Spatial organization of projection neurons in the mouse auditory cortex identified by *in situ* barcode sequencing

Xiaoyin Chen[1], Huiqing Zhan[1], Justus M Kebschull[1,2], Yu-Chi Sun[1], and Anthony M Zador[1*]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

[2]Watson School of Biological Sciences, Cold Spring Harbor, NY 11724, USA

*Correspondence: zador@cshl.edu

Summary

Understanding neural circuits requires deciphering the interactions of myriad cell types defined by anatomy, spatial organization, gene expression, and functional properties. Resolving these cell types requires both single neuron resolution and high-throughput, a combination that is challenging to achieve with conventional anatomical methods. Here we introduce BARseq, a method for mapping the projections of thousands of spatially resolved neurons by combining the high throughput of DNA sequencing with the high spatial resolution of microscopy. We used BARseq to determine the projections of 1309 neurons in mouse auditory cortex to 11 targets. We observed 264 distinct projection patterns. Hierarchical clustering confirmed the major classical classes of projection neurons, segregated across cortical laminae. Further analysis revealed 25 subclasses, largely intermingled across laminae. Unlike cell types defined by gene expression, projection subclasses beyond the major classes were rarely enriched in specific laminae, raising the possibility that the organization of projection patterns in mature neurons is orthogonal to that of gene expression. In this way, downstream brain areas could receive information from multiple cell types through parallel pathways. By sequencing *in situ*, BARseq has the potential to bridge anatomical, transcriptomic, functional, and other approaches at single neuron resolution with high throughput, and thereby offer unprecedented insight of the structure and function of a neural circuit.

An important challenge in neuroscience is to relate diverse characteristics of single neurons, in a co-registered fashion, within single brains[1]. Even simultaneous co-registration of two characteristics can be challenging, and has led to insights about the functional organization of neural circuits[2,3]. A high-throughput method capable of such multimodal co-registration would yield a "Rosetta Brain"—an integrative dataset that could constrain theoretical efforts to bridge across levels of structure and function in the nervous system[1].

As a first step toward this goal we began with MAPseq[4,5] (Fig. 1A, *left*), a sequencing-based method capable of mapping long-range projections of thousands of single neurons in a single brain. MAPseq achieves multiplexing by uniquely labeling individual neurons with random RNA sequences, or "barcodes". Because MAPseq, like most other sequencing methods, relies on tissue homogenization, it cannot resolve the spatial organization of the neuronal somata. This spatial organization, however, potentially allows the registration of distinct neuronal characteristics. We therefore sought to develop a method that would preserve the spatial organization of barcoded somata. By combining MAPseq with *in situ* sequencing[6-8] of somatic barcodes (Fig 1B), BARseq (Barcoded Anatomy Resolved by Sequencing; Fig. 1A, *right*) preserves the spatial organization of neurons during projection mapping, and has the potential to link multiple sources of data in a single specimen.

We applied BARseq to study the organization of the auditory cortex. The spatial organization of projection neurons into layers reflects the functional organization of circuitry in the neocortex and other brain areas. The major classes of neocortical excitatory neurons—corticothalamic (CT), pyramidal tract (PT), and intratelencephalic (IT)—are defined by their long-range axonal projections, and are segregated within cortical laminae[9,10]. More recently, single cell sequencing has uncovered remarkable diversity in gene expression patterns[11-14]. Many classes of projection neurons identified by gene expression are restricted to specific laminae[13], but it remains unclear whether projection patterns show comparable diversity, and whether projection diversity similarly respects laminar organization.

To provide a bridge between the laminar position of a barcoded soma and its distal axonal projections, we adapted BaristaSeq (Supp. Fig. S1), a targeted approach for sequencing barcodes *in situ* developed in non-neuronal cells[8], to neurons. BaristaSeq enabled efficient amplification and sequencing of barcodes in cultured neurons co-expressing barcodes and GFP (Fig. 1C, D). We then further adapted BaristaSeq to brain slices. We identified a reaction chamber system that was both physically and chemically compatible with our samples (Supp. Fig. S2A; Supp. Note 1) and increased tissue accessibility (Supp. Fig. S2B; Supp. Note 1). With these modifications, we obtained highly efficient amplification of barcodes in brain slices (Fig 2A; Supp. Fig. S2C; Supp. Note 1). We also optimized Illumina sequencing chemistry for brain slices, resulting in a ~10-fold improvement in signal-to-noise ratio (SNR) compared with sequencing by ligation (SOLiD; Supp. Fig. S2D, E; Supp. Note 1), which had previously been used for *in situ* sequencing[6,7].

To evaluate barcode sequencing in brain slices, we sequenced 25 bases in a sample infected with a diverse ($>10^6$) pool of barcoded Sindbis virus[5,15] injected in the auditory cortex (See *Methods*; Fig. 2B). Basecall quality (Fig. 2C; Supp. Fig. S2F; see *Methods*) and signal intensity (Supp. Fig. S2G) remained high through all 25 cycles. We also observed no bias towards any particular base (Supp. Fig. S2H). Read accuracy was high: barcodes in 50/51 (98%) randomly selected cells matched perfectly to known barcodes in the library, and the remaining cell had only a single base mismatch to the closest barcode in the library (Fig. 2D, E). Because the barcodes in this library

represents a small fraction (~10^{-9}) of all possible 25-mer barcodes (>10^{15}), random 25-nt sequences on average had seven mismatches to the closest known barcode, indicating that our barcode reads were unlikely to be false positive matches by chance. These results indicate that *in situ* sequencing of RNA barcodes in brain slices is both accurate and efficient.

To assess the sensitivity and specificity of BARseq, we compared it to standard retrograde tracing using cholera toxin subunit B (CTB) by mapping contralateral projections (Fig. 3A; Supp. Fig. S3A; Supp. Table S1). We injected a highly diverse (~10^7 barcodes) viral library into the auditory cortex and CTB in the contralateral auditory cortex. We then microdissected the contralateral auditory cortex into "cubelets" and homogenized, and sequenced the barcodes within. By matching the barcode sequences obtained at the somatic injection site *in situ* with the sequences obtained at the contralateral auditory cortex, we could associate the precise position of CTB labeled neuron with its axonal projections to distant brain areas as mapped by BARseq. The majority (12/13, 92%) of CTB labeled neurons were detected by BARseq, whereas only 28% (12/43) of contralaterally projecting neurons identified by BARseq were labeled with CTB (Fig. 3B; Supp. Table S1). The contralateral projections were much stronger than the noise threshold defined by projection strengths in the olfactory bulb (Supp. Fig. S3B), indicating that the higher apparent sensitivity of BARseq was not due to false positives resulting from contaminating barcodes. These results indicate that BARseq retains the high specificity and sensitivity of MAPseq[5] (91 ± 6%), and may exceed the sensitivity of conventional CTB tracing.

Having established the sensitivity and specificity of BARseq, we applied it to the organization of long-range projections from the mouse auditory cortex. We focused on 11 auditory cortex projection target areas[16], using the olfactory bulb as a negative control. We analyzed 6391 neurons with high-quality projection data, consisting of 1309 neurons from two brains using BARseq and 5082 neurons from one brain using conventional MAPseq (Supp. Note 2; Supp. Fig. S4; Supp. Table S2). Consistent with conventional bulk GFP tracing[16] (Fig. 3C; Pearson correlation coefficient $r = 0.94$, $p < 0.0001$), projections to the thalamus, the tectum, the contralateral auditory cortex, and the caudal striatum were particularly strong (Fig. 3D). Projections determined by BARseq were thus consistent with those obtained by conventional bulk labeling techniques.

The projection patterns of individual neurons were remarkably diverse. After binarizing projections, we observed 264 distinct patterns, more than an eighth of the approximately 2^{11} = 2048 possible patterns (Supp. Note 3). Individual neurons projected to up to nine areas, and most (64%) neurons projected to two or more areas (Fig. 3E). Because we did not sample all potential targets, the actual number of targets per neuron and distinct projection patterns are likely higher. This diversity is consistent with previous reports[4], and may be a general feature of cortical projection patterns.

To explore the structure underlying this diversity we clustered the non-binarized projection patterns hierarchically (Supp. Fig. S5A-F, S6A-E; Supp. Note 4). This clustering partitioned neurons into established cortical classes[9]: CT, PT and IT. Further clustering revealed two classes of IT neurons, those that have a contralateral projection (ITc) and those that only project ipsilaterally (ITi; Fig. 3F). Interestingly, although both IT subclasses projected to both the caudal striatum—an area implicated in auditory decision making[17]—and the rostral striatum, PT neurons did not project to the rostral striatum. We confirmed this finding by triple retrograde tracing (Supp. Fig. S7A): a significant fraction of neurons (66/236) projecting to the rostral

striatum also projected to the caudal striatum, but none (0/236) projected to the tectum (Supp. Fig. S7B, C). The results of BARseq were thus consistent with classical approaches.

The large size of the pooled MAPseq and BARseq samples (6391 neurons) enabled us to look for projection subclasses within the established classes. Clustering revealed 25 statistically significant (see *Hierarchical clustering* in *Methods*) "subclasses" of projections neurons, at four clustering levels (Fig. 3G, H; Supp. Fig. S6E; Supp. Note 5 for details on the subclasses). These results indicate that BARseq identified candidate subclass-level structures in the auditory cortex, beyond the major divisions of projection neuron types.

The tripartite classification of cortical excitatory projection patterns into PT, IT and CT classes is supported by the corresponding segregation of these neurons into distinct laminae. Because BARseq recovers the positions of barcoded somata, we could associate the projection pattern of each neuron in our dataset with its laminar position. As expected, the major classes segregated strongly with laminar position: superficial layers contained predominantly IT neurons, whereas deep layers contained predominantly CF (PT and CT) neurons[9] (Fig 4A; see Supp. Note 6 and Supp. Fig. S8 for the estimation of layers). Thus major classes of cortical projection neurons defined by BARseq had distinct laminar distributions, consistent with observations using conventional methods[18].

Many cell types defined by gene expression are highly enriched in single laminae[12,13,19], but it is unclear if projection patterns are similarly organized into laminae. We therefore further examined whether PT and IT subclasses defined by projections were restricted to single layers (CT neurons were restricted to L6, and therefore were not examined). Most subclasses defined by projections spanned many layers (Fig. 4B, C; Supp Note S7; Supp. Fig. S9A, B). We identified no IT subclass restricted to L2/3, and only a small fraction of IT neurons (45 out of 805, 6%) were found in L5 or L6 enriched subclasses (Leaf 43, 19, and 45). Similarly, we identified no PT subclass specific to L5 or L6, in contrast to previous reports of layer-specific PT classes defined by morphological and electrophysiological properties[20] (Supp. Note 7; Supp Fig. S9C-E). Thus classes defined by projections do not appear to be restricted to single laminae, suggesting that projection patterns may not correspond to cell types defined by other neuronal features.

There are at least three possible interpretations of our observation that projection patterns, unlike patterns of gene expression in adult animals, did not appear to be well explained by laminar position. First, we cannot rule out the possibility that we failed to discover the "true" underlying projection subclasses that comprise auditory cortex, and that these subclasses do in fact segregate according to laminae. Second, it could be that although the connections are structured within a class (Supp. Fig. S9F, G), these structures are not entirely categorical and hence will not be revealed by clustering. Finally, our results raise the intriguing possibility that the organization of projection patterns in mature neurons is orthogonal to that of gene expression. Such an intersectional organization, especially in the IT classes, is consistent with scRNAseq results of retrogradely labeled neurons[13] and could arise either from temporal changes in gene expression over development, or through probabilistic or activity-dependent pruning of projections during development. Thus downstream brain areas may receive information from multiple cell types through parallel pathways.

Neuronal classes are defined by the combination of multiple features[14,19,21], including anatomical characteristics such as morphology and connectivity[22,23], molecular characteristics such as

gene[11-13] or protein expression, and functional characteristics such as behaviorally-evoked activity[21]. Most current neuronal taxonomies focus on just a single feature, but partitioning based on one feature alone does not reveal the interaction among neuronal features that result in circuit functions. The spatial organization of neurons can be used to register multiple features at cellular resolution. Because BARseq preserves the spatial organization of projection neurons, it has the potential to be combined with other *in situ* techniques [e.g., *in situ* sequencing of mRNAs (Supp. Fig. S10A) and functional imaging (Supp. Fig. S10B)]. BARseq can thus potentially achieve a "Rosetta Brain"[1]—a comprehensive multi-modal description of the vertebrate brain at cellular resolution.

Methods

Viruses, constructs, and oligos

The plasmid encoding the Sindbis barcode library (JK100L2, https://benchling.com/s/EKtQttOe) is available from Addgene (#79785). The RT primer (XC1215), the padlock probe / sequencing primer (XC1164), and the fluorescent probe for visualization (XC92) were described previously[8].

For validation of BaristaSeq in brain slices, we used a barcode library previously described by Kebschull et al.[5]. The library contained 1.5 million known 30-nt random barcode sequences, which represented ~97% of all barcodes in the library. For BARseq experiments, we used a separate diverse barcode library with ~10^7 diversity[4]. This library was not fully sequenced *in vitro*.

Animals and tissue processing:

Eight to nine week old male C57BL/6 mice were injected in the left auditory cortex at -4.8 mm ML, -2.6 mm AP from bregma, with 140nL Sindbis virus at each of the following depths (200 µm, 400 µm, 600 µm, and 800 µm) at 30˚ angle. For samples prepared for BaristaSeq only, we transcardially perfused the animal with 10% formalin, then postfixed the tissues for 24 hrs. We then cryo-protected the brain in PBS with 10% sucrose for 12 hrs, 20% sucrose PBS for 12 hrs, and 30% sucrose PBS for 12 hrs. We then embedded the brain in OCT (Electron Microscopy Sciences) and cryo-sectioned to 14 µm slices onto SuperFrost Plus slides (VWR).

For BARseq samples, we transcardially perfused the animal with PBS 43-45 hrs post-injection. We cut out the left auditory cortex from the brain and post-fixed it in 10% formalin at 4□ for 8 hrs, and snap-froze the rest of the brain on a razor blade on dry ice. The snap-frozen brain (without the injection site) was then processed for conventional MAPseq as described[5]. The post-fixed auditory cortex was cryo-protected, embedded, and cryo-sectioned as described above.

Fluorescent *in situ* hybridization (FISH)

FISH was performed using Panomics ViewRNA ISH cell assay kits (Thermo Fisher Scientific) according to the manufacturer's protocols. Type 1 probes against Cux2 and Fezf2 (Thermo Fisher Scientific) were used to visualize mRNA expression of the two genes, respectively.

BaristaSeq

BaristaSeq on cultured neurons was performed as described[8]. Briefly, the neurons were fixed in 10% formalin, washed in PBST (PBS with 0.5% tween-20), and dehydrated in 70%, 85%, and 100% ethanol for an hour. After rehydration in PBST, we incubated the samples in 0.1M HCl for 5 mins, followed by three PBST washes. We then reverse transcribed the samples [1 U/µl RiboLock RNase inhibitor (Thermo Fisher Scientific), 0.2 µg/µl BSA, 500 µM dNTPs (Thermo Fisher Scientific), 1 µM RT primer, and 20 U/µl RevertAid H Minus M-MuLV reverse transcriptase (Thermo Fisher Scientific) in 1× RT buffer] at 37 ˚C overnight. After reverse transcription, we crosslinked the cDNAs in 50 mM $BS(PEG)_9$ for 1 hour and neutralized with 1M Tris-HCl for 30 mins. We then gap-filled and ligated padlock probes [100 nM padlock probe, 50 µM dNTPs, 1 U/µl RiboLock RNase inhibitor, 20% formamide (Thermo Fisher Scientific), 0.5 U/µl Ampligase (Epicentre), 0.4 U/µl RNase H (Enzymatics), and 0.2 U/µl Phusion DNA polymerase (Thermo Fisher Scientific) in 1× ampligase buffer supplemented with additional 50 mM KCl] for 30 mins at 37 ˚C and 45 mins at 45 ˚C. Following PBST washes, we

performed rolling circle amplification (RCA) [20 uM aadUTP, 0.2 ug/ul BSA, 250 uM dNTPs, and 1 U/ul □29 DNA polymerase in 1× □29 DNA polymerase buffer supplemented with 5 % additional glycerol] overnight at room temperature. After crosslinking the rolonies using $Bs(PEG)_9$ and neutralized with Tris-HCl, we hybridized 2.5 µM sequencing primers or 0.5 µM fluorescent probes in 2× SSC with 10 % formamide, washed three times in the same buffer, and proceeded to sequencing or imaging.

For BaristaSeq on brain tissues, all slides with brain slices were first sealed in HybriWell-FL chambers (22 mm x 22 mm x 0.25 mm; Grace Bio-labs) for reactions. The brain slices were washed three times in PBS supplemented with 0.5% Tween-20 (PBST), followed by a 3 min digestion by 0.2% pepsin in 0.1M HCl. We then proceeded with ethanol dehydration, followed by reverse transcription, padlock gap-filling and ligation, and RCA as described above.

Sequencing of the rolonies was performed as described previously[8] with minor modifications using the Illumina HiSeq SBS kit v4. We reduced the CRM reaction time to two minutes each, which was sufficient for the heat transfer, and increased the PBST washes to four to eight times after the IRM reactions.

Imaging

All imaging except for the experiments in Fig. 2B-E was performed on an UltraView VoX spinning disk confocal microscope (Perkin Elmer) as previously described[8]. The sequencing channels and their calibrations were described previously[8]. The sequencing channel filters were as previously described[8]. The experiments in Fig. 2B-E were produced on a Zeiss LSM 710 laser scanning confocal microscope as previously described[8].

Base-calling

Base-calling was performed as described[8], except that the images were first processed through a median filter and a rolling ball background subtraction to remove noise and background fluorescence. The sequencing quality score is defined as the intensity of the called channel divided by the root sum square of all four channels.

Comparison to other *in situ* sequencing techniques

The original padlock probe based barcode amplification was performed similar to BaristaSeq amplification as described above, except that the Stoffel fragment (DNA Gdansk) was used in place of Phusion DNA polymerase and the cDNA crosslink was done using 4% paraformaldehyde in PBST for 10 mins[7].

To perform targeted FISSEQ in brain slices, we processed the sample in the same way as in BaristaSeq in brain slices to the cDNA crosslink step. After cDNA crosslinking, we digested the RNAs [10 µl RiboShredder (Epicentre) and 5 µl RNase H in 1× RNase H buffer] for 1 hour at 37 ˚C. After washing the samples twice in water, we circularized the cDNAs [0.5 mM DTT, 1M Betaine, 2.5 mM $MnCl_2$, and 1 U/µl Circligase II in 1× Circligase buffer] for 1 hour at 60 ˚C. After washing the samples with PBST, we hybridized 1.5 µM RCA primers in 2× SSC with 10% formamide for 1 hour. We then washed the samples three times with the same buffer, twice more with PBST, and proceeded to RCA as in BaristaSeq.

To compare Illumina sequencing chemistry to SOLiD sequencing chemistry for *in situ* sequencing in tissues, we performed SOLiD sequencing as described previously[6,8]. To calculate the signal to noise ratio, we first converted the four-channel images into one single channel,

taking the maximum value of the four channels for each pixel. We selected areas containing barcoded cells or areas containing tissues but no barcoded cells by thresholding as the signal and background areas. We then subtracted the black point of the camera from both the signals and the backgrounds and calculated the SNR. The SNR was calculated using the same selected areas in all six cycles.

BARseq

Animals were injected and processed as described above. Cryo-sectioned brain slices were first imaged to generate DIC and GFP images, before they were processed for BaristaSeq. The GFP images from neighboring slices were aligned to each other manually to remove deformation during sectioning. The sequencing images were then registered back to the GFP images to locate the positions of the neurons within the slice. Each basecall ROI was thus registered back to the aligned GFP images.

We dissected 12 projection sites from the frozen brains for sequencing. These 12 sites include the olfactory bulb, the orbitofrontal cortex, the motor cortex, the rostral and caudal striatum, the somatosensory cortex, the ipsilateral and contralateral visual cortex, the contralateral auditory cortex, the thalamus, and the tectum. For several brain areas, we limited the areas collected to prevent cross contamination. These areas include the somatosensory cortex (restricted to the upper-limb, lower-limb, and the trunk areas), the two visual cortices (restricted to mostly area pm and am), and the striatum (the rostral striatum and the caudal striatum samples were separated by two brain slices, or 600 µm). The contralateral auditory sample also included the neighboring temporal association area. Sample slice images from the brain XC9 after dissection indicating the locations of the collected projection sites are available at Dryad (see *Data and software availability* in *Methods*). The projection sites were sequenced as described for MAPseq[5].

We first filtered the MAPseq generated barcodes so that all barcodes had at least 10 molecules but no more than 10000 molecules at the strongest projection site. We recovered 26840 barcodes using these criteria from the three brains (Supp. Table 1). We then matched these barcode sequences at the projection sites to those at the injection site, allowing three mismatches for conventional MAPseq or one mismatch for BaristaSeq. In the conventional MAPseq brain (XC14), 5082 out of 8418 barcodes were confirmed to be from the auditory cortex and were used for the subsequent analyses.

In the two BARseq brains (XC9 and XC28), we sequenced 3237 cells *in situ*. Of all sequenced cells, 1806 (56%) cells had corresponding sequences at any projection site. The remaining cells had either low read qualities (possibly from having more than one barcode in the cell), or did not project to the examined areas (e.g., local interneurons, excitatory neurons that project to secondary auditory areas, and non-neuronal cells). We further filtered out barcodes with fewer than 10 molecules in the maximum projection area, removed neurons below the bottom of the cortex (these are likely persistent subplate neurons in the collosal commissure) and neurons in highly distorted slices (as judged by an abnormal cortical thickness). After filtering, 1309 neurons were used in the analyses.

The projection strengths to each area were obtained by normalizing the barcode molecule counts to the amount of spike-in RNAs[5] recovered in each area. When counting the number of projections per neuron, we only counted projections with the number of barcode molecules above the noise floor. The noise floor was defined as the maximum number of barcode molecule

count in the negative control area (the olfactory bulb). For all subsequent analyses of the projection data, we used the logarithm of the barcode molecule counts after normalization to spike-in RNAs and pooled the data from three brains together.

Comparison of BARseq to retrograde tracer

We injected 140 nL Alexa 647 labeled cholera toxin subunit B (CTB) into the right auditory cortex at 4.8 mm ML, -2.6 mm AP at multiple depths (200 µm, 400 µm, 600 µm, and 800 µm) at 30˚ angle. After 48 hrs, we injected 140 nL JK100L2 virus into the left auditory cortex at -4.8 mm ML, -2.6 mm AP, with 140nL Sindbis virus at each depth (200 µm, 400 µm, 600 µm, and 800 µm) at 30˚ angle. After incubation for another 44 hrs, we processed the animal for BARseq, but collected four target sites, including the olfactory bulb (negative control), the contralateral auditory cortex (i.e. the center of the CTB injection), the remaining areas where CTB is visible to the naked eye, and the surrounding areas where CTB is visible under fluorescent microscope. The last three samples thus formed concentric rings around the CTB injection site. All three samples gave consistent results regarding contralateral projections (Supp. Table S1)

Before library preparation for BaristaSeq, we also imaged the Alexa 647 channel to locate retrograde-labeled neurons. These images were aligned to the sequencing images. To find the fraction of BARseq identified projection neurons that are also labeled by CTB, we picked barcoded neurons with a minimum sequencing quality of 0.75 and counted the number of neurons labeled by CTB. To find the fraction of CTB labeled neurons that are also labeled by BARseq, we identified neurons labeled by both CTB and barcodes with a minimum sequencing quality of 0.75, and counted the number of neurons with barcodes in the contralateral auditory cortex above the noise floor. The noise floor is set to be the maximum count of individual barcodes recovered in the olfactory bulb.

Validation of BARseq identified cell types using retrograde tracing

To validate the striatal projections of PT and IT neurons, we injected red RetroBeads (LumaFluor) diluted 1:1 in PBS in the superior colliculus at -4.8 mm AP, -0.7 mm ML at depths 500 µm, 700 µm, 900 µm, 1100 µm, and 1300 µm (70 µl per depth) from the surface of the brain, Alexa 488 labeled CTB in the caudal striatum at -1.6 mm AP, -3.2 mm ML at depths 2.5 mm and 3 mm (50 µl per depth) from the surface of the brain, and Alexa 647 labeled CTB in the rostral striatum at 0.6 mm AP, -2 mm ML at depths 2.5 mm and 3 mm (50 µl per depth) from the surface of the brain. After 96 hrs, we perfused the animal and sliced the auditory cortex coronally into 70 µm slices. We then imaged the slices on an UltraView VoX spinning disk confocal microscope (Perkin Elmer).

Comparison of BARseq bulk projection pattern to bulk GFP tracing

For bulk projection comparison to GFP tracing data, we used the bulk GFP tracing data from five brains in the Allen connectivity database[16] (experiments 116903230, 100149109, 120491896, 112881858, and 146858006; © 2011 Allen Institute for Brain Science. Allen Mouse Brain Connectivity Atlas. Available from: http://connectivity.brain-map.org/). All five brains had cells labeled in the primary auditory cortex and no labeling in non-auditory area. For both BARseq/MAPseq bulk projections and GFP bulk tracing data, we normalized the projection strengths to individual areas to total projection strengths in all examined areas for that brain first, and then averaged across brains (five brains for GFP tracing and three brains for

BARseq/MAPseq). We then calculated the correlation coefficients between the GFP tracing and BARseq/MAPseq bulk projection strengths of the corresponding brain areas.

Hierarchical clustering of projection data

All clustering analyses were done using the logarithm of the spike-in-normalized projection strength. We first filtered the projection data using non-negative matrix factorization (NMF)[24] in Matlab. We approximated the projection pattern X of m neurons to n areas as the product of Y, an m by k matrix containing the loadings of the k projection modules for each neuron, and A, a k by n matrix containing the projection pattern of each projection module. Each of the k projection modules represents a set of projections that correlate with each other. To determine the number of projection modules k, we looked for k values that did not result in a significant increase in the correlation among neurons using the filtered projection data X' and resulted in similar classification of neurons in the first two hierarchy. This resulted in $k = 6$. We then reconstructed the filtered projection data $X' = Y * A$. The filtered projection data X' was used for clustering.

During each step of the hierarchical clustering, we split each node into two groups using k-means clustering on the squared Euclidean distance of the projection patterns. We then calculated the significance of the split using a Matlab implementation of SigClust (http://www.unc.edu/~marron/marron_software.html). We kept the new clusters if the split was significant after Bonferroni correction and the sizes of the resulting clusters were larger than 1% of all data points. This procedure was repeated for each new node until no new clusters were found.

We then validated our data using random forest as described[12]. For each pair of clusters, we trained a random forest classifier on 80% of the data. We then used the classifier to classify the remaining 20% of the data. We repeated this process five times, each time using a mutually exclusive group of 20% of data, so that all data were classified once. We repeated this whole process 10 times for each pair of clusters, so that all data were classified between each pair of clusters 10 times. For each barcode, we then removed all cluster memberships that were scored 0 out of 10 in any one of the pairs involving that cluster. The remaining clusters (i.e. ones that have scored at least 1 out of 10 comparisons in any pairwise comparison) were assigned to the barcode, with the main identity as the cluster with the highest sum of scores across all pairwise comparisons involving that cluster.

To compare clusters obtained using k-means with those obtained using Louvain community detection[25], we performed Louvain community detection hierarchically using a similarity matrix $\boldsymbol{S}$, in which each element $\boldsymbol{S}_{ij}$ is the difference between the Euclidean distance of the two data points and the maximum Euclidean distance of any two data points in the dataset. Louvain community detection was performed using a MATLAB implementation of the algorithm (https://perso.uclouvain.be/vincent.blondel/research/Community_BGLL_Matlab.zip).

To compare clusters obtained using k-means with those using spectral clustering[26], we substituted k-means with normalized spectral clustering [27] using the same similarity matrix $\boldsymbol{S}$ as described above. Spectral clustering was performed using a MATLAB implementation of the algorithm (https://www.mathworks.com/matlabcentral/fileexchange/34412-fast-and-efficient-spectral-clustering).

t-SNE[28] was performed using the MATLAB implementation of the standard t-SNE (https://lvdmaaten.github.io/tsne/) using either the log projection data as inputs.

Normalized entropy for the laminar distribution of a group of neurons

To calculate the entropy of the laminar distribution of a group of neurons, we discretized the laminar location of the neurons into 13 bins, each covering 100 µm. We then calculated the entropy of the discrete distribution of laminar locations $\boldsymbol{E} = -\sum_{i=1}^{13} P_i \log_{10} P_i$, where P_i is the probability of neurons falling into the ith bin. We then normalized $\boldsymbol{E}$ to the maximum possible $\boldsymbol{E}$ for 13 bins to obtain the normalized entropy $\boldsymbol{E}' = -\boldsymbol{E}/\log_{10}(1/13)$. The normalized entropy thus equals 0 when all neurons fall into one bin, and 1 when the neurons randomly distribute across all 13 bins.

QUANTIFICATION AND STATISTICAL ANALYSIS

All *p* values reported were after Bonferroni correction except in Supp. Fig. S9A, B. In these two figures, however, the significance level shown in the figures were corrected for multiple testing. The statistical tests used for individual experiments were noted in the main text.

DATA AND SOFTWARE AVAILABILITY

All *in vitro* high throughput sequencing datasets are being deposited to SRA. All *in situ* sequencing data and scripts for data processing are being deposited to Dryad.

Author contributions

X.C. and A.M.Z conceived the study. X.C., H.Z. and Y.S. performed BaristaSeq. J.M.K, and H.Z. performed MAPseq. X.C. and A.M.Z. analyzed the data. X.C. and A.M.Z. wrote the paper.

Acknowledgements

The authors would like to acknowledge Josh Huang, Gordon Shephard, Jessica Tollkuhn, and Jesse Gillis for useful discussions, and Barry Burbach, Nour El-Amine, and Stephen Hearn for technical support. This work was supported by the following funding sources: National Institutes of Health [5RO1NS073129 to A.M.Z., 5RO1DA036913 to A.M.Z.]; Brain Research Foundation [BRF-SIA-2014-03 to A.M.Z.]; IARPA MICrONS [D16PC0008 to A.M.Z.]; Simons Foundation [382793/SIMONS to A.M.Z.]; Paul Allen Distinguished Investigator Award [to A.M.Z.]; postdoctoral fellowship from the Simons Foundation to X.C. This work was performed with assistance from CSHL Shared Resources, which are funded, in part, by the Cancer Center Support Grant 5P30CA045508.

Conflict of interests

The authors declare no conflict of interests.

Figures:

Fig. 1. Multiplexed projection mapping using *in situ* sequencing. (A) Workflow of MAPseq (*left*) and BARseq (*right*). In both MAPseq and BARseq, a barcoded viral library is delivered to the area of interest. The source area and several target areas are then dissected. In MAPseq, barcodes in all dissected areas are sequenced using Next-Gen sequencing. Barcodes at the source site are then matched to those at the targets to find the projection patterns of individual neurons. In BARseq, the injection site is sequenced *in situ*, thus preserving spatial information during projection mapping. (B) Next-Gen sequencing (NGS) vs. *in situ* sequencing. In conventional NGS (*top*), DNAs are anchored to a flow cell (blue), amplified locally, and sequenced. During each sequencing cycle, one fluorescence-labeled base is incorporated using the amplified DNA as templates. The sequence is read out from the sequence of fluorescence over cycles. In *in situ* sequencing (*bottom*), RNA is reverse transcribed and amplified *in situ* on a slide (yellow). The amplified cDNAs are then sequenced using the same chemistry as NGS, and imaged *in situ*. Sequences can be read out together with spatial information. (C) Representative images of barcode rolonies (yellow) generated in primary hippocampal neuronal culture coexpressing barcodes and GFP (cyan). All GFP positive neurons were filled with barcode amplicons, indicating efficient somatic barcode amplification. (D) Images of the first six sequencing cycle of the same neurons shown in (C). The bases corresponding to the four colors and the sequences of the three neurons circled in (D) are indicated to the left of the images in (D). In all images, scale bars = 50 µm.

Fig. 2. *In situ* barcode sequencing in brain slices using BaristaSeq. (A) Representative low-magnification images of a barcoded brain slice expressing GFP (*left*), rolonies generated in the same brain slice (*middle*), and the first sequencing cycle of the brain slice (*right*). In all images, scale bars = 100 µm. (B) Representative low-resolution images of the indicated cycles of barcode sequencing in a brain slice. The sequences of the three cells basecalled are indicated below the images. Only bases corresponding to the images shown are capitalized and color-coded. Scale bars = 100 µm. (C) The quality of the base calls over 25 cycles of Illumina sequencing *in situ* on the barcoded brain slice. The quality score is defined as the intensity of the called channel divided by the root sum square of all four channels. A quality score of 1 (best) indicates sequencing signal in only one channel, and a score of 0.5 (worst) indicate same intensity across all four channels. (D) Histogram of the number of mismatches between the *in situ* reads and their closest matches from *in vitro* reads (*in situ*) and the number of mismatches between random sequences and their closest matches from *in vitro* reads (Random). (E) An example barcode read *in situ* and its closest match *in vitro*, and a random sequence and its closest match *in vitro*. Red indicates mismatches.

Fig. 3. Interrogation of the projection neurons of the mouse auditory cortex using BARseq. (A) Comparison of BARseq to CTB retrograde labeling. (B) Venn diagram showing the number of GFP expressing neurons labeled with (magenta) or without (white) CTB and/or neurons found to project contralaterally using BARseq (cyan). (C) Conventional bulk GFP tracing intensities were plotted against the bulk projection strength obtained from MAPseq. Error bars indicate SEM. N = 5 for GFP tracing and N = 3 for MAPseq. Pearson correlation coefficient $r = 0.94$, $p < 0.0001$. (D) The distribution of projection intensity in each projection area. The y-axis indicates the logarithms of raw barcode counts in each area, and the x-axis indicate the number of cells. (E) Histogram of the number of projections per neuron. (F) The mean projection patterns of nodes corresponding to major classes of neurons. Line thickness indicates projection strength

normalized to the strongest projection for that class. Blue arrows indicate projections to contralateral brain areas and black arrows indicate projections to ipsilateral brain areas. The corresponding class and node numbers are indicated below each projection pattern. CT: corticothalamic neurons, PT: pyramidal tract neurons, ITc: contralaterally projecting intratelencephalic neurons, ITi: ipsilaterally projecting intratelencephalic neurons. (G) Hierarchical clustering of single-cell projection data. Left: dendrogram of the hierarchical structure of the clusters. An index is assigned to each node/leaf as indicated. Right: the mean projection patterns sorted by cluster identity. Each row represents a leaf cluster. The number of neurons in each cluster is indicated on the right. Each column represents projection to the indicated area. OB: olfactory bulb; OFC: orbitofrontal cortex; Motor: motor cortex; Rstr: rostral striatum; SSctx: somatosensory cortex; Cstr: caudal striatum; Amyg: amygdala; VisIp: ipsilateral visual cortex; VisC: contralateral visual cortex; AudC: contralateral auditory cortex; Thal: thalamus; Tect: tectum. (H) t-SNE plot of the projection neurons using the projection data as inputs. The neurons are color-coded by their first level subclass identities post-hoc.

Fig. 4. Laminar distribution of projection neurons in the mouse auditory cortex. (A) The sequenced projection neurons from a brain (XC9) are color-coded by class identities and plotted at their locations in the cortex. The top and bottom of the cortex are indicated by the red and blue dashed lines, respectively. The lamina and their boundaries are marked. Scale bar = 100 µm. Inset: histograms of each class of projection neurons in the pooled dataset of XC9 and XC28. The y-axis indicates cortical depth and the x-axis indicates the number of neurons at that depth in a particular class. Laminar layers are indicated on the right and the boundaries between two layers are marked by dashed lines. (B) The normalized entropy of nodes/leaves (y-axis) in the indicated clustering hierarchy (x-axis). The branch of the subclass nodes/leaves were color coded as indicated. Grey bars indicate mean ± stdev of all nodes/leaves of a specific hierarchy. Hierarchy 1-3 correspond to class divisions and hierarchy 4-7 correspond to subclass divisions. (C) The laminar distribution of all subclasses. Individual neurons that were well classified (red) or ambivalently classified (blue) are superimposed on top of the distribution plots (black). The subclasses are arranged in the same order as in Fig. 3G. The class to which each subclass belongs is indicated below.

References

1 Marblestone, A. H., Daugharthy, E.R., Kalhor,R., Peikon, I.D., Kebschull, J.M., Shipman, S.L., Mishchenko, Y., Lee, J.H., Kording, K.P., Boyden, E.S., Zador, A.M., Church, G.M. Rosetta Brains: A Strategy for Molecularly-Annotated Connectomics. **arXiv** (2014).

2 Sorensen, S. A. *et al.* Correlated gene expression and target specificity demonstrate excitatory projection neuron diversity. *Cereb Cortex* **25**, 433-449, doi:10.1093/cercor/bht243 (2015).

3 Bock, D. D. *et al.* Network anatomy and in vivo physiology of visual cortical neurons. *Nature* **471**, 177-182, doi:10.1038/nature09802 (2011).

4 Han, Y., Kebschull, J.M., Campbell, R.A.A., Cowan, D., Imhof, F., Zador, A.M., Mrsic-Flogel, T.D. A single-cell anatomical blueprint for intracortical information transfer from primary visual cortex. *Biorxiv* (2017).

5 Kebschull, J. M. *et al.* High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. *Neuron* **91**, 975-987, doi:10.1016/j.neuron.2016.07.036 (2016).

6 Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360-1363, doi:10.1126/science.1250212 (2014).

7 Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* **10**, 857-860, doi:10.1038/nmeth.2563 (2013).

8 Chen, X., Sun, Y. C., Church, G. M., Lee, J. H. & Zador, A. M. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Res*, doi:10.1093/nar/gkx1206 (2017).

9 Harris, K. D. & Shepherd, G. M. The neocortical circuit: themes and variations. *Nat Neurosci* **18**, 170-181, doi:10.1038/nn.3917 (2015).

10 Shepherd, G. M. Corticostriatal connectivity and its role in disease. *Nat Rev Neurosci* **14**, 278-291, doi:10.1038/nrn3469 (2013).

11 Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138-1142, doi:10.1126/science.aaa1934 (2015).

12 Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* **19**, 335-346, doi:10.1038/nn.4216 (2016).

13 Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *bioRxiv*, doi:10.1101/229542 (2017).

14 Paul, A. *et al.* Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity. *Cell* **171**, 522-539 e520, doi:10.1016/j.cell.2017.08.032 (2017).

15 Kebschull, J. M., Garcia da Silva, P. & Zador, A. M. A New Defective Helper RNA to Produce Recombinant Sindbis Virus that Infects Neurons but does not Propagate. *Front Neuroanat* **10**, 56, doi:10.3389/fnana.2016.00056 (2016).

16 Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207-214, doi:10.1038/nature13186 (2014).

17 Znamenskiy, P. & Zador, A. M. Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature* **497**, 482-485, doi:10.1038/nature12077 (2013).

18 Custo Greig, L. F., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat Rev Neurosci* **14**, 755-769, doi:10.1038/nrn3586 (2013).

19 Economo, M. *et al.* Distinct descending motor cortex pathways and their roles in movement. *bioRxiv*, doi:10.1101/229260 (2017).

20 Slater, B. J., Willis, A. M. & Llano, D. A. Evidence for layer-specific differences in auditory corticocollicular neurons. *Neuroscience* **229**, 144-154, doi:10.1016/j.neuroscience.2012.10.053 (2013).

21 Cadwell, C. R. *et al.* Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat Biotechnol* **34**, 199-203, doi:10.1038/nbt.3445 (2016).

22 Economo, M. N. *et al.* A platform for brain-wide imaging and reconstruction of individual neurons. *Elife* **5**, e10566, doi:10.7554/eLife.10566 (2016).

23 Gerfen, C. R., Economo, M. N. & Chandrashekar, J. Long distance projections of cortical pyramidal neurons. *J Neurosci Res*, doi:10.1002/jnr.23978 (2016).

24 Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788-791, doi:10.1038/44565 (1999).

25 Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008, doi:10.1088/1742-5468/2008/10/p10008 (2008).

26 von Luxburg, U. A tutorial on spectral clustering. *Stat Comput* **17**, 395-416, doi:10.1007/s11222-007-9033-z (2007).

27 Ng, A. Y., Jordan, M. I. & Weiss, Y. On spectral clustering: Analysis and an algorithm. *Adv Neur In* **14**, 849-856 (2002).

28 van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).

29 Linden, J. F. & Schreiner, C. E. Columnar transformations in auditory cortex? A comparison to visual and somatosensory cortices. *Cereb Cortex* **13**, 83-89 (2003).

30 Oviedo, H. V., Bureau, I., Svoboda, K. & Zador, A. M. The functional asymmetry of auditory cortex is reflected in the organization of local cortical circuits. *Nat Neurosci* **13**, 1413-1420, doi:10.1038/nn.2659 (2010).

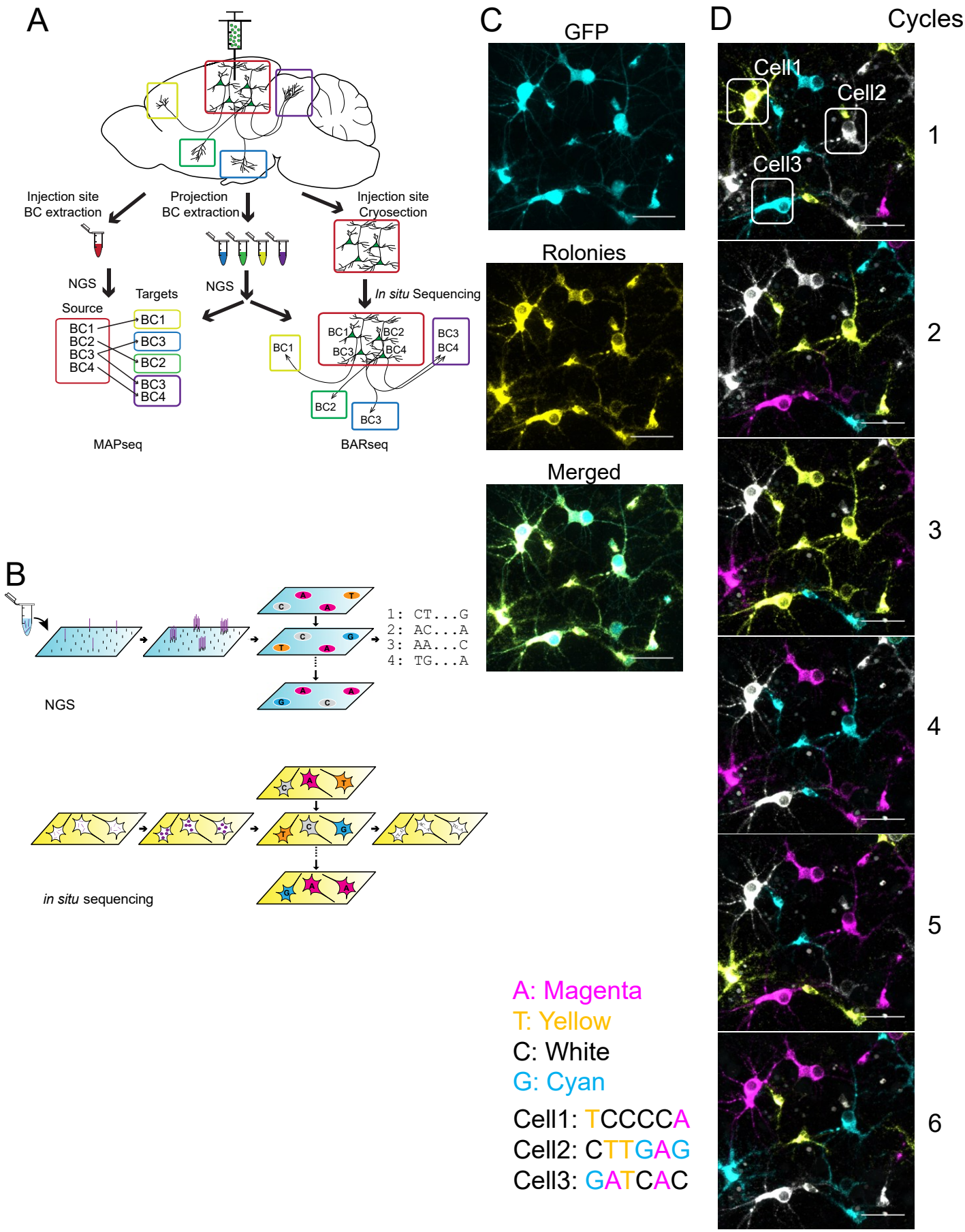
A
Injection site
BC extraction
Projection
BC extraction
Injection site
Cryosection
NGS
Targets
Source
BC1
BC2
BC3
BC4
BC1
BC3
BC2
BC3
BC4
NGS
In situ Sequencing
BC1
BC1
BC2
BC3
BC4
BC3
BC4
BC2
BC3
MAPseq
BARseq
B
NGS
1: CT...G
2: AC...A
3: AA...C
4: TG...A
in situ sequencing
C
GFP
Rolonies
Merged
D
Cycles
Cell1
Cell2
Cell3
1
2
3
4
5
6
A: Magenta
T: Yellow
C: White
G: Cyan
Cell1: TCCCCA
Cell2: CTTGAG
Cell3: GATCAC

Figure 1

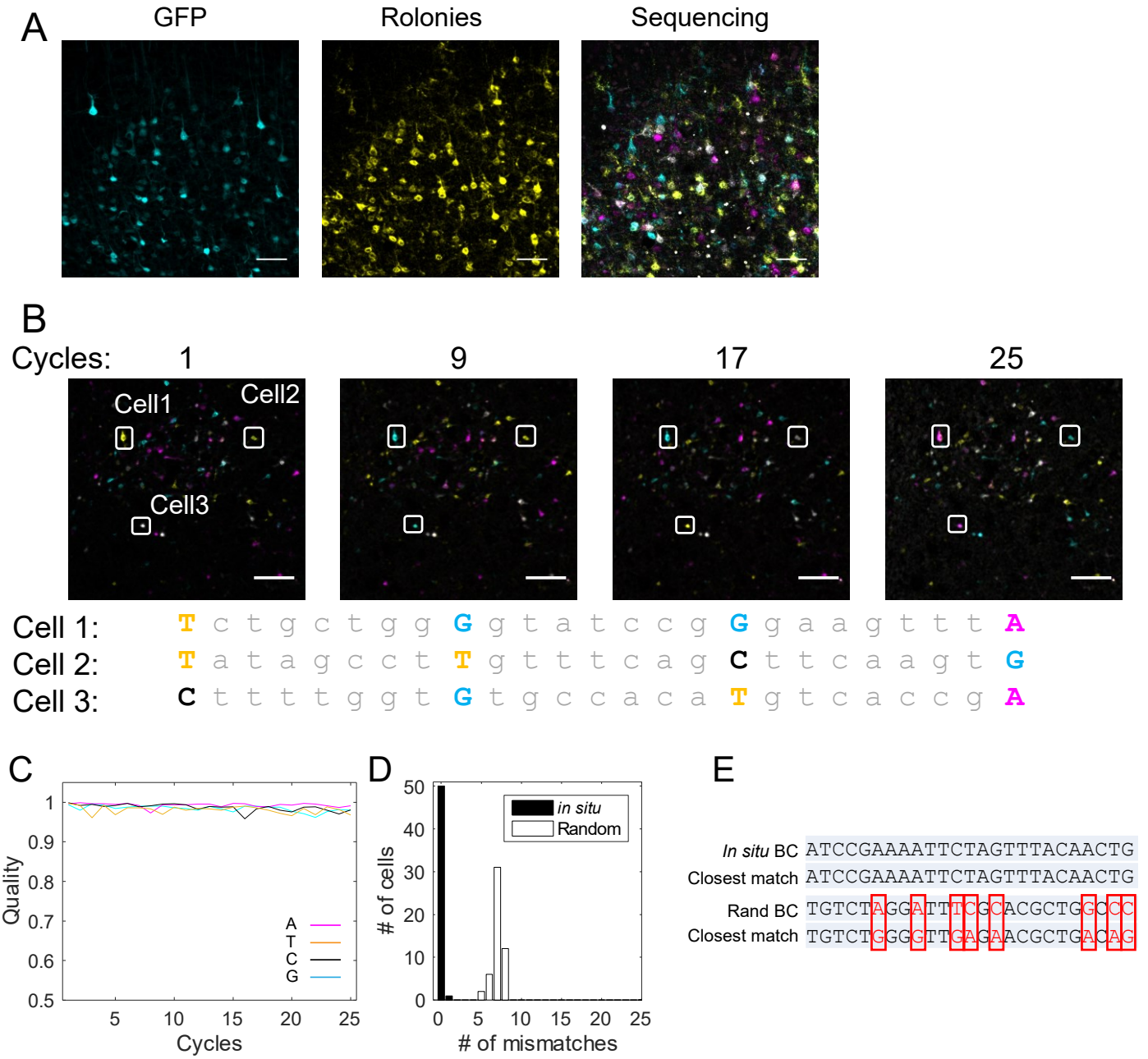
A
GFP
Rolonies
Sequencing
B
Cycles:
1
9
17
25
Cell1
Cell2
Cell3
Cell 1: T c t g c t g g G g t a t c c g G g a a g t t t A
Cell 2: T a t a g c c t T g t t t c a g C t t c a a g t G
Cell 3: C t t t t g g t G t g c c a c a T g t c a c c g A
C
Quality
Cycles
A
T
C
G
D
# of cells
# of mismatches
in situ
Random
E
In situ BC ATCCGAAAATTCTAGTTTACAACTG
Closest match ATCCGAAAATTCTAGTTTACAACTG
Rand BC TGTCTAGGATTTCGCACGCTGGCCC
Closest match TGTCTGGGGTTGAGAACGCTGACAG

Figure 2

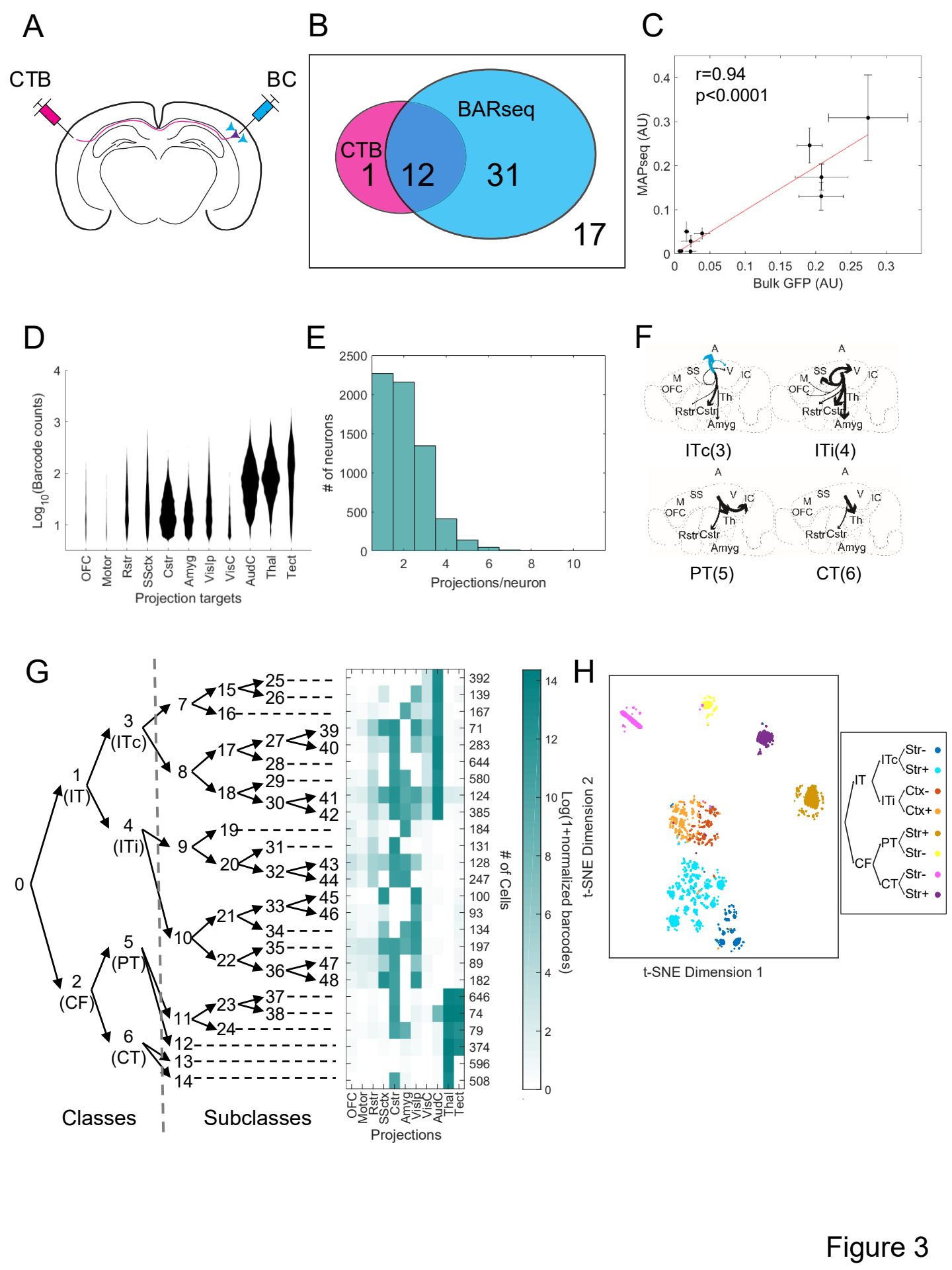
A
CTB
BC
B
CTB
BARseq
1
12
31
17
C
r=0.94
p<0.0001
MAPseq (AU)
Bulk GFP (AU)
D
Log10(Barcode counts)
OFC
Motor
Rstr
SSctx
Cstr
Amyg
VisIp
VisC
AudC
Thal
Tect
Projection targets
E
# of neurons
Projections/neuron
F
ITc(3)
ITi(4)
PT(5)
CT(6)
G
0
1 (IT)
2 (CF)
3 (ITc)
4 (ITi)
5 (PT)
6 (CT)
Classes
Subclasses
Projections
# of Cells
Log(1+normalized barcodes)
H
t-SNE Dimension 1
t-SNE Dimension 2
IT
ITc
ITi
CF
PT
CT
Str-
Str+
Ctx-
Ctx+

Figure 3

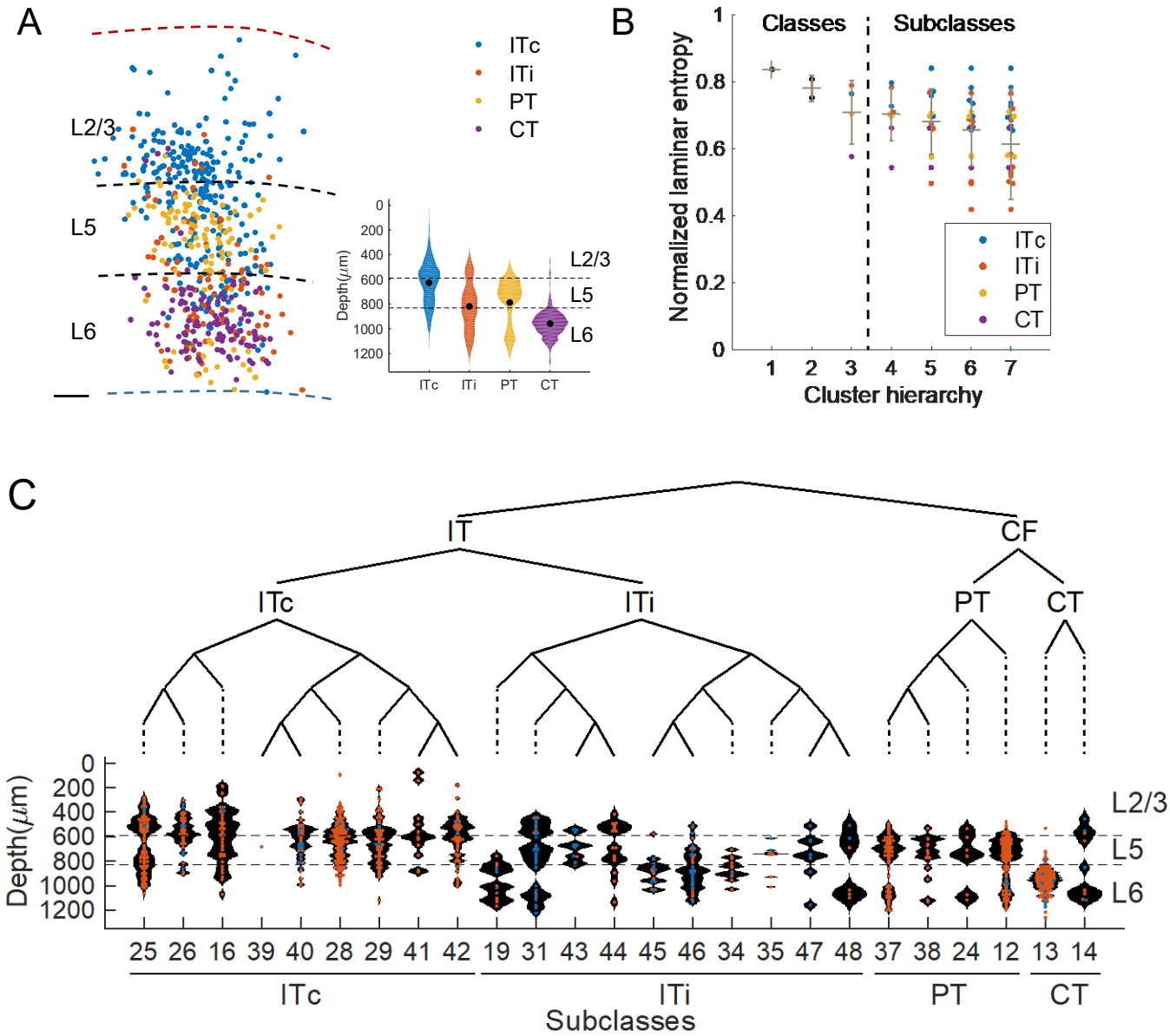
A
L2/3
L5
L6
ITc
ITi
PT
CT
Depth(μm)
0
200
400
600
800
1000
1200
B
Classes
Subclasses
Normalized laminar entropy
Cluster hierarchy
C
IT
CF
ITc
ITi
PT
CT
25 26 16 39 40 28 29 41 42 19 31 43 44 45 46 34 35 47 48 37 38 24 12 13 14
Subclasses

Figure 4