm. To reduce the impact of tissue distortion during sequencing, we used mid-sequencing cycles (Seq08/09) as reference cycles. Vis of section #76 was excluded during alignment.

For stitching, we combined and aligned the rolony coordinates from nearby tiles across sequencing cycles to the neighboring tiles. To minimize errors, we stitched tiles with a lower number of rolonies to tiles with a higher number of rolonies.

Rolony base-calling

Our rolony base-calling pipeline allowed for a degree of error during alignment. No non-linear transformations were performed during alignment. Base-calling was performed by matching nearby local maxima (dots) across sequencing cycles (SupFig. 3C). Dots were first given unique IDs in each sequencing cycle, and dots from later cycles were one-to-one matched to the closest available dots in the previous cycle within a 5-pixel range. The sequence of dots was then assembled, and the nucleotides associated with the dots were identified as the rolony barcode.

During sequencing, rolony signals may be lost, shifted, or near a strong non-specific noise signal. To maximize continuity in the sequencing results, each rolony was matched to rolonies in three previous cycles. To assemble the sequence, matches were merged sequentially from a 0 to 3 cycle interval in ascending order (e.g. 3-4, 2-4, 1-4, 4-5, 3-5, 2-5, etc.). Non-base-called nucleotides were assigned to intervals when two matching cycles were not consecutive. If there was a disagreement between the current match and the existing sequence, the previous sequence was duplicated to include the new match (SupFig. 3C, blue). Barcodes with more than 3 continuous non-base-called nucleotides were discarded. However, due to the high density of signals and resulting higher error rate near the injection site, we excluded rolonies close to the soma (within 20 μm of ≥ 35 soma pixels in > 2 cycles) from the analysis.

Soma base-calling

Since barcoded somata were larger in size compared to individual rolonies, we base-called somata using pixel location rather than local maxima. We identified the barcodes as the channels with the highest intensity across sequencing cycles using stitched images from the AudI. For technical reasons, the stitched images used for soma base-calling only included one tile for overlap areas.

During later sequencing cycles, we noticed that barcoded somata showed phasing signals, but single rolonies did not. This may be because somata contain a larger number of barcoded single-stranded DNA, and the protocol was not optimized for soma barcode sequencing. To digitally correct this, we subtracted the pixel intensity from the previous cycle (50% for max intensity and 100% for the rest), which improved the signal-to-noise ratio (SupFig. 3D).

Registering image to brain volume

We identified and imaged targeted brain regions separately for *in situ* sequencing. After alignment and stitching, we registered the images to the whole coronal section using nuclear signals by phase correlation. We then aligned the coronal sections into a 3D volume using control point pairs. These point pairs were selected manually between nearby sections, and displacement fields were generated from polynomial2 and piecewise linear transformation within manually defined limits and corrections. One brain section (section #108) was excluded due to severe distortion.

Codebook and lookup table

We used the results of the *in situ* sequencing to construct a list (codebook) of infected barcodes. The two sources of barcode combinations are axons and soma. As this study focused primarily on axonal projections, we chose only axonal barcodes for our codebook (SupFig. 4A).

We tolerated 1-2 nt errors in our rolony base-calling protocol, since the base-calling process could have errors either due to single mutations during sample preparation or misalignments during analysis. This ensured the accuracy of the codebook/lookup table and minimized data loss.

As described above, the rolony base-calling procedure can base-call a rolony multiple times, resulting in a set of barcodes with and without errors. To construct our final codebook for this dataset, we made the following assumptions: (1) a true barcode can be found in multiple rolonies (≥ 3 rolonies); (2) a true barcode has a higher count than its erroneous versions; (3) there are no pairs of true barcodes within 1-hamming distance (<0.1%, SupFig. 4B); (4) a true barcode consists of different nucleotides (< 14 same nucleotides) and meets length limits (with ≥ 13 of 15 nucleotides base-called, ≤ 3 continuous non-base-called nucleotides). Based on these parameters, our codebook consisted of 13919 barcodes.

Lookup tables were used to match individual axonal and somatic barcodes to the codebook. A hamming distance of 2 was set as a cut-off to match axonal and somatic barcodes to the codebook barcodes. Barcodes that matched more than one codebook barcode within the minimum Hamming distance were discarded. During this process, non-base-called nucleotides were treated as a match to all four nucleotides if there was no mismatch; otherwise, they were treated as a mismatch. All nucleotides were included in the Hamming distance calculation at this step. To minimize misassignment, we constructed the codebook and lookup table before filtering, as eliminating a potential barcode early on may result in its axonal rolonies being assigned to another barcoded cell within the maximum Hamming distance.

Axonal barcode correction

During axonal base-calling, it was possible for a single axonal rolony in one sequencing cycle to link to more than one rolony in another sequencing cycle. This can result in (1) one rolony belonging to multiple different barcodes; (2) the same axonal rolony being called multiple times and linked to different rolonies in other cycles, but belonging to the same barcode. To address these issues, we took the following steps: (1) a rolony in a cycle linked to more than one barcode was excluded and the cycle was assigned as non-base-called; (2) barcodes that did not meet the requirements for length and interval were excluded; (3) barcodes with similar sets of rolonies were condensed into the one with the most base-called digits.

The Hamming distance between a pair of barcodes was calculated as the total number of mismatches between them. By default, non-base-called nucleotides were treated as a match to all four nucleotides. In this SINV library, the 9th and 10th nucleotides were fixed and therefore excluded from the Hamming distance calculation, unless stated otherwise.

Soma identification

Soma barcode counts were determined by counting the number of pixels associated with a specific barcode at the injection site (AudI). However, these counts alone were not always reliable for identifying barcoded somata in our current setup, potentially due to the following factors: (1) some somata were cut and split into two neighboring sections during sectioning; (2) loss of surface area of the section during sample preparation (e.g. due to over-digestion by pepsin); (3) weak signals in deep areas of the section due to insufficient permeabilization during sample preparation; (4) low soma barcode counts in some cells; (5) base-called areas appearing smaller than they should be due to alignment, stitching errors, and phasing; (6) soma base-calling being sensitive enough to identify dendritic, and occasionally axonal, rolonies in AudI.

To identify valid soma locations, we identified the brain section with the highest sum intensity of soma pixels as the soma section, and within this section, we identified the XY coordinates of the brightest pixels as soma locations. A soma needed at least 80 counts of its barcode within 100 μm of its location. Barcodes that did not meet these criteria were identified as barcodes without soma locations. In the registrated data, the median distance between somata and the injeciton center is 267 μm.

Barcode Filters

Filtering out error-prone barcodes. To reduce the number of nonspecific barcodes, we applied the following filters: (1) barcodes with > 6 continuous identical nucleotides were excluded (152 out of 13919); (2) barcodes with more than 14-nt in Ch1/2 or Ch3/4 were excluded (127 out of 13767); (3) barcodes with more rolonies with 1 or 2 mismatches compared to no mismatches were excluded (82 out of 13640). Non-base-called nucleotides were treated as a match to all four nucleotides at this step. Only barcodes that passed the count filter (1) with ≥ 10 rolonies in at least one target region, (2) with ≤1000 and ≥ 3 axonal barcode counts, and (3) with ≤ 7000 somal barcode counts were included for analysis (9185 out of 13558 were included).

Secondary infection exclusion. We observed secondary infection in target brain regions, and most of the infected cells had a glial morphology. We manually identified 17 barcodes from these cells in all regions except AudI. Barcodes within a 4-Hamming distance of these identified barcodes were excluded from the analysis (92 out of 9185).

Repeated rolony exclusion. To avoid double-counting, we excluded repeated rolonies in overlapping imaging fields, such as the cortex and striatum. Specifically, in these overlapping areas, we only included one copy per barcode from different fields (exclusion range < 25 μm).

Floating rolonies identification and exclusion. We observed that rolonies could sometimes float out of the soma and settle within a surrounding area. Among barcodes without a soma location, we also observed this floating rolony effect. Since the soma locations were unknown, we could not use the soma section to exclude these floating rolonies. Therefore, we used an alternative method to identify sections with

floating rolonies. We used two criteria for identifying these sections: (1) the slide (and sometimes its neighboring slide) was the only one with rolonies in specific areas, and (2) the rolonies on the slides were widely and sparsely distributed. To identify rolonies that meet criterion (1), we excluded rolonies with neighbors (< 140 μm) in other sections (> 1 section away). To test whether criterion (2) was met, we identified a section to have enough rolonies that were far apart (≥ 3 rolonies/clusters with a distance beyond 50 μm). For barcodes with more than one such section, we selected the one with the widest rolony coverage. We used AudI, Thal, and Vis to find slides with floating rolonies.

We used this algorithm to identify sections with floating rolonies in cells with and without soma locations. Verification using cells with soma locations showed that the algorithm identified floating rolonies in 37.1% of cells. Within these positive barcodes, the algorithm had an accuracy of 94.8% for identifying the range of sections for the soma (± 1 section). For cells without soma locations, the algorithm detected floating rolonies in 14.9% of barcodes. We manually validated 75 positive barcodes and found a false positive rate of 22.7%. We achieved 100% accuracy for identifying the range of sections for floating rolonies. False positives would result in the exclusion of true rolonies in a 40-60 μm area in selective targets, but since projections usually extend more than 200 μm, this had a limited effect on downstream analysis.

To exclude floating rolonies, we excluded axonal rolonies in AudI, Thal, and Vis from 2-3 sections around the soma sections for barcodes with soma locations, and from the floating rolony section for barcodes without soma locations. After applying these exclusions, 8838 barcodes passed the count filters.

Non-neural cell exclusion. SINV can infect both neurons and non-neural cells(Shahar et al. 1990). However, we were unable to distinguish between these cell types due to the lack of cell type markers. To exclude non-neural cells from our analysis, we applied distance and counts criteria. Specifically, cells needed to have ≥ 3 axonal rolonies ≥ 200 μm from the soma or center of axonal rolonies in AudI, and 50 cells were excluded using this criterion. This criterion was applied because non-neural cells typically do not have long projections. It is worth noting that this process may also filter out neurons with short local projections.

Additional filtering after CCF registration. After registering the data to the CCFv3 reference frames we applied the following additional filters. We first deleted rolonies outside the CCFv3 brain area and performed additional floating rolony elimination in the hippocampus, ventricle and fimbria of CCFv3. Next, we set a minimum rolony counts for five major targets: 5 for the ipsilateral/contralateral cortex and thalamus, 3 for the striatum and midbrain. After filtering, we excluded 167 barcoded cells with < 10 counts in any imaging region, as well as 16 non-neural cells. It is worth noting that these steps are optional and can be skipped. After the above-mentioned steps, we identified 8620 barcodes, including 3700 with soma location. Four barcodes had single-digit non-base-called nucleotides.

Manual assessment of base-calling results. To assess sensitivity and accuracy of our automated base-calling pipeline, we compared to manual base-calling. To evaluate the sensitivity of axonal rolony base-calling, we calculated the percentage of base-called rolonies in 17 randomly selected Seq14 images from target areas (3-6 ROIs per image, 300 x 300 pixels). Sensitivity was 44.5%±9%. To estimate the accuracy of axonal rolony base-calling, we randomly selected 18 barcodes and found that 0 out of 60 (0%) rolonies had > 2 nt disagreements between the codebook and evaluator.

To evaluate the efficiency of soma base-calling, we manually selected 50-112 somata per image in 7 randomly selected sequencing images, and found that 63.9%±7.6% of somata were base-called. To estimate the accuracy of soma base-calling, we randomly selected 40 barcoded somata and manually base-called them, and found that 5 (12.5%) had > 2 nt disagreements between the codebook and evaluator. To evaluate the accuracy of soma location, we randomly selected 90 barcoded somata and

found that 79 (87.8%) were in agreement with the evaluator's assessment. It is worth noting that there was high signal density near the injection site, which may have contributed to some uncertainty in the evaluator's assessments.

Registering to Allen mouse brain CCFv3

To align the 3D data volume with the CCFv3(Wang et al. 2020), we used a manual linear registration process. We then applied nonlinear adjustments to the coronal plates using control point pairs, similar to the method used for image registration to the brain volume. We used the nissl reference map for this process. All reference maps used in this study (nissl, average template, and annotation map) had a voxel resolution of 25 µm.

Cortical flatmap and ML/AP/depth-coordinates

To compare projections across cortical regions and hemispheres, we generated a lookup table for a cortical flatmap from the CCFv3 (SupFig. 4D-I). The concept of this flatmap is similar to that of(Wang et al. 2020)), but with some differences. The flatmap coordinates consisted of three axes, with one axis oriented in the same direction as the cortical columns and the other two on a plate perpendicular to the columns. We defined the outer and inner cortical boundaries using the outer boundaries of layers 1 and 6, respectively. To determine the direction of the cortical columns, we calculated the lines from each outer boundary voxel to the closest inner boundary voxel. The depth percentage was the percentage of cortical depth along individual column lines. Due to cortical curvature, the distance between two column lines may vary at different depths (i.e., the distance is larger in upper layers compared to lower layers). We chose the mid-cortical plate (~50% depth) as the reference plate for the other two axes. The values of the other two axes were calculated as the cumulative sum of voxel-to-voxel distances on the plate, and column lines were assigned to the values of their intersections at the reference plate. To obtain continuous, smooth values in all three axes, we applied an average filter.

Following general practice, we divided the reference plate into medial-lateral (ML) and anterior-posterior (AP) axes. Our goal in creating the flatmap was to simply flatten the cortex for physical distance, rather than attempting to represent biological gradients. We determined the definition of each axis with the following considerations: (1) rotating the brain around the x-axis in 3D space changes the direction and voxel value of the AP-axis; (2) similar to earth mapping, it is impossible to get a flat and continuous cortical plate without distorting the direction of the axes or the point-to-point distance due to cortical curvature. Although the midline is generally considered the 'medial' part of the brain, we found that setting the midline as a fixed value to flatten the cortex with this algorithm caused relatively more distortion near the lateral region. Therefore, we used principal component analysis (PCA) on the reference plate of the right hemisphere to define the 1st axis as AP and the 2nd axis as ML, and verified this visually. The contour line of the median AP and ML values was used as the reference line for flattening. The AP value was assigned as the distance to the AP reference line on the reference plate, and the ML value was assigned as the distance to the ML reference line with the minimum AP value change along the reference plate. The minimum value in both axes per hemisphere was set to 1. This flattening method was unable to differentiate between cortical regions folded towards the midline and the increased distortion at the lateral edge. However, since these were not the target areas of this dataset, the flatmap algorithm did not adjust for them. Additionally, this flatmap preserved the subtle voxel difference between the left and right hemispheres in CCFv3.

We used these lookup tables with interpolation to convert the registered rolony and soma locations into ML/AP/depth coordinates. To minimize interpolation error near the edges, we applied a 50 µm non-zero average filter to the outer edge of the cortex. For visualization purposes, we assigned the ML-values of

the left cortex as negative and the right cortex as positive, and excluded the range of the AP-axis without cortical rolonies. Two of the 3700 somata were excluded during this process.

After registration, we defined the areas of four major projection targets (cortex, thalamus, striatum, and midbrain) using the CCFv3. We further divided the cortex into ipsilateral and contralateral cortexes using the midline (SupFig. 6A). We drew the brain boundaries based on the parents of the 11th level of CCFv3. In the cortical flatmap, we represented the area boundaries by boundaries between 45-55% cortical depth, unless specified otherwise.

Simulated retrograde tracing

We manually identified three injection centers for simulated retrograde tracing on the flatmap. We defined a range of 300 µm around the center as the injection/patch region. To be considered positive for retrograde tracing, a cell must meet the following criteria: (1) be from the IT class, (2) have ≥ 10 rolonies within the patch region, and (3) have a soma location. To identify cells that specifically project to a contralateral patch, the patch/CtxC rolony count ratio must be ≥ 75%.

Single-cell tracing reconstruction

We created reconstructions by connecting the registered xyz-coordinates of rolonies and soma from the same barcoded cells. We first connected data points (including rolonies and soma) to their closest neighbor to form clusters, then connected each cluster to the nearest cluster via the closest data points until all clusters were connected. We set the maximum distance for connecting two data points at 1000 µm. We only included cells with soma location in the reconstruction, and dilated the somata for visualization purposes.

We computed the transparent outline of the brain using the CCFv3 annotation map. In coronal view images, we excluded stacks anterior or posterior to the current dataset (e.g., olfactory bulb and cerebellum) for visualization purposes.

Grouping major cell types

We divided barcoded neurons into CF and IT cells based on projections to the ipsilateral thalamus (Thal) and superior colliculus (SupCol). CF cells had projections to either the Thal or SupCol (Fig. 3A) with minimum rolony counts as described above; all other cells were assigned as IT cells.

We used a two-step approach to classify Thal+ cells as either ET or CT cells. In the first step, we grouped cells based on the presence of rolonies in the striatal-thalamic fiber and thalamic reticular nucleus. The AUD axons in this region could be divided into two bundles, an upper and a lower bundle in the coronal view and corresponding to axons from CT and ET cells ((Mitrofanis and Baker 1993); also see Allen mouse connectivity: C57BL/5J 158314278 image 72, 115958825 image 68; Syt6-Cre 124060405 image 71; Ntsr1-Cre 266963362 image 69; Rbp4-Cre 182090318 image 72, 606100558 image 77;(Oh et al. 2014)). We manually defined a region of interest in this region using registered xyz-coordinates (x: 8250-9000 µm; y: 3500-5000 µm; z: 6750-7500 µm; red region in SupFig. 7A) and divided the rolonies within it into two groups based on their y-axis location: the top half were classified as CT cells and the

bottom half were classified as ET cells. We then assigned each individual rolony to the most frequent group of its nearest 10 neighbors and followed this by assigning each barcoded cell to the most frequent group. This process was repeated until convergence or after 100 iterations. The cell type for each barcoded cell was represented by the most frequent group of rolonies within the region of interest. The results of this initial grouping are shown in SupFig. 7B (930 of 1134 Thal+ cells were assigned to the CT/ET group; CT: 581; ET: 349). In the second step, we assigned all thalamic rolonies to the most frequent group of their nearest 10 neighbors, and the cell type for each barcoded cell was represented by the most frequent group of thalamic rolonies. 58 of the 930 cells were assigned to a different group in this step compared to the first step. The final results of the ET/CT grouping are shown in SupFig. 7C (CT: 713; ET: 421). Overall, this approach was able to classify the cell types for 91.2% (321/352) of Thal+ cells that project to the thalamus and the superior colliculus as ET cells and 8.8% (31) as CT cells, indicating that axonal BARseq can effectively identify cell types. We also observed that CT cells have rolonies in the striatum due to their axons traveling through the region to reach the thalamus. Visual examination showed that the majority of cells with rolonies in the striatum belong to the ET group (SupFig. 7D).

IT cells were divided into two subtypes: ITi and ITc. IT cells with ≥ 5 rolonies in the contralateral cortex were assigned to the ITc group, while the rest were assigned to the ITi group.

Visualization and quantification of soma laminar distribution

To visualize the distribution of somata across groups, we plotted them using flatmap coordinates: x-axis, ML; y-axis, depth %. The plotting sequences were randomly shuffled across groups. Note that the ratio of the x-axis to the y-axis is not equal for visualization purposes.

To quantify the proportion of somata from different groups at different depths, we binned the soma depths into 5% bins and calculated the percentages for each bin: (group count) / (all group counts) × 100%.

Visualization and quantification of rolony laminar distribution

To visualize the rolonies in the CCFv3, we plotted them in registered xyz-coordinates using a coronal view. The brain outline was shown in gray.

To visualize the laminar distribution in the cortex, we presented the data as heatmaps, unless otherwise specified. The frequency of rolony depth was calculated for each cell or bin within a region, using bin sizes of 1% or 5% along the depth. The darkness of the grid represented the relative probability.

Cortical rolony analysis

In our experiments, we excluded cortical rolonies that were deeper than 95% due to their proximity to the fiber tract and potential registration errors. Most infected somata were localized to the middle layers. We excluded rolonies near the somata based on their xyz-coordinates, which may create uniform exclusion across cortical depth. Thus, for the cortical analysis (Fig. 4-5, SupFig. 8-9), we excluded rolonies that were located < 95 percentile of the injection center on the ML-AP plate (indicated by the black disk; the

injection center is the median soma location). As a control for this exclusion, we also excluded rolonies in the same region on the other hemisphere (referred to as the LatC local-exclusion control).

Although the carrier protein VAMP2nl was based on VAMP2, which localizes presynaptically, we found barcodes in dendrites as well. We excluded rolonies close to the somata using two filters as described above, so we believe the impact of dendritic barcodes on the cortical analysis is minimal.

The separation of Met and Lat targets was described in Fig. 4B, and we required a minimum of 5 rolonies per target for LatI/MedI/MedC/LatC-projection unless stated otherwise.

Description and comparison of cortical laminar patterns

To compare the projections between cortical regions, we used the frequency and cumulative distribution function (CDF, bin size: 1% depth, SupFig. 8A-D). When comparing projections between the Med and Lat areas, we combined the MedI and C as Med due to their similar projection patterns.

To compare the laminar patterns between two regions, we considered several factors. The cortex consists of layers that can be seen as categorical differences along the depth, so comparing the distance between two continuous distributions may not be the most suitable method. For example, the distances between layers 2 and 3, and between layers 1 and 2 may be similar, but they may have different biological implications. However, biological and experimental variability can cause the boundaries between layers to be blurry, resulting in projections that follow a continuous distribution within a certain range. Additionally, there are other factors that influence the laminar distribution of rolonies, such as the need for upper layer projections to pass through lower layers on the contralateral side. Given these considerations, we used the Kolmogorov-Smirnov test (KS test) based on the CDF to determine whether the two laminar patterns are different.

The p-value represented the probability that the two laminar patterns come from the same distribution. We calculated the p-value using permutation tests, as the possibility of shuffle KS distances $\geq$ data KS distance. During shuffling, the rolony depths from two regions were randomly shuffled across regions for each cell, and the KS distance was calculated for each shuffle (total 1000 shuffles per cell). If the p-value ≤ 0.05, we classified the two laminar patterns as different, otherwise we considered them to be the same. The distribution of p-values for the four groups is shown in SupFig. 8I.

The percentage of cells with different projections was calculated for each group as: (cell counts with p-value ≤ 0.05) / (cell counts) $\times$ 100%. We computed the confidence intervals as described below.

Lat-projecting IT cell grouping

We selected Lat-projecting IT cells with a minimum of 5 rolonies in LatI/C and divided them into three groups based on soma depth (Fig. 5A). However, not all the barcoded cells have a known soma location (as described above), so we only included Lat-projecting cells with a known soma location in the analysis. As a control, we grouped barcoded cells without soma information into two groups based on projection depth, as shown in SupFig. 9I-M.-

Focal projection distance

For each barcoded cell, we calculated the all-to-all distances of the LatI/C rolonies on the ML-AP plate. For each rolony, we selected the shortest 33% of distances and used the mean of these distances to represent the 'focal projection distance' of the rolony. We then used the mean of the rolony focal projection distances to represent the distance for each cell.

We selected the shortest 33% distance as a measurement based on the following considerations: We wanted to capture the small range and tightness of the projection, so we calculated the focal projection distance using a subset of the nearest rolonies. We observed that cells can have more than one focal projection per target (e.g. the LatC projection from the orange cell in Fig. 5D), so we set a threshold (i.e. 33%) to exclude rolonies from other clusters. Based on our observations, most cells did not have many focal projections, and it was difficult to distinguish a cell with many focal projections from one with a sparse projection. Therefore, we assumed that cells have a maximum of three focal projections and calculated the distance from the closest 33% of rolonies.

To estimate the effect of rolony number on focal projection distance per cell, we randomly subsampled different numbers of rolonies from each barcoded cell and calculated the focal projection distance for each subsample. The ground truth distance was calculated using all rolonies from the same cell. We defined the sampling error as:

$$\mathrm{abs}(\mathrm{distance}_{\mathrm{sample}} - \mathrm{distance}_{\mathrm{groundtruth}})/\mathrm{distance}_{\mathrm{groundtruth}} \times 100\%$$

We performed 100 random samplings per cell, and the median error represents the error per cell. The results are shown in SupFig. 9F. Thus, we only used cells with ≥ 55 Lat rolonies for the following analysis (Fig. 5H, SupFig. 9G-H, L-M).

Software and statistical analysis

We used MATLAB, ImageJ/FIJI, and MIJ(Sage et al. 2012) for data processing and visualization. We used MATLAB and GraphPad Prism 9 for statistical analyses, as indicated in the text. **, p-value < 0.01, ***, p-value < 0.001, ****, p-value < 0.0001, unless stated otherwise.

To estimate confidence intervals (2.5 and 97.5%), bootstrap was performed for each group (2000 iterations). * indicated confidence intervals of two groups have no overlap.

Data availability

Raw images will be deposited to a public database. Codes for image processing and analysis are available in GitHub.

Acknowledgements

The authors would like to acknowledge Alex Vaughan and Justus M. Kebschull for VAMP2nl library, Kathleen Lucere, Eugene Fong and John M. Bolger for virus injection and animal processing, Yu-Chi Sun, Wiktor Wadolowski and Barry Burbach for technical support, Longwen Huang, Daniel Fürth and Anand Suresh for useful discussion, chatGPT for manuscript editing. This work was supported by the National Institutes of Health (RF1MH123403; U19NS123716) to A.M.Z; IARPA MICrONS (D16PC0008 to A.M.Z.), and Robert Lourie award (to A.M.Z.). In conducting research using animals, the investigator adheres to the laws of the United States and regulations of the Department of Agriculture.

Author Contributions

L.Y., X.C., and A.M.Z. conceived the study. L.Y. performed experiments and data analysis. L.Y. and A.M.Z. wrote the paper.

Declaration of Interests

A.M.Z. is a founder and equity owner of Cajal Neuroscience and a member of its scientific advisory board.

Supplementary tables

SupTable 1, List of brain sections and imaged areas

SupTable 2, List of reagents and optics

Supplementary video

SupVideo, Single-cell reconstruction from 100 barcoded neurons

References

Chen, Xiaoyin, Yu-Chi Sun, George M. Church, Je Hyuk Lee, and Anthony M. Zador. 2018. "Efficient in Situ Barcode Sequencing Using Padlock Probe-Based BaristaSeq." *Nucleic Acids Research* 46 (4): e22–e22.

Chen, Xiaoyin, Yu-Chi Sun, Huiqing Zhan, Justus M. Kebschull, Stephan Fischer, Katherine Matho, Z. Josh Huang, Jesse Gillis, and Anthony M. Zador. 2019. "High-Throughput Mapping of Long-Range Neuronal Projection Using In Situ Sequencing." *Cell* 179 (3): 772–86.e19.

Chen, Yushu, Xiaoyin Chen, Batuhan Baserdem, Huiqing Zhan, Yan Li, Martin B. Davis, Justus M. Kebschull, Anthony M. Zador, Alexei A. Koulakov, and Dinu F. Albeanu. 2022. "High-Throughput Sequencing of Single Neuron Projections Reveals Spatial Organization in the Olfactory Cortex." *Cell* 185 (22): 4117–34.e28.

Coleman, Jason E., Marc Nahmani, Jeffrey P. Gavornik, Robert Haslinger, Arnold J. Heynen, Alev Erisir, and Mark F. Bear. 2010. "Rapid Structural Remodeling of Thalamocortical Synapses Parallels Experience-Dependent Functional Plasticity in Mouse Primary Visual Cortex." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 30 (29): 9670–82.

Gao, Le, Sang Liu, Lingfeng Gou, Yachuang Hu, Yanhe Liu, Li Deng, Danyi Ma, et al. 2022. "Single-Neuron Projectome of Mouse Prefrontal Cortex." *Nature Neuroscience* 25 (4): 515–29.

Gergues, Mark M., Kasey J. Han, Hye Sun Choi, Brandon Brown, Kelsey J. Clausing, Victoria S. Turner, Ilia D. Vainchtein, Anna V. Molofsky, and Mazen A. Kheirbek. 2020. "Circuit and Molecular Architecture of a Ventral Hippocampal Network." *Nature Neuroscience* 23 (11): 1444–52.

Hackett, Troy A., Amanda R. Clause, Toru Takahata, Nicholas J. Hackett, and Daniel B. Polley. 2016. "Differential Maturation of Vesicular Glutamate and GABA Transporter Expression in the Mouse Auditory Forebrain during the First Weeks of Hearing." *Brain Structure & Function* 221 (5): 2619–73.

Han, Yunyun, Justus M. Kebschull, Robert A. A. Campbell, Devon Cowan, Fabia Imhof, Anthony M. Zador, and Thomas D. Mrsic-Flogel. 2018. "The Logic of Single-Cell Projections from Visual Cortex." *Nature* 556 (7699): 51–56.

Harris, Julie A., Stefan Mihalas, Karla E. Hirokawa, Jennifer D. Whitesell, Hannah Choi, Amy Bernard, Phillip Bohn, et al. 2019. "Hierarchical Organization of Cortical and Thalamic Connectivity." *Nature* 575 (7781): 195–202.

Harris, Kenneth D., and Gordon M. G. Shepherd. 2015. "The Neocortical Circuit: Themes and Variations." *Nature Neuroscience* 18 (2): 170–81.

Hausmann, Frances S., John M. Barrett, Megan E. Martin, Huiqing Zhan, and Gordon M. G. Shepherd. 2022. "Axonal Barcode Analysis of Pyramidal Tract Projections from Mouse Forelimb M1 and M2." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 42 (41): 7733–43.

Herculano-Houzel, Suzana, Bruno Mota, and Roberto Lent. 2006. "Cellular Scaling Rules for Rodent Brains." *Proceedings of the National Academy of Sciences of the United States of America* 103 (32): 12138–43.

Hirai, Yasuharu, Mieko Morishima, Fuyuki Karube, and Yasuo Kawaguchi. 2012. "Specialized Cortical Subnetworks Differentially Connect Frontal Cortex to Parahippocampal Areas." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 32 (5): 1898–1913.

Huang, Longwen, Justus M. Kebschull, Daniel Fürth, Simon Musall, Matthew T. Kaufman, Anne K. Churchland, and Anthony M. Zador. 2020. "BRICseq Bridges Brain-Wide Interregional Connectivity to Neural Activity and Gene Expression in Single Animals." *Cell* 183 (7): 2040.

Joffroy, Bastian, Yavuz O. Uca, Domen Prešern, Jonathan P. K. Doye, and Thorsten L. Schmidt. 2018. "Rolling Circle Amplification Shows a Sinusoidal Template Length-Dependent Amplification Bias." *Nucleic Acids Research* 46 (2): 538–45.

Kebschull, Justus M., Pedro Garcia da Silva, Ashlan P. Reid, Ian D. Peikon, Dinu F. Albeanu, and Anthony M. Zador. 2016. "High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA." *Neuron* 91 (5): 975–87.

Klingler, Esther, Ugo Tomasello, Julien Prados, Justus M. Kebschull, Alessandro Contestabile, Gregorio L. Galiñanes, Sabine Fièvre, et al. 2021. "Temporal Controls over Inter-Areal Cortical Projection Neuron Fate Diversity." *Nature* 599 (7885): 453–57.

Livet, Jean, Tamily A. Weissman, Hyuno Kang, Ryan W. Draft, Ju Lu, Robyn A. Bennis, Joshua R. Sanes, and Jeff W. Lichtman. 2007. "Transgenic Strategies for Combinatorial Expression of Fluorescent Proteins in the Nervous System." *Nature* 450 (7166): 56–62.

Llano, Daniel A., and S. Murray Sherman. 2008. "Evidence for Nonreciprocal Organization of the Mouse Auditory Thalamocortical-Corticothalamic Projection Systems." *The Journal of Comparative Neurology* 507 (2): 1209–27.

Lyamzin, Dmitry, and Andrea Benucci. 2019. "The Mouse Posterior Parietal Cortex: Anatomy and Functions." *Neuroscience Research* 140 (March): 14–22.

Mathis, Victor P., Maya Williams, Clementine Fillinger, and Paul J. Kenny. 2021. "Networks of Habenula-Projecting Cortical Neurons Regulate Cocaine Seeking." *Science Advances* 7 (45): eabj2225.

Mitrofanis, J., and G. E. Baker. 1993. "Development of the Thalamic Reticular and Perireticular Nuclei in Rats and Their Relationship to the Course of Growing Corticofugal and Corticopetal Axons." *The Journal of Comparative Neurology* 338 (4): 575–87.

Muñoz-Castañeda, Rodrigo, Brian Zingg, Katherine S. Matho, Xiaoyin Chen, Quanxin Wang, Nicholas N. Foster, Anan Li, et al. 2021. "Cellular Anatomy of the Mouse Primary Motor Cortex." *Nature* 598 (7879): 159–66.

Nahmani, Marc, and Alev Erisir. 2005. "VGluT2 Immunochemistry Identifies Thalamocortical Terminals in Layer 4 of Adult and Developing Visual Cortex." *The Journal of Comparative Neurology* 484 (4): 458–73.

Oh, Seung Wook, Julie A. Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, et al. 2014. "A Mesoscale Connectome of the Mouse Brain." *Nature* 508 (7495): 207–14.

Oswald, Manfred J., Malinda L. S. Tantirigama, Ivo Sonntag, Stephanie M. Hughes, and Ruth M. Empson. 2013. "Diversity of Layer 5 Projection Neurons in the Mouse Motor Cortex." *Frontiers in Cellular Neuroscience* 7 (October): 174.

Peng, Hanchuan, Peng Xie, Lijuan Liu, Xiuli Kuang, Yimin Wang, Lei Qu, Hui Gong, et al. 2021. "Morphological Diversity of Single Neurons in Molecularly Defined Cell Types." *Nature* 598 (7879): 174–81.

Sage, Daniel, Dimiter Prodanov, Jean-Yves Tinevez, and Johannes Schindelin. 2012. "MIJ: Making Interoperability Between ImageJ and Matlab Possible." bigwww.epfl.ch. 2012. http://bigwww.epfl.ch/sage/soft/mij/Poster-MIJ-Daniel-Sage.pdf.

Sato, Haruka, Jun Hatakeyama, Takuji Iwasato, Kimi Araki, Nobuhiko Yamamoto, and Kenji Shimamura. 2022. "Thalamocortical Axons Control the Cytoarchitecture of Neocortical Layers by Area-Specific Supply of VGF." *eLife* 11 (March): e67549.

Shahar, A., S. Lustig, Y. Akov, Y. David, P. Schneider, and R. Levin. 1990. "Different Pathogenicity of Encephalitic Togaviruses in Organotypic Cultures of Spinal Cord Slices." *Journal of Neuroscience Research* 25 (3): 345–52.

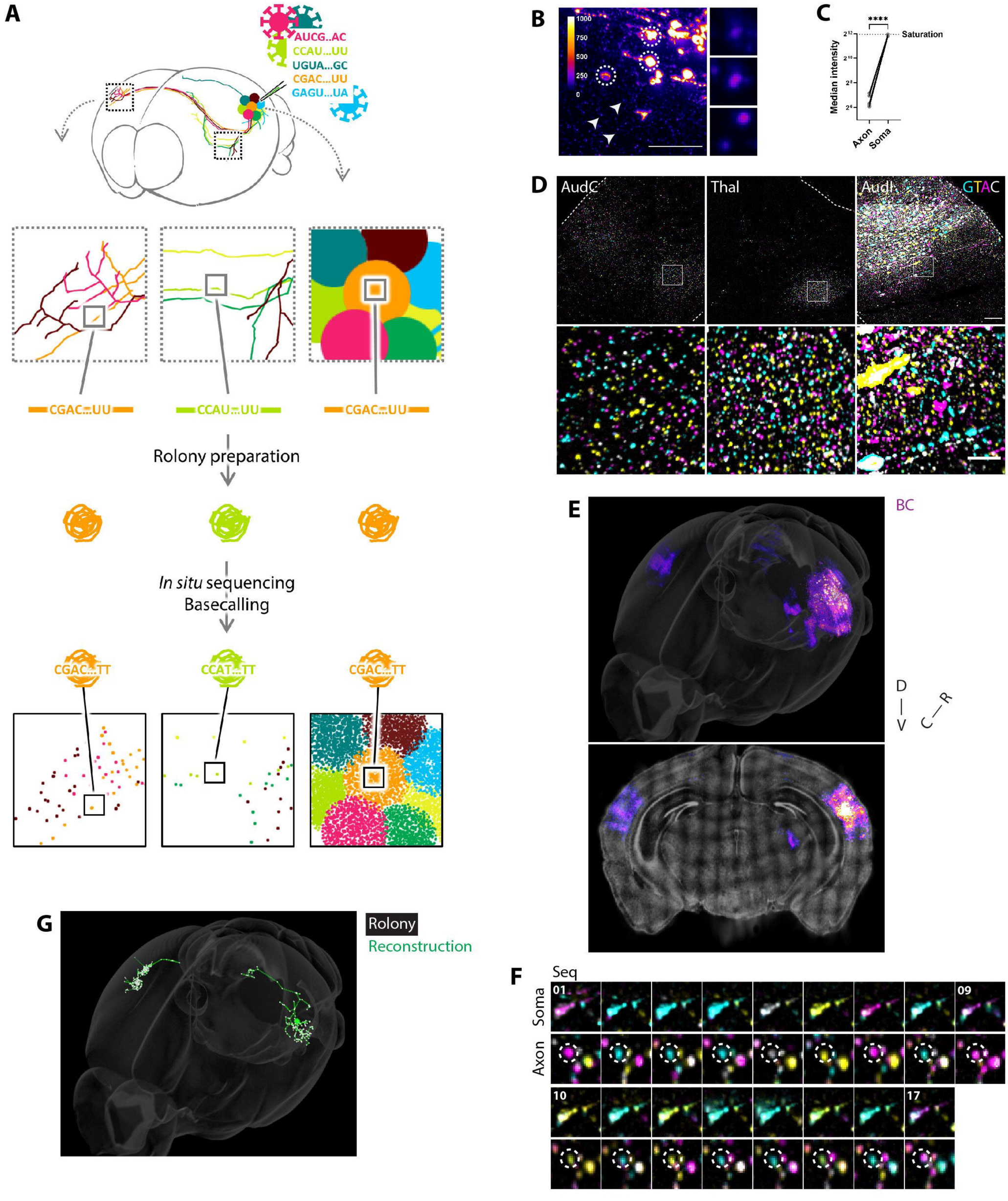
Sun, Yu-Chi, Xiaoyin Chen, Stephan Fischer, Shaina Lu, Huiqing Zhan, Jesse Gillis, and Anthony M. Zador. 2021. "Integrating Barcoded Neuroanatomy with Spatial Transcriptional Profiling Enables Identification of Gene Correlates of Projections." *Nature Neuroscience* 24 (6): 873–85.

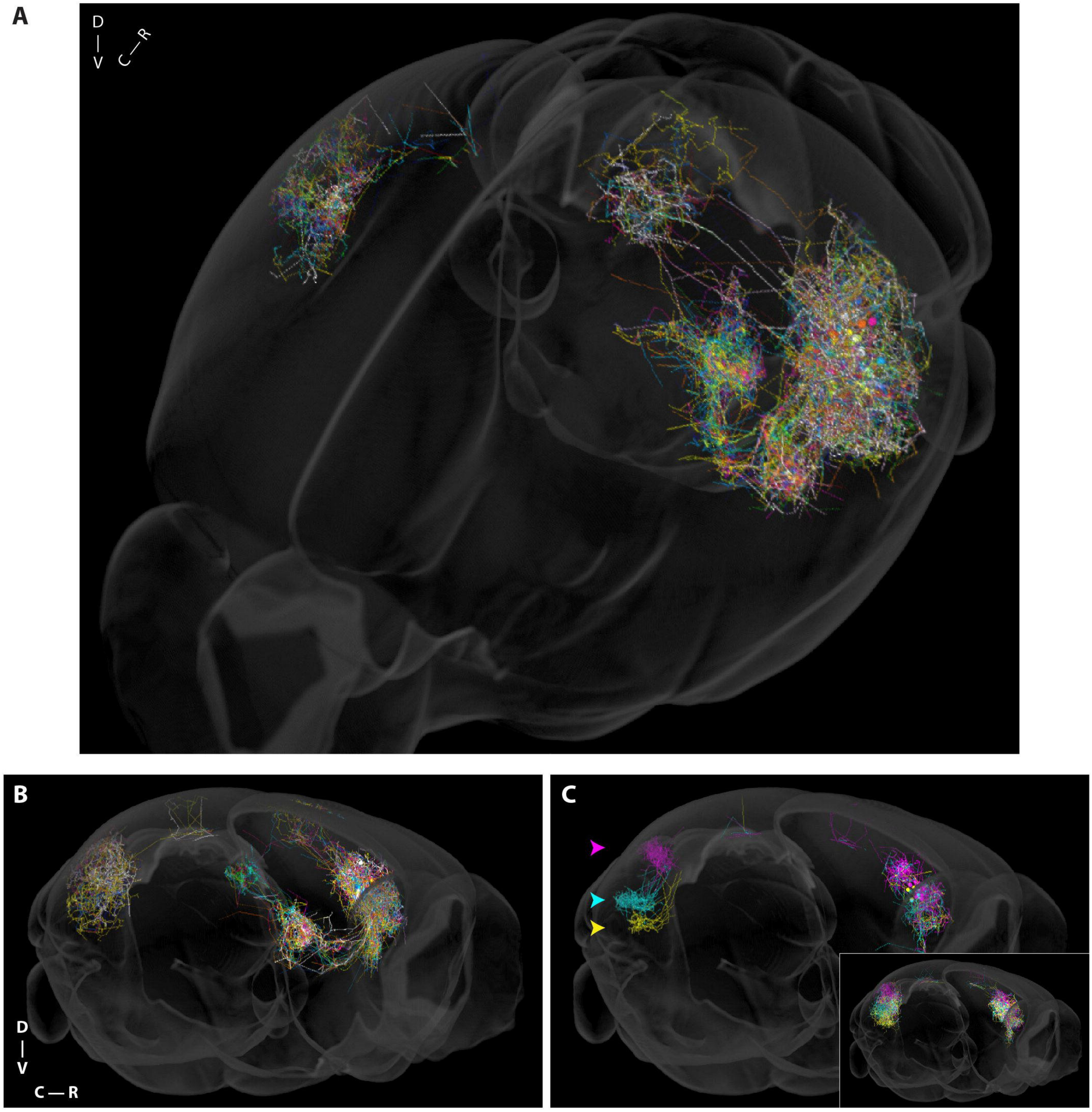
Taniguchi, Hiroki, Miao He, Priscilla Wu, Sangyong Kim, Raehum Paik, Ken Sugino, Duda Kvitsiani, et al. 2011. "A Resource of Cre Driver Lines for Genetic Targeting of GABAergic Neurons in Cerebral Cortex." *Neuron* 71 (6): 995–1013.

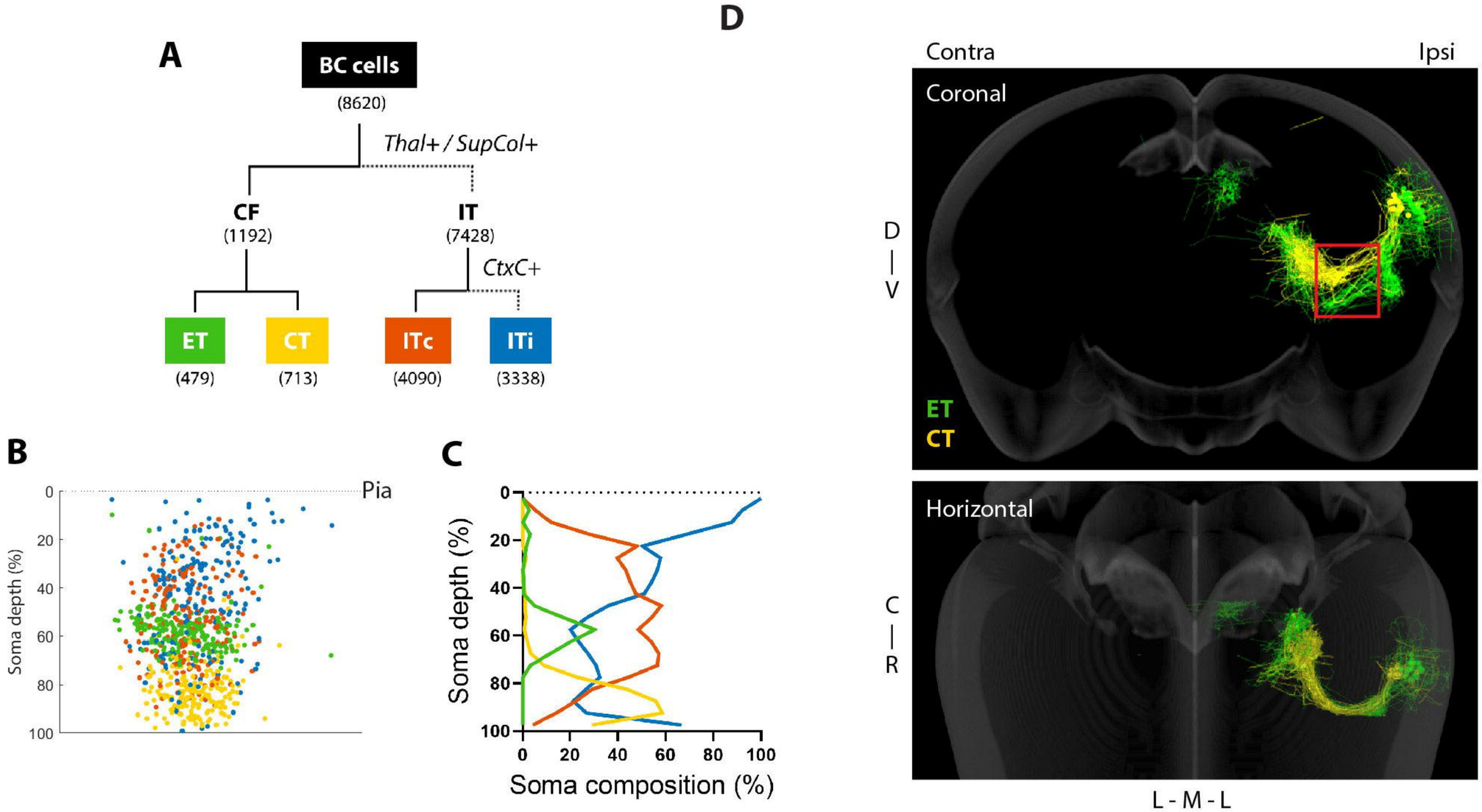
Tasic, Bosiljka, Vilas Menon, Thuc Nghi Nguyen, Tae Kyung Kim, Tim Jarsky, Zizhen Yao, Boaz Levi, et al. 2016. "Adult Mouse Cortical Cell Taxonomy Revealed by Single Cell Transcriptomics." *Nature Neuroscience* 19 (2): 335–46.

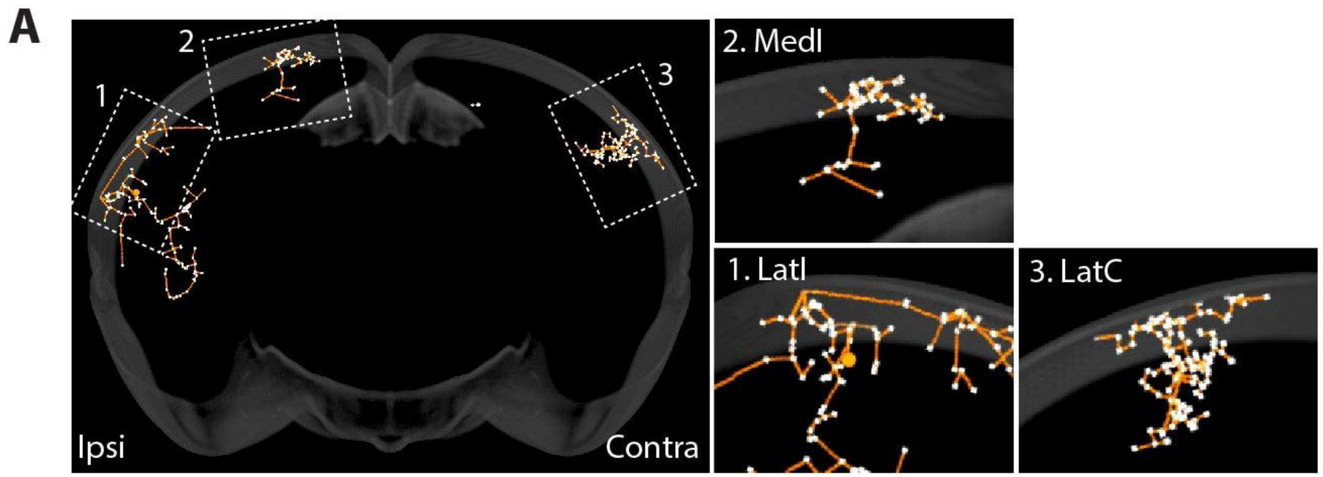
Tasic, Bosiljka, Zizhen Yao, Lucas T. Graybuck, Kimberly A. Smith, Thuc Nghi Nguyen, Darren Bertagnolli, Jeff Goldy, et al. 2018. "Shared and Distinct Transcriptomic Cell Types across

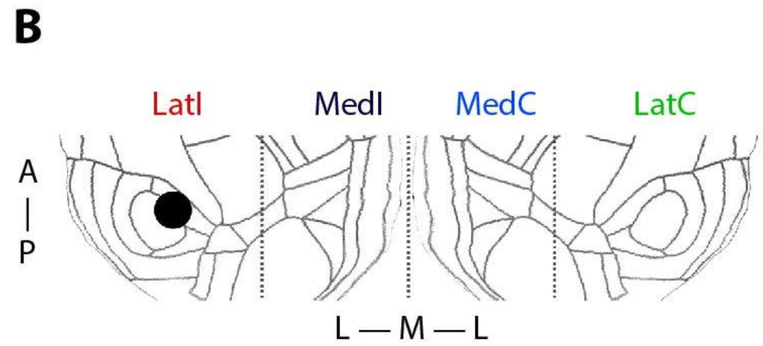
Neocortical Areas." *Nature* 563 (7729): 72–78.
Tsoi, Sau Yee, Merve Öncül, Ella Svahn, Mark Robertson, Zuzanna Bogdanowicz, Christina McClure, and Gülşen Sürmeli. 2022. "Telencephalic Outputs from the Medial Entorhinal Cortex Are Copied Directly to the Hippocampus." *eLife* 11 (February): e73162.
Wang, Quanxin, Song-Lin Ding, Yang Li, Josh Royall, David Feng, Phil Lesnar, Nile Graddis, et al. 2020. "The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas." *Cell* 181 (4): 936–53.e20.
Webb, John M., Mingyang Ma, Chen Yin, Louis J. Ptáček, and Ying-Hui Fu. 2022. "An Excitatory Peri-Tegmental Reticular Nucleus Circuit for Wake Maintenance." *Proceedings of the National Academy of Sciences* 119 (34): e2203266119.
Winnubst, Johan, Erhan Bas, Tiago A. Ferreira, Zhuhao Wu, Michael N. Economo, Patrick Edson, Ben J. Arthur, et al. 2019. "Reconstruction of 1,000 Projection Neurons Reveals New Cell Types and Organization of Long-Range Connectivity in the Mouse Brain." *Cell* 179 (1): 268–81.e13.
Zeisler, Zachary R., Liza London, William G. Janssen, J. Megan Fredericks, Catherine Elorette, Atsushi Fujimoto, Huiqing Zhan, et al. 2023. "High-Throughput Sequencing of Macaque Basolateral Amygdala Projections Reveals Dissociable Connectional Motifs with Frontal Cortex." *bioRxiv*. https://doi.org/10.1101/2023.01.18.524407.

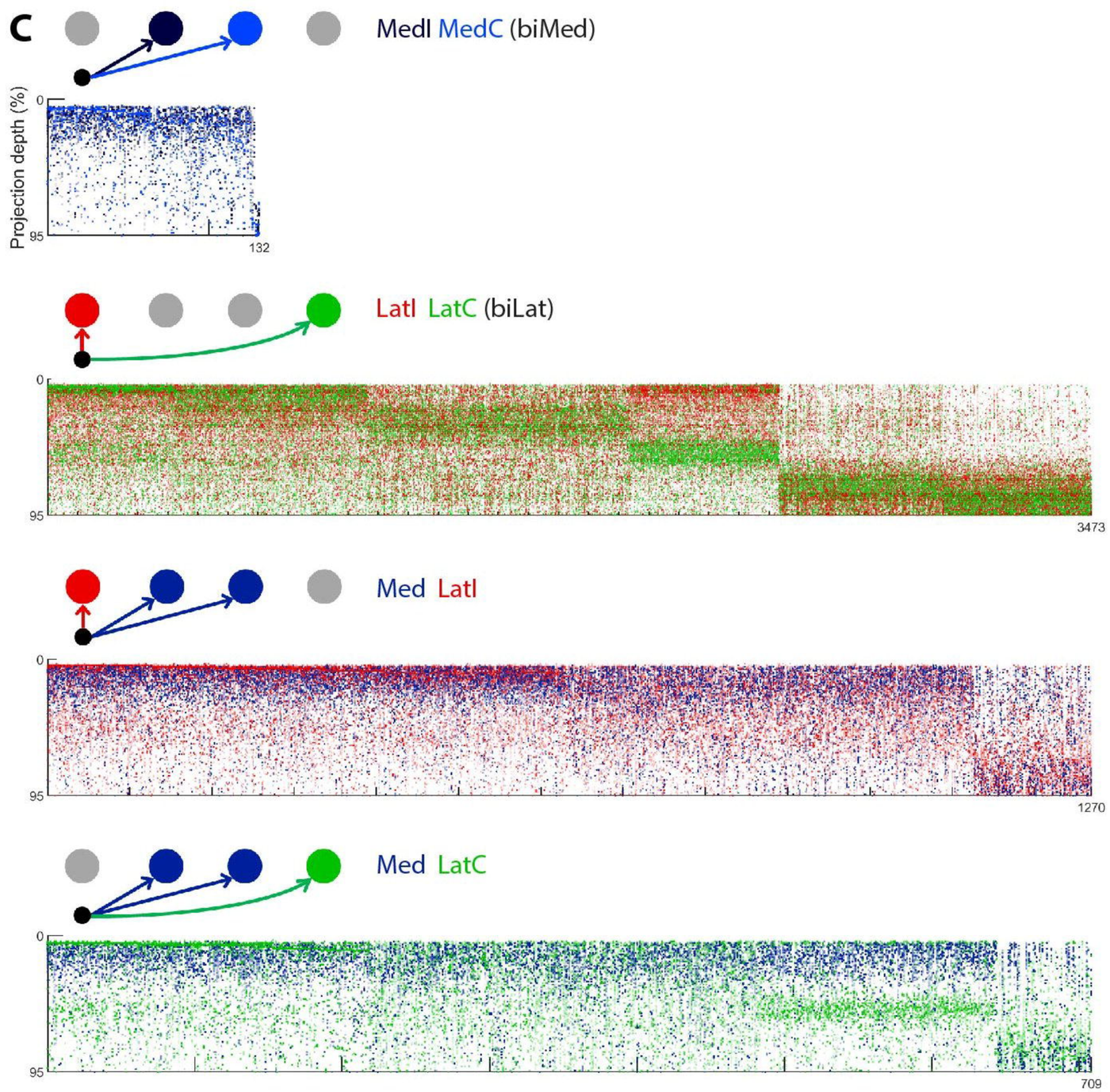

A
AUCG..AC
CCAU...UU
UGUA...GC
CGAC...UU
GAGU...UA
CGAC...UU
CCAU...UU
CGAC...UU
Rolony preparation
In situ sequencing
Basecalling
CGAC...TT
CCAT...TT
CGAC...TT
B
1000
750
500
250
0
C
****
Saturation
Median intensity
Axon
Soma
D
AudC
Thal
AudI
GTAC
E
BC
D
V
C
R
G
Rolony
Reconstruction
F
Seq
Soma
Axon
01
09
10
17

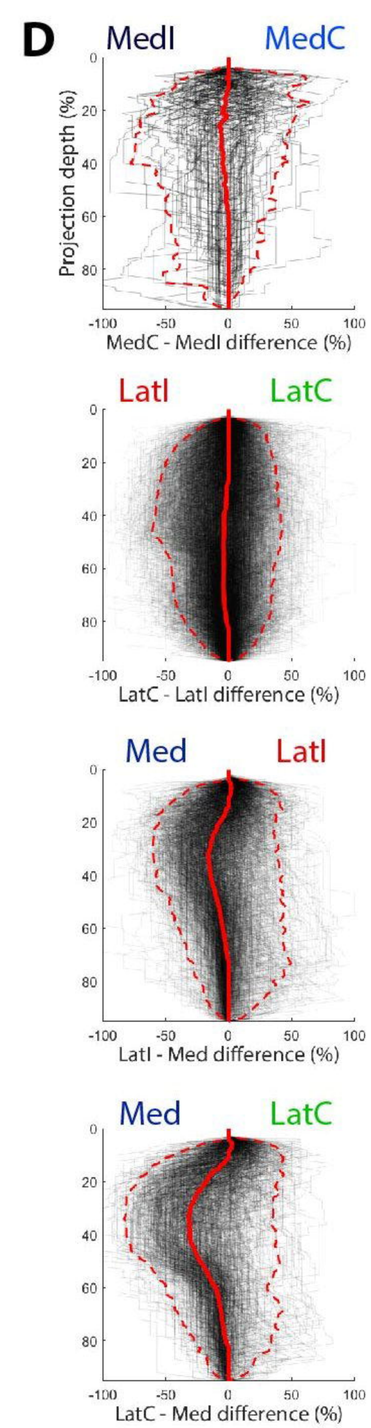

A
D
V
C
R
B
D
V
C — R
C

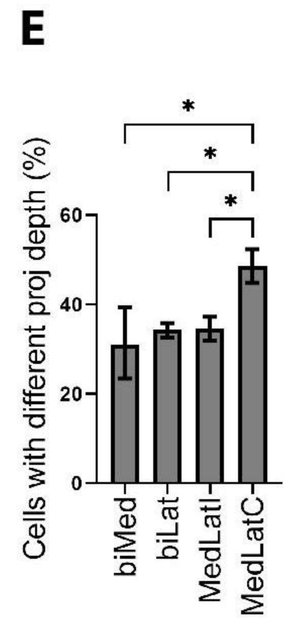

A
BC cells
(8620)
Thal+ / SupCol+
CF
(1192)
IT
(7428)
CtxC+
ET
(479)
CT
(713)
ITc
(4090)
ITi
(3338)
B
Pia
Soma depth (%)
0
20
40
60
80
100
C
Soma depth (%)
Soma composition (%)
0
20
40
60
80
100
D
Contra
Ipsi
Coronal
D
V
ET
CT
Horizontal
C
R
L - M - L

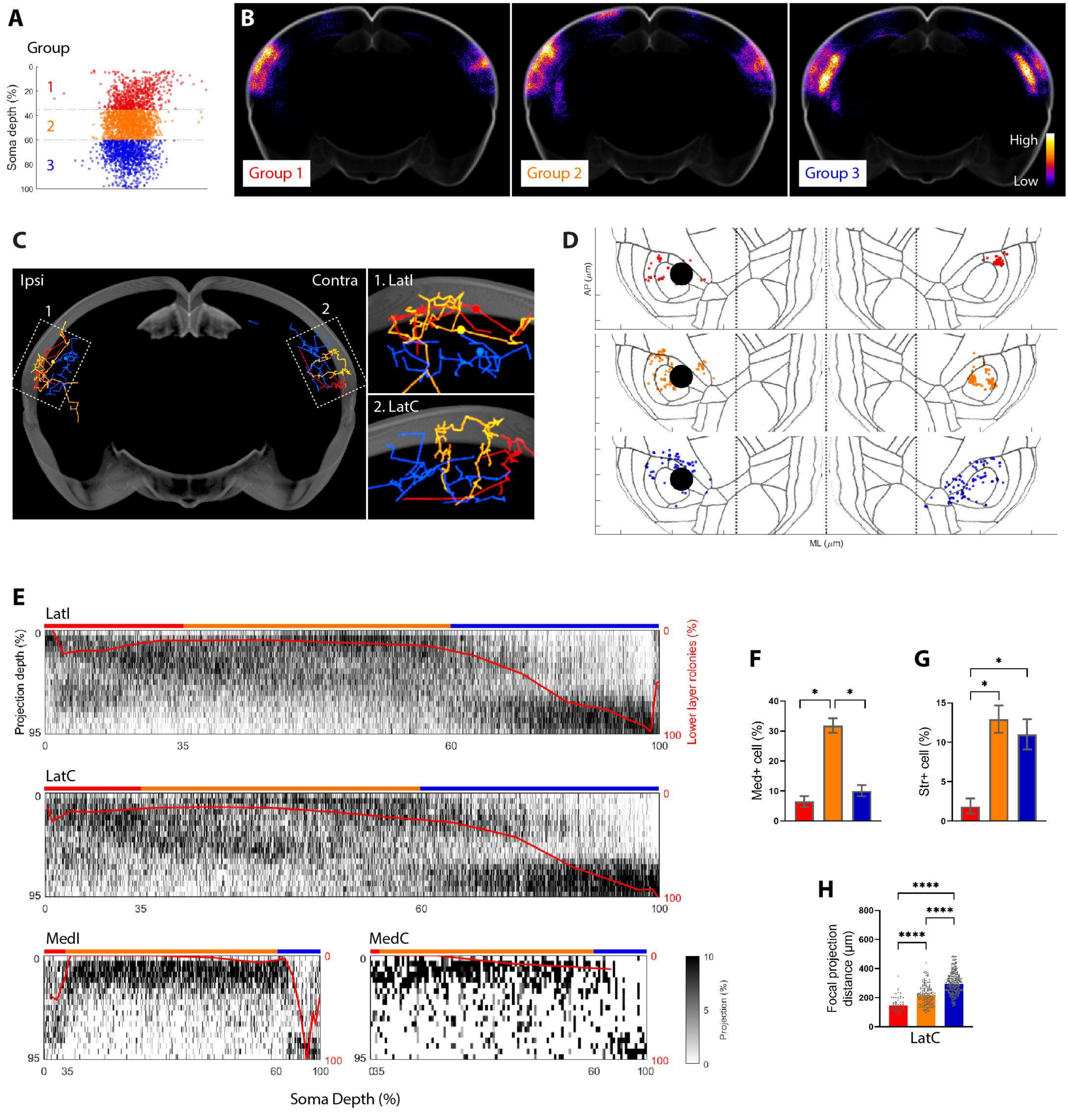

A
1
2
3
Ipsi
Contra
2. MedI
1. LatI
3. LatC

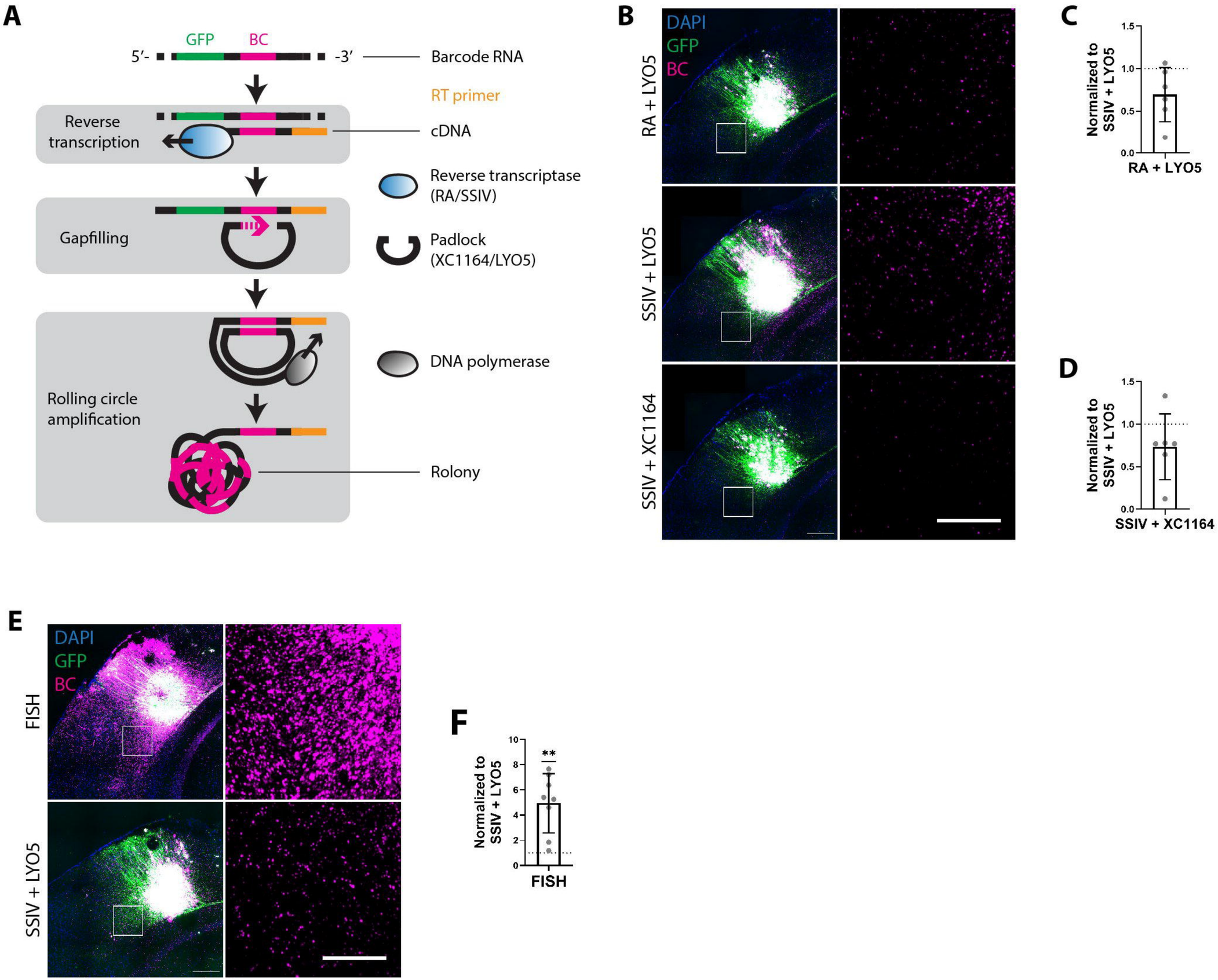

B
LatI
MedI
MedC
LatC
A
P
L — M — L

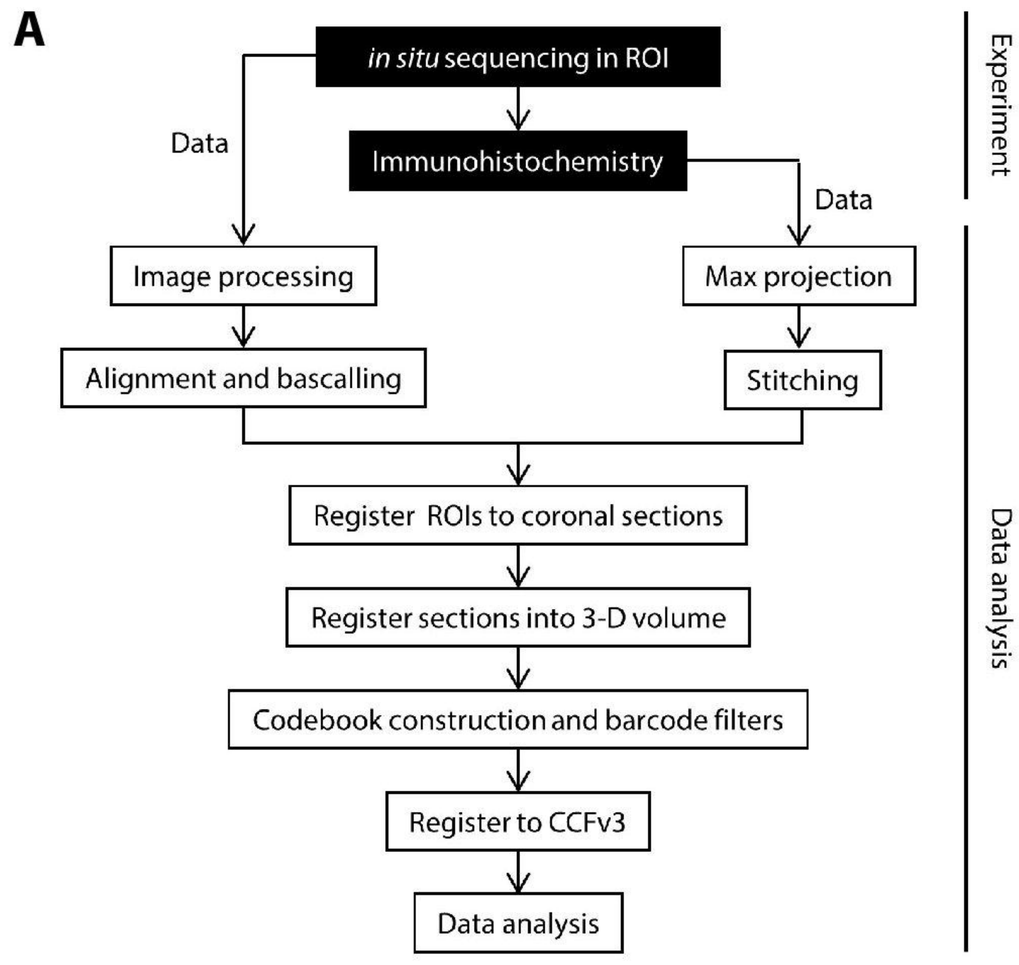

C
MedI MedC (biMed)
Projection depth (%)
0
95
132
LatI LatC (biLat)
3473
Med LatI
1270
Med LatC
709

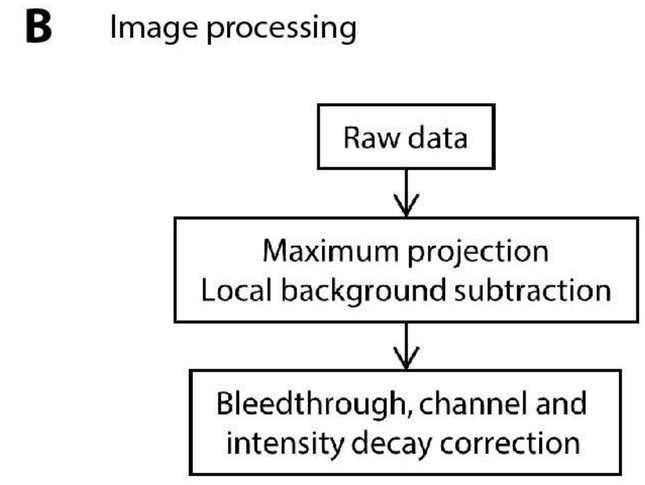

D
MedI
MedC
Projection depth (%)
MedC - MedI difference (%)
LatI
LatC
LatC - LatI difference (%)
Med
LatI
LatI - Med difference (%)
Med
LatC
LatC - Med difference (%)

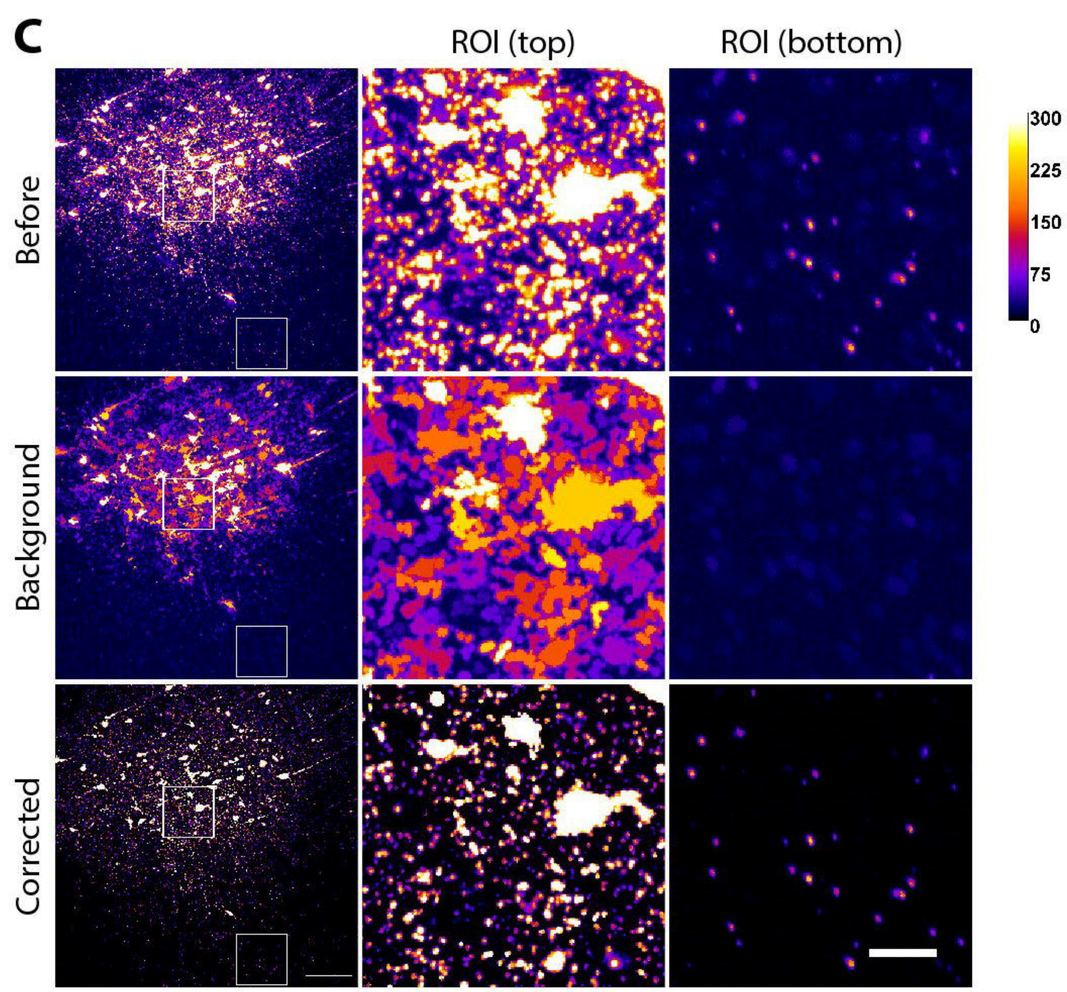

E
Cells with different proj depth (%)
0
20
40
60
*
*
*
biMed
biLat
MedLatI
MedLatC

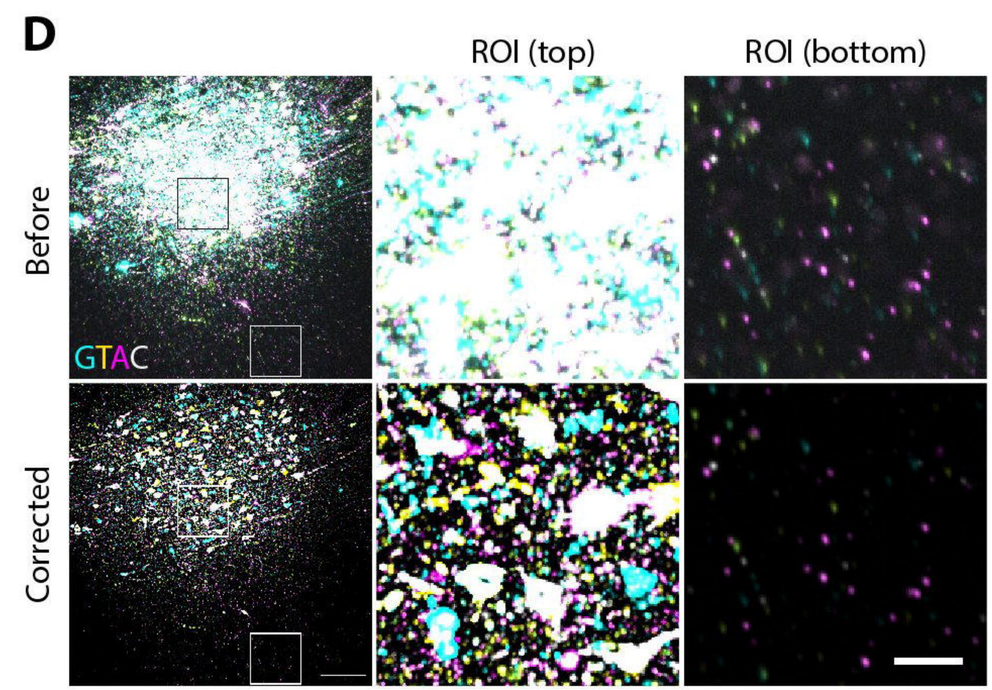

A
Group
Soma depth (%)
1
2
3
B
Group 1
Group 2
Group 3
High
Low
C
Ipsi
Contra
1
2
1. LatI
2. LatC
D
AP (μm)
ML (μm)
E
LatI
LatC
MedI
MedC
Projection depth (%)
Lower layer rolonies (%)
Projection (%)
Soma Depth (%)
F
Med+ cell (%)
G
Str+ cell (%)
H
Focal projection distance (μm)
LatC

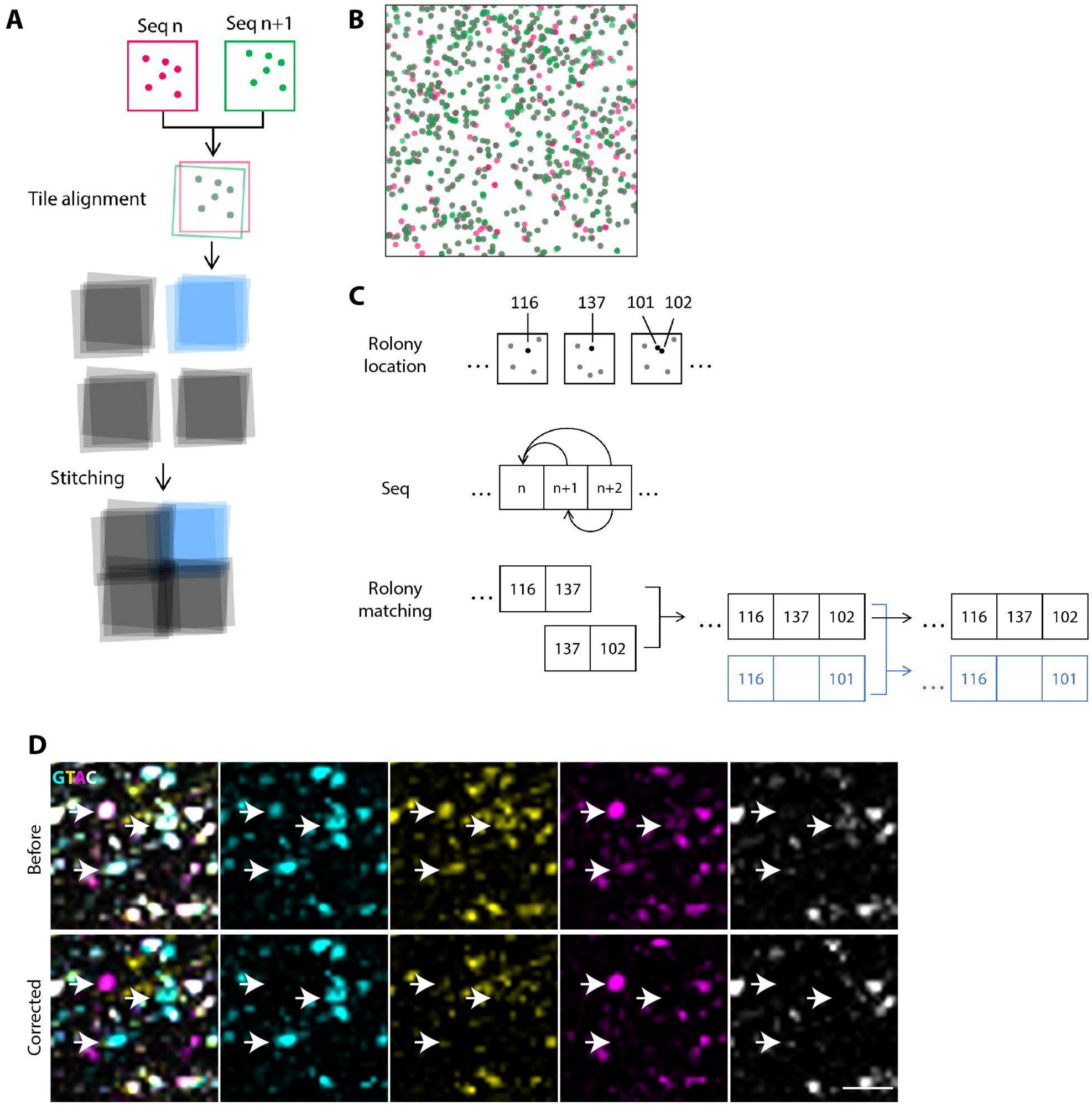

A
GFP
BC
5′-
-3′
Barcode RNA
RT primer
Reverse transcription
cDNA
Reverse transcriptase (RA/SSIV)
Gapfilling
Padlock (XC1164/LYO5)
DNA polymerase
Rolling circle amplification
Rolony
B
DAPI
GFP
BC
RA + LYO5
SSIV + LYO5
SSIV + XC1164
C
Normalized to SSIV + LYO5
RA + LYO5
D
Normalized to SSIV + LYO5
SSIV + XC1164
E
DAPI
GFP
BC
FISH
SSIV + LYO5
F
Normalized to SSIV + LYO5
**
FISH

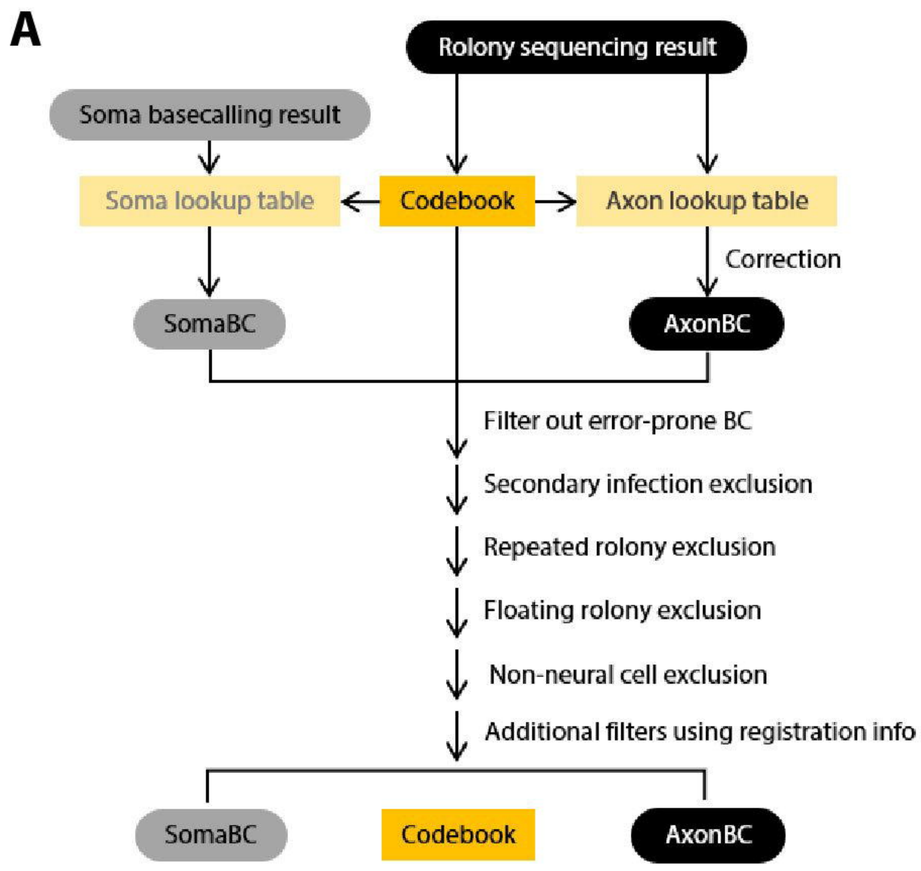

A
in situ sequencing in ROI
Immunohistochemistry
Data
Data
Image processing
Max projection
Alignment and bascalling
Stitching
Register ROIs to coronal sections
Register sections into 3-D volume
Codebook construction and barcode filters
Register to CCFv3
Data analysis
Experiment
Data analysis

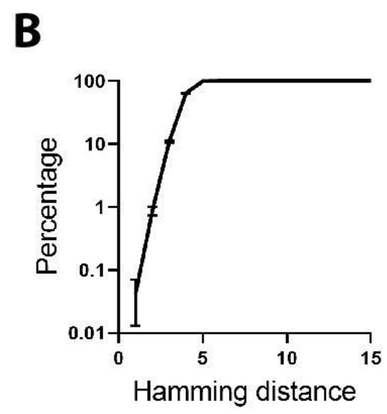

B
Image processing
Raw data
Maximum projection
Local background subtraction
Bleedthrough, channel and
intensity decay correction

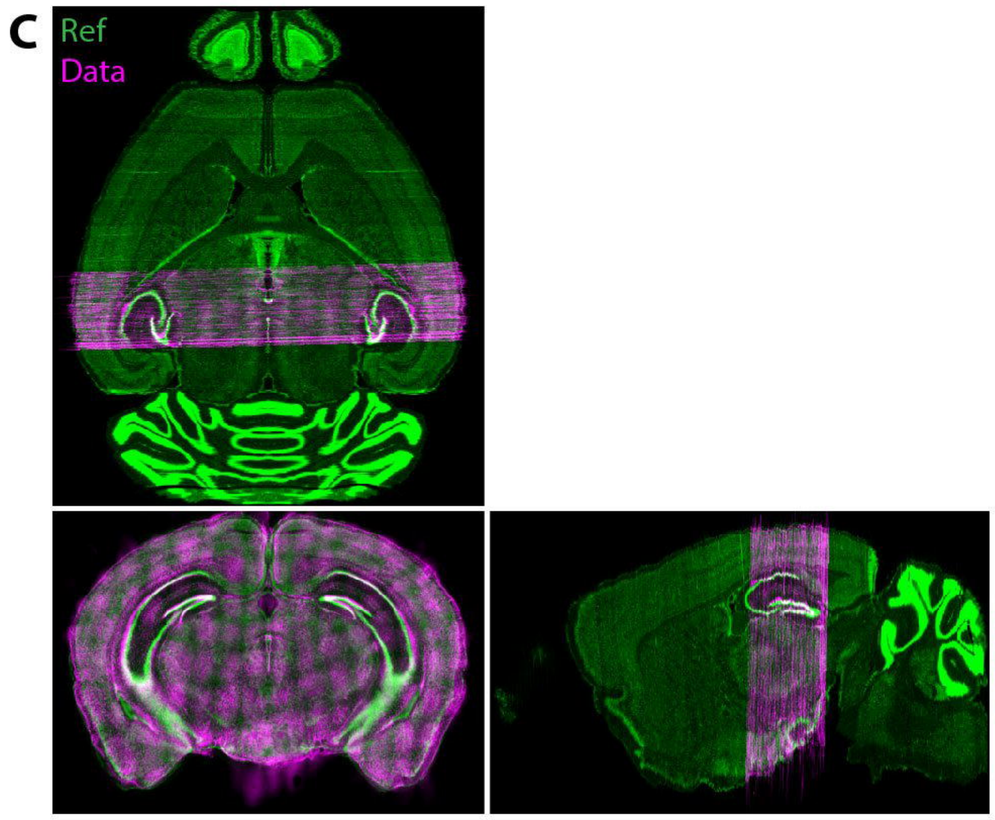

C
ROI (top)
ROI (bottom)
Before
Background
Corrected
300
225
150
75
0

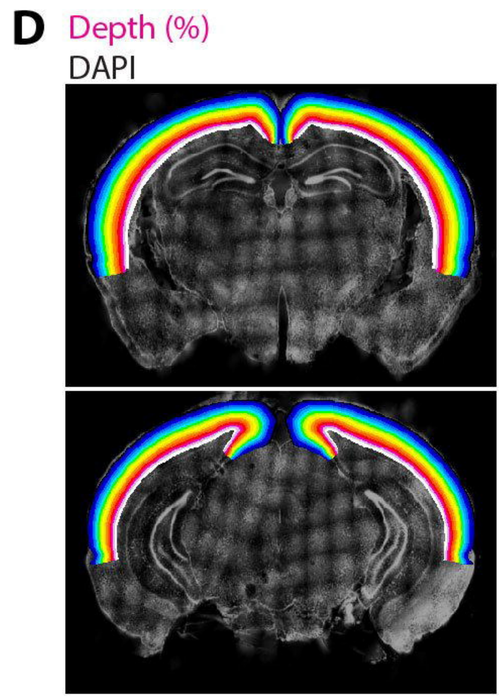

D
ROI (top)
ROI (bottom)
Before
Corrected
GTAC

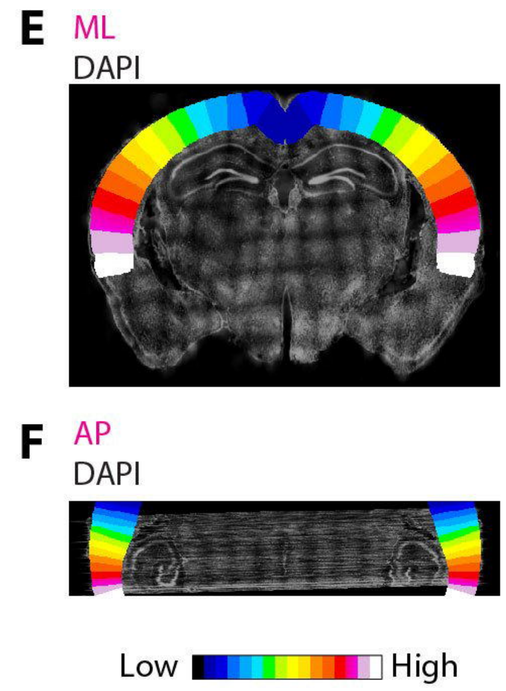

A
Seq n
Seq n+1
Tile alignment
Stitching
B
C
116
137
101
102
Rolony location
Seq
n
n+1
n+2
Rolony matching
116
137
137
102
116
137
102
116
101
116
137
102
116
101
D
GTAC
Before
Corrected

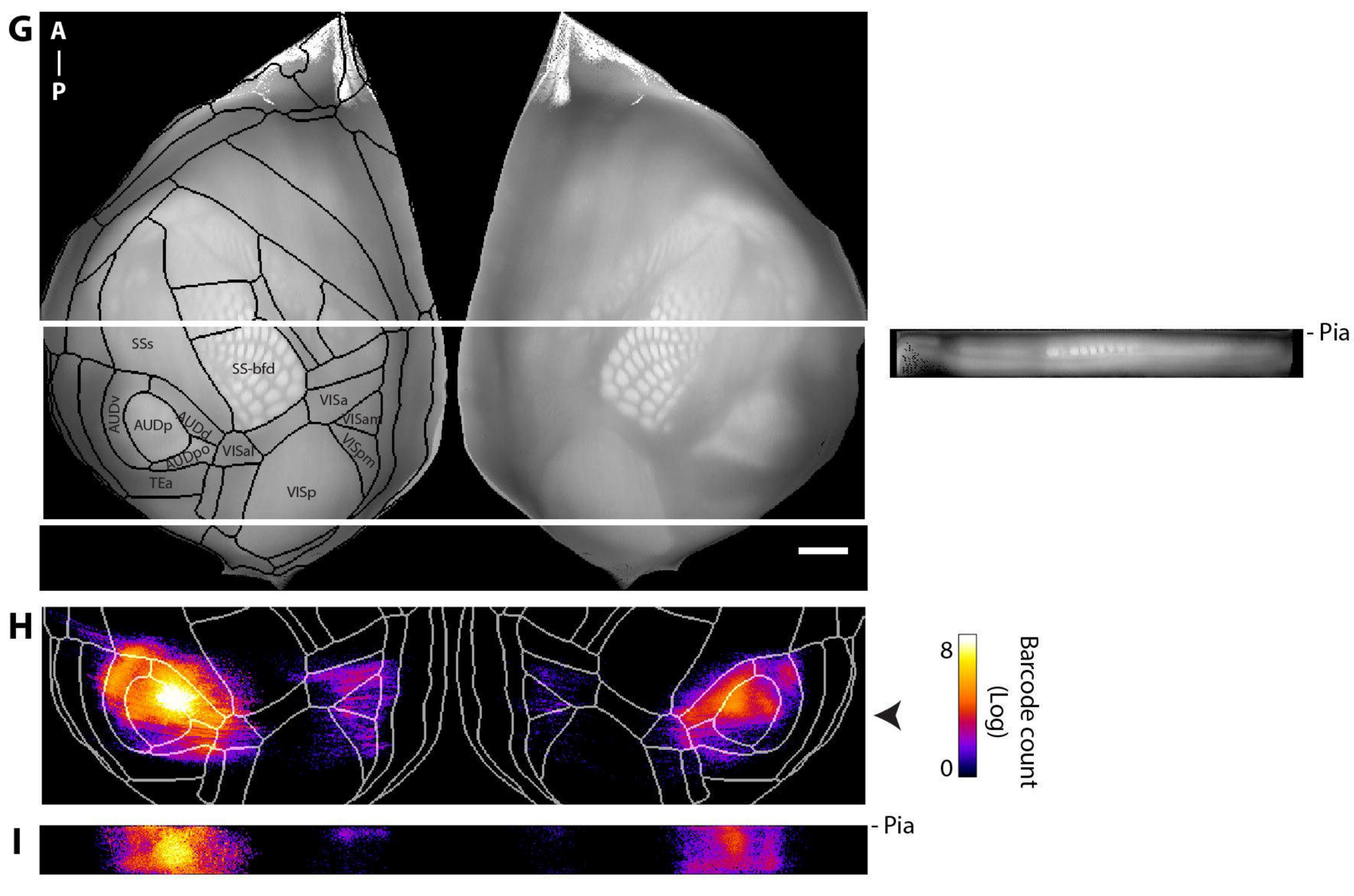

A
Rolony sequencing result
Soma basecalling result
Soma lookup table
Codebook
Axon lookup table
Correction
SomaBC
AxonBC
Filter out error-prone BC
Secondary infection exclusion
Repeated rolony exclusion
Floating rolony exclusion
Non-neural cell exclusion
Additional filters using registration info
SomaBC
Codebook
AxonBC

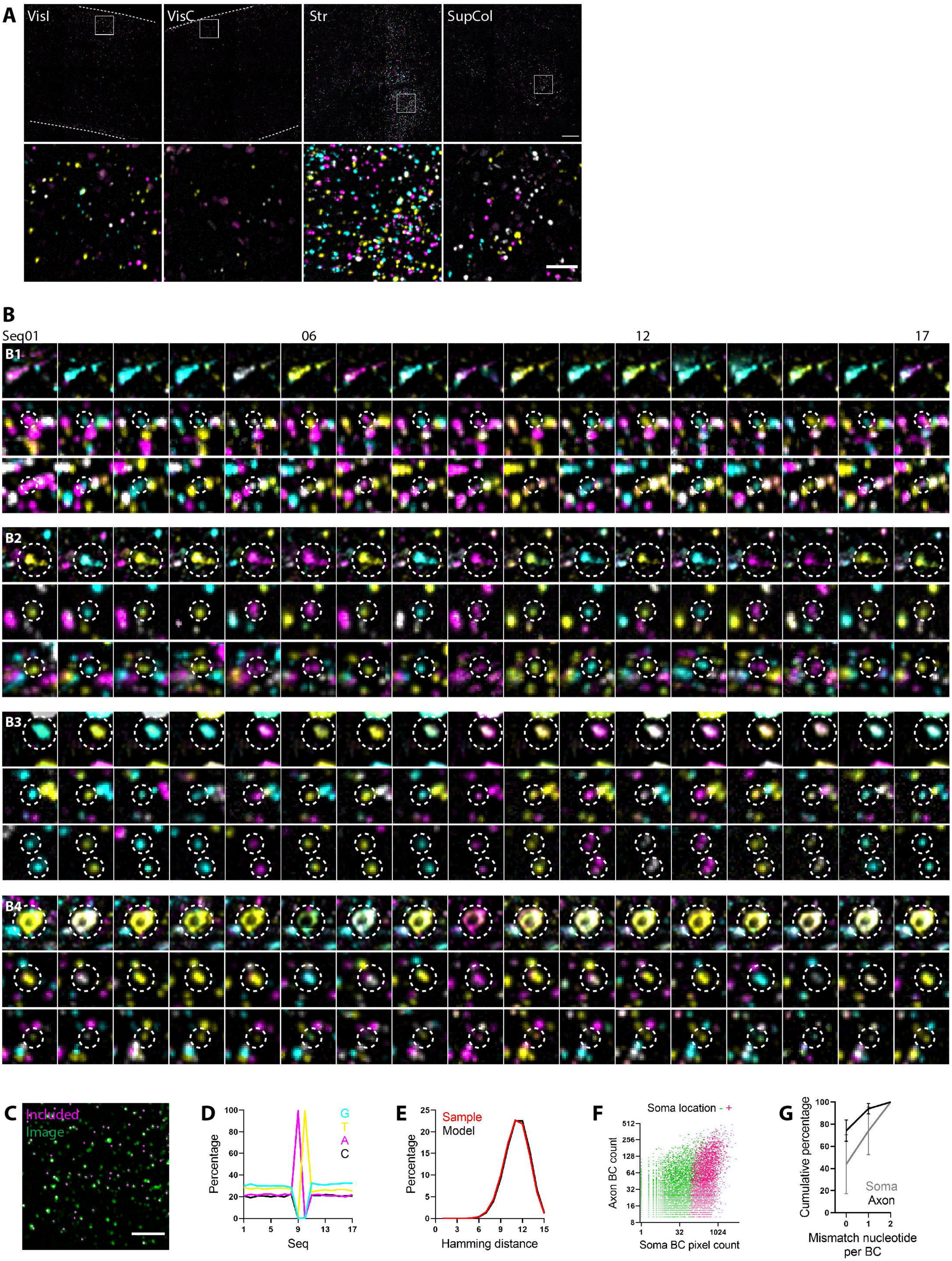

B
Percentage
100
10
1
0.1
0.01
0
5
10
15
Hamming distance

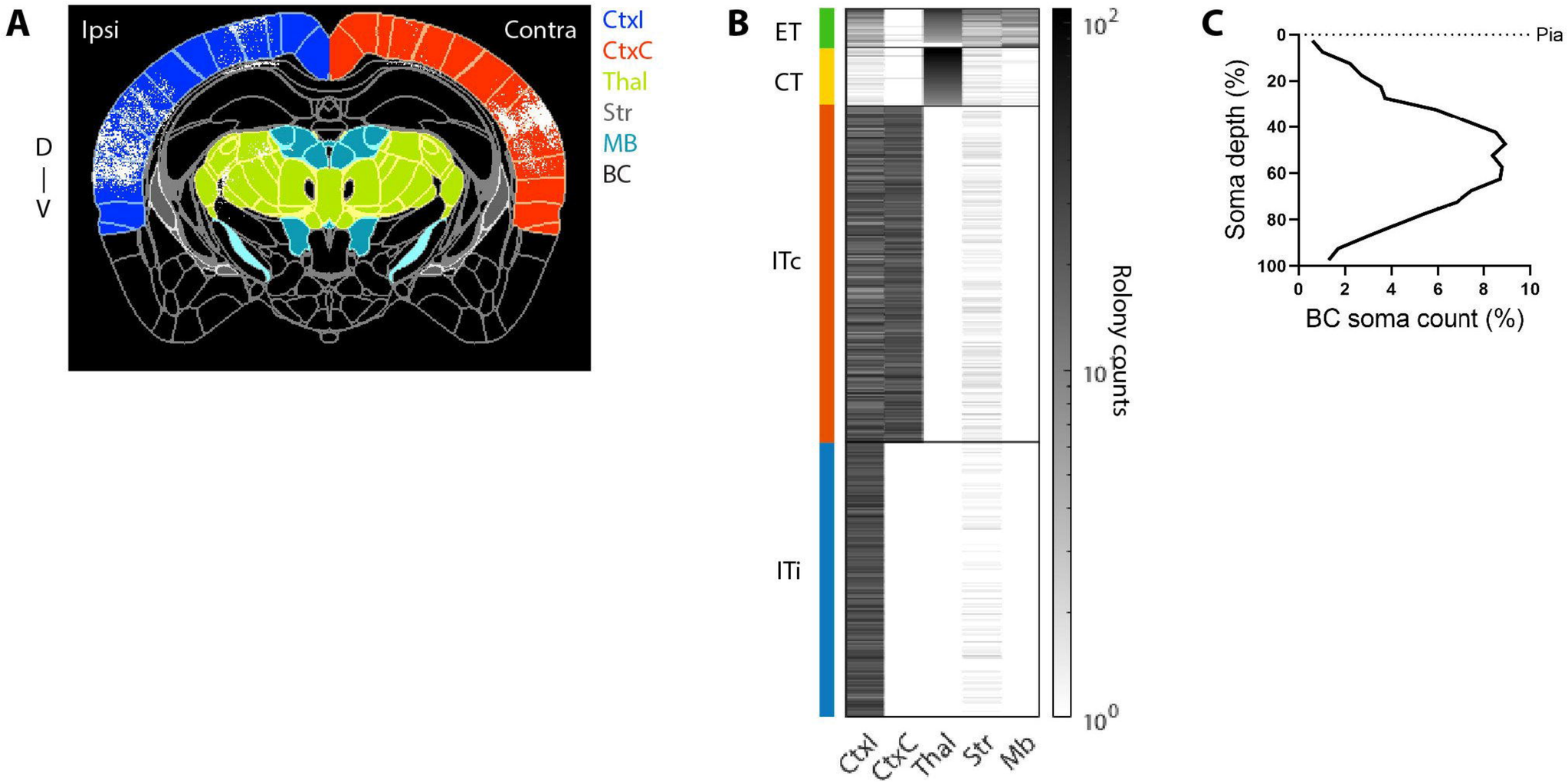

C
Ref
Data

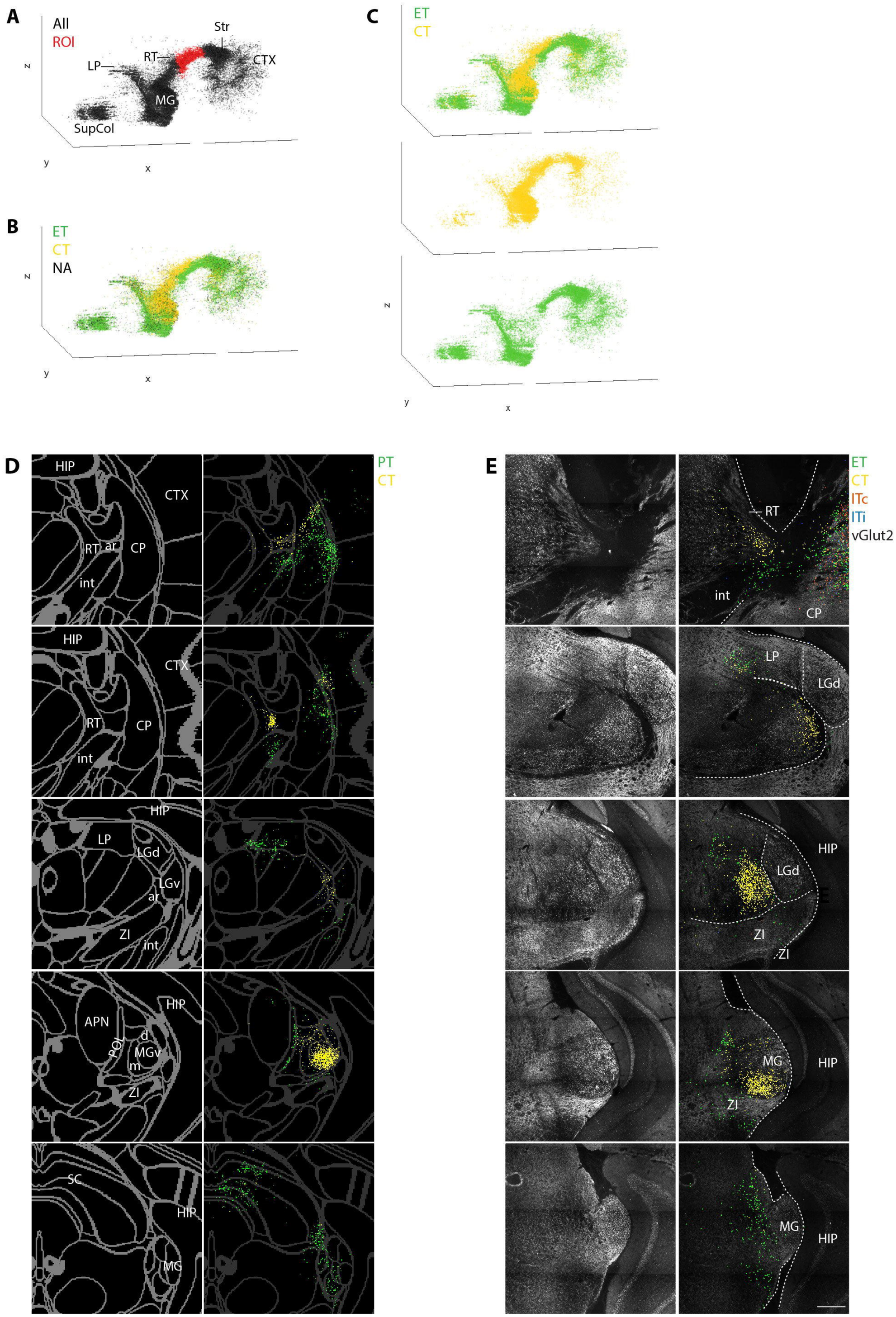

D
Depth (%)
DAPI

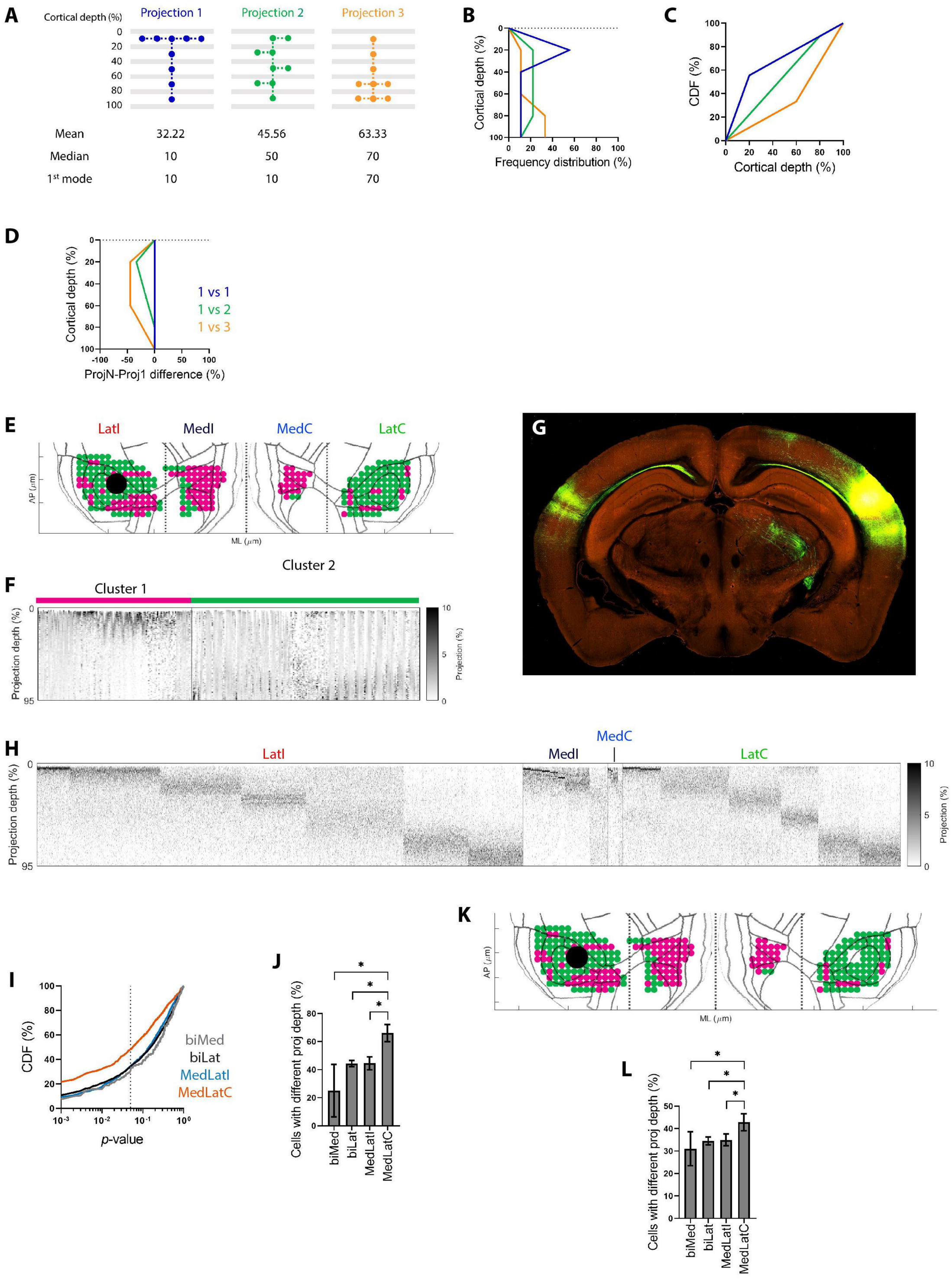

E
ML
DAPI
F
AP
DAPI
Low
High

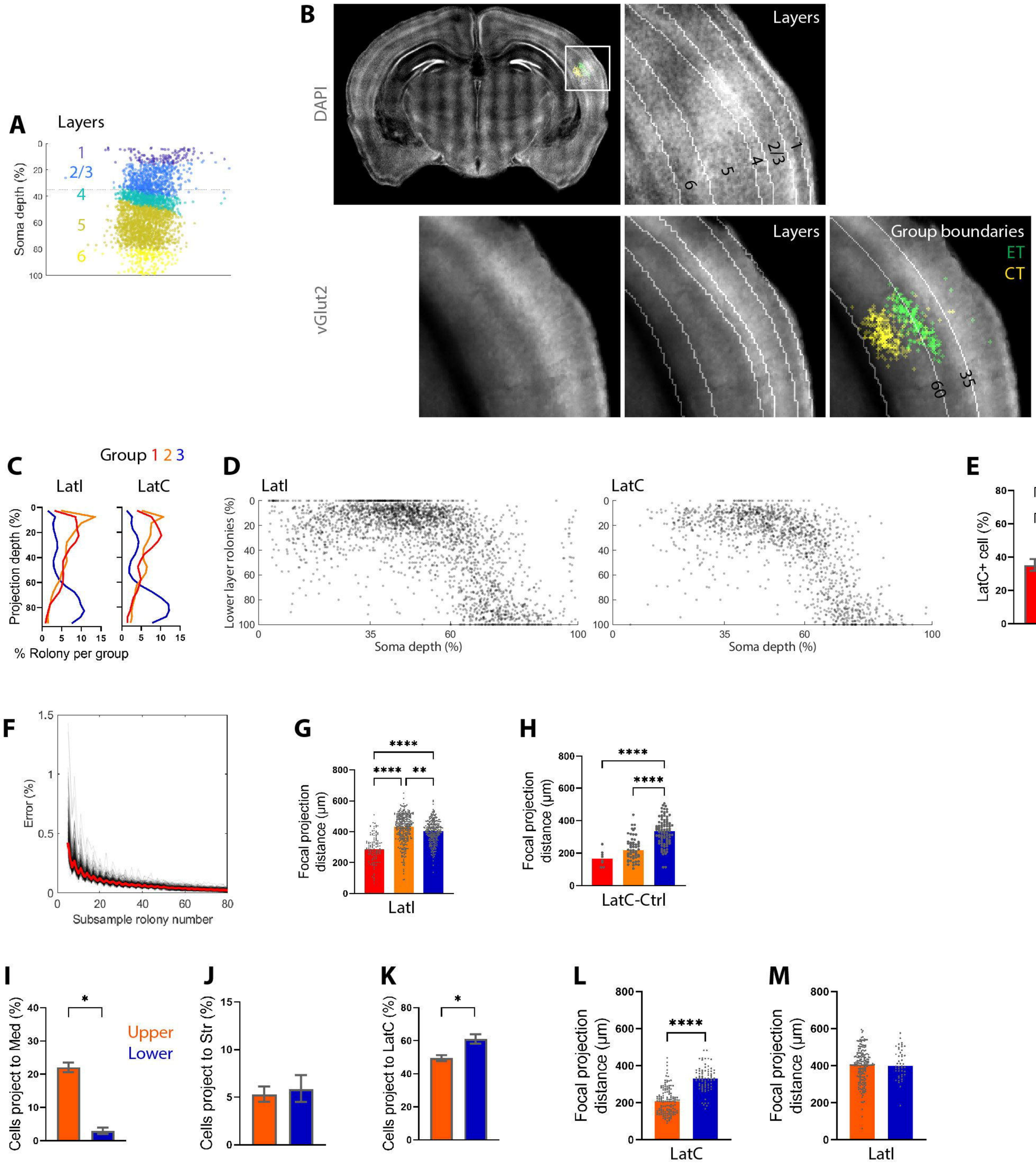

G
A
P
SSs
SS-bfd
AUDv
AUDp
AUDd
AUDpo
VISa
VISam
VISal
VISpm
TEa
VISp
- Pia
H
8
0
Barcode count
(Log)
I
- Pia


A
VisI
VisC
Str
SupCol
B
Seq01
06
12
17
B1
B2
B3
B4
C
Included
Image
D
Percentage
Seq
G
T
A
C
E
Sample
Model
Percentage
Hamming distance
F
Soma location - +
Axon BC count
Soma BC pixel count
G
Cumulative percentage
Soma
Axon
Mismatch nucleotide
per BC


A
Ipsi
Contra
CtxI
CtxC
Thal
Str
MB
BC
D
V
B
ET
CT
ITc
ITi
CtxI
CtxC
Thal
Str
Mb
10^2
10^1
10^0
Rolony counts
C
Pia
Soma depth (%)
0
20
40
60
80
100
0
2
4
6
8
10
BC soma count (%)


A
All
ROI
Str
RT
LP
CTX
MG
SupCol
z
y
x
B
ET
CT
NA
z
y
x
C
ET
CT
z
y
x
D
PT
CT
HIP
CTX
RT
ar
CP
int
HIP
CTX
RT
CP
int
HIP
LP
LGd
LGv
ar
ZI
int
HIP
APN
POL
d
MGv
m
ZI
SC
HIP
MG
E
ET
CT
ITc
ITi
vGlut2
RT
int
CP
LP
LGd
HIP
LGd
ZI
ZI
MG
HIP
ZI
MG
HIP


A
Cortical depth (%)
Projection 1
Projection 2
Projection 3
Mean 32.22 45.56 63.33
Median 10 50 70
1st mode 10 10 70
B
Cortical depth (%)
Frequency distribution (%)
C
CDF (%)
Cortical depth (%)
D
Cortical depth (%)
1 vs 1
1 vs 2
1 vs 3
ProjN-Proj1 difference (%)
E
LatI
MedI
MedC
LatC
AP (μm)
ML (μm)
F
Cluster 1
Cluster 2
Projection depth (%)
Projection (%)
G
H
LatI
MedI
MedC
LatC
Projection depth (%)
Projection (%)
I
CDF (%)
biMed
biLat
MedLatI
MedLatC
p-value
J
Cells with different proj depth (%)
biMed
biLat
MedLatI
MedLatC
K
AP (μm)
ML (μm)
L
Cells with different proj depth (%)
biMed
biLat
MedLatI
MedLatC


A
Layers
Soma depth (%)
1
2/3
4
5
6
B
DAPI
vGlut2
Layers
Group boundaries
ET
CT
35
60
C
Group 1 2 3
LatI
LatC
Projection depth (%)
% Rolony per group
D
LatI
LatC
Lower layer rolonies (%)
Soma depth (%)
E
LatC+ cell (%)
F
Error (%)
Subsample rolony number
G
Focal projection distance (µm)
LatI
H
LatC-Ctrl
I
Cells project to Med (%)
Upper
Lower
J
Cells project to Str (%)
K
Cells project to LatC (%)
L
LatC
M
LatI