High-throughput mapping of long-range neuronal projection using *in situ* 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 connectivity, spatial organization, gene expression, and other 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 RNA barcoding. 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, as well as further subdivisions within each class. We further combined BARseq with *in situ* detection of mRNA to relate neuronal gene expression with projections. By bridging high-throughput neuroanatomy with high spatial resolution, BARseq provides a path towards a systematic multi-modal description of the nervous system.

Introduction

The nervous system consists of networks of neurons, organized into areas, nuclei, lattices, laminae and other structures. Within these structures are uncounted different neuronal types, each characterized by its own pattern of connections, gene expression and other properties. The spatial organization of neuronal types generally reflects the functional compartmentalization of neural circuits. Understanding how diverse neurons are organized thus requires methods that can map various neuronal characteristics to their locations at cellular resolution, with high-throughput, in single brains.

A particular challenge, unique to neuroscience, is the need to map the long-range axonal projection patterns that form the physical basis for neuronal circuits across brain areas. Traditional neuroanatomical methods based on microscopy can be used to visualize neuronal morphology, including long-range projections, but the throughput of these methods remains low despite recent advances in methodology (up to ~50 neurons per single cortical area) (Economo et al., 2016; Gong et al., 2016; Guo et al., 2017; Zheng et al., 2013). Furthermore, these methods usually rely on dedicated imaging platforms specifically designed for neuronal tracing, which are not readily accessible to non-neuroanatomy laboratories.

To allow high throughput projection mapping, we previously developed MAPseq (Han et al., 2018; Kebschull et al., 2016a) (Fig. 1A, left), a sequencing-based method capable of mapping projections of thousands of single neurons in a single brain. MAPseq achieves multiplexing by uniquely labeling individual neurons with random RNA sequences, or "barcodes." However, like most other sequencing methods, the original MAPseq protocol relies on tissue homogenization, so it cannot register the mapped neurons to specific locations in the brain. This lack of spatial information obscures the spatial organization of projection neurons, and further prevents association of projection patterns with other neuronal properties *in situ*.

Here we describe BARseq (Barcoded Anatomy Resolved by Sequencing; Fig. 1A, right), a method that combines MAPseq with *in situ* sequencing (Chen et al., 2018; Ke et al., 2013; Lee et al., 2014) of cellular barcodes. BARseq preserves the spatial organization of neurons during projection mapping, and further allows co-registration of neuronal projections with mRNA expression through space in a single specimen. We show that BARseq is able to recapitulate the known spatial organization of major classes of neocortical excitatory neurons—corticothalamic (CT), pyramidal tract (PT), and intratelencephalic (IT) neurons (Harris and Shepherd, 2015; Shepherd, 2013)—in the mouse auditory cortex, and additionally to uncover previously unknown organization beyond these major classes. BARseq thereby complements traditional neuro-anatomical method by providing a high-throughput method for linking axonal projections, spatial organization, gene expression, and potentially other neuronal properties that define circuits.

Results

In what follows, we first demonstrate robust *in situ* sequencing of cellular barcodes in neuronal culture. We then extend this approach to achieve efficient sequencing of barcodes in cortical slices. We next combine *in situ* sequencing with MAPseq to create BARseq, and characterize the performance of BARseq. We show that BARseq can be combined with *in situ* hybridization to correlate neuronal projections with gene expression. Finally, we apply BARseq to projection neurons in the auditory cortex, and identify a structured organization of projections beyond known classes of projection neurons.

In situ barcode sequencing in neuronal culture.

We set out to develop a method capable of reading out highly diverse ensembles of sequences *in situ*, with cellular and even subcellular spatial resolution. Although several highly sensitive multiplexed *in situ* hybridization methods have been described [e.g., merFISH (Chen et al., 2015), seqFISH (Shah et al., 2016), and STARmap (Wang et al., 2018)], these methods lack single-nucleotide specificity, and can distinguish only up to 10^2-10^3 different transcript sequences, each with a length of 10^2-10^3 nucleotides. By contrast, cellular barcodes used for projection mapping are much shorter (30-nt), more diverse (~10^6 to 10^7), and may differ from each other by just a few nucleotides. Cellular barcodes thus cannot be detected adequately using currently available multiplexed *in situ* hybridization approaches. Spatial transcriptomics (Stahl et al., 2016) can read out diverse short barcodes, but at present the spatial resolution (100-200 um) is insufficient to resolve single neurons. Laser micro-dissection combined with RNA sequencing can provide cellular resolution, but the throughput is too low for high-throughput projection mapping. We therefore focused on an *in situ* sequencing approach (Chen et al., 2018; Ke et al., 2013) to achieve both the spatial resolution and the throughput needed for barcode identification *in situ.*

In situ sequencing is conceptually similar to conventional *in vitro* Illumina DNA sequencing (Fig. 1B), both of which consist of two basic steps: amplification and sequencing. In the amplification step, the RNA is converted to cDNA by reverse transcription, and the cDNA is amplified using rolling circle amplification, resulting in the formation of a small (< 1 µm) nanoball of DNA called a "rolony." In the sequencing step, the rolonies are sequenced in parallel using four fluorescently-labeled nucleotides. The nucleotide sequences are thus transformed into color sequences, and are read out using multi-channel fluorescence microscopy.

To capture short diverse barcode sequences, we needed an approach in which the actual target barcode sequence is captured and amplified (Fig. 1Ca-f). Although some *in situ* hybridization methods use sequencing to read out gene-specific tags to allow multiplexed detection of RNAs (Ke et al., 2013; Wang et al., 2018), they cannot directly sequence the RNAs of interest and are unsuitable for barcode sequencing (Fig. 1Ccd). Circligase-based FISSEQ (Fig. 1Ce) can readout mRNA sequences, but is not sufficiently sensitive for barcode sequencing. We therefore used a targeted approach (Chen et al., 2018; Ke et al., 2013) in which reverse transcription was used to convert the barcode sequence into cDNA, after which padlock probe hybridization followed by gap-filling and ligation was used to form a circular template for rolling circle amplification (Fig. 1Cf). The gap-filling padlock approach allowed for efficient amplification of barcodes in neurons co-expressing barcodes and GFP (Fig. 1D). We also adapted Illumina Sequencing By

Synthesis (SBS) chemistry for *in situ* reactions. Taken together, these formed the basis for a protocol that allowed efficient and robust sequencing of barcoded cultured neurons (Fig. 1E).

In situ barcode sequencing in brain slices

In situ sequencing in brain slices presents additional challenges compared to sequencing in cultured neurons. We developed a chamber system which enabled convenient handling and was compatible with the amplification reactions (Fig. S1A; see STAR Methods). Because barcoded neurons express high levels of GFP (which interferes with imaging during sequencing), we developed a protocol for tissue digestion to reduce the GFP signal from barcoded cells (Fig. S1B; see STAR Methods) and to increase RNA accessibility. Together these optimizations resulted in highly efficient amplification of barcodes (Fig. S1C) compared to other methods (Ke et al., 2013; Lee et al., 2014). We also optimized Illumina sequencing chemistry for brain slices, resulting in a ~10-fold improvement in signal-to-noise ratio (SNR) compared to sequencing by ligation approaches (Ke et al., 2013; Lee et al., 2014) (Fig. S1D, E; see STAR Methods). These modifications resulted in highly efficient and specific barcode amplification (Fig. 2A) and sequencing (Fig. 2B) in barcoded neuronal somata in brain slices.

To evaluate barcode sequencing in brain slices, we sequenced 25 bases in a sample infected with a diverse ($>10^6$ unique sequences) pool of barcoded Sindbis virus (Kebschull et al., 2016a; Kebschull et al., 2016b) injected in the auditory cortex (Fig. 2C). Basecall quality (Fig. 2D; Fig. S1F; see STAR Methods) and signal intensity (Fig. S1G) remained high through all 25 cycles. We observed no bias toward any particular base (Fig. S1H). 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. 2E, F). The per-base error rate of $1 / (25 \times 51) = 0.0078$ corresponds to a Phred score of 31. Because the barcodes in this library represent a small fraction ($\sim 10^{-9}$) 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.

Projection mapping using BARseq

We next combined *in situ* barcode sequencing with MAPseq to create BARseq. In MAPseq (Fig. 1A, *left*), neurons are barcoded using a Sindbis virus library. After 44 hours, both the source area containing neuronal somata and target projection areas are micro-dissected into “cubelets” and sequenced. Barcodes from the target areas are then matched to those at the source area to reveal projection patterns. The spatial resolution of MAPseq is thus determined by the size of the cubelets. In BARseq, we perform *in situ* sequencing of barcoded somata at the source; the target projection areas are still micro-dissected and sequenced as cubelets (Fig. 1A, *right*). BARseq thus inherits the throughput and cubelet resolution of projections of MAPseq, but allows the precise origin of each axonal projection to be determined with cellular spatial resolution.

The high sensitivity and accuracy of MAPseq is well established (Han et al., 2018; Kebschull et al., 2016a; Klingler et al., 2018). In particular, the sensitivity of MAPseq is indistinguishable from that of conventional GFP-based single neuron reconstruction (Han et al., 2018). To confirm

that BARseq maintains the high sensitivity and accuracy of MAPseq, we compared it to conventional retrograde cholera toxin subunit B (CTB) tracing of contralateral projections. We injected a highly diverse (~10^7 barcodes) viral library into the auditory cortex and CTB in the contralateral auditory cortex (Fig. 3A, B). We used BARseq to identify contralaterally projecting neurons and the precise locations of their somata. We then identified barcoded somata that were also labeled with CTB. 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. 3C; Table S1). The strength of the contralateral projections was well above the noise threshold (defined by barcode counts in the olfactory bulb, an area to which the auditory cortex does not project) (Fig. 3D), 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 previously reported (Kebschull et al., 2016a) high specificity and sensitivity of MAPseq (91 ± 6%), and may exceed the sensitivity of conventional CTB tracing.

Relating projection patterns to gene expression using BARseq

The *in situ* nature of BARseq allows us to relate axonal projection patterns with other single neuron characteristics measured *in situ* in the same sample. In particular, BARseq can be combined with *in situ* detection of gene expression. Simultaneous high-throughput measurement of gene expression and projection pattern could help establish the relationship between projections and transcriptomic cell types. Such approaches are particularly appealing in light of recent developments in multiplexed *in situ* hybridization (Chen et al., 2015; Shah et al., 2016; Wang et al., 2018) and *in situ* sequencing techniques (Lee et al., 2014), including the non-gapfilling version of the *in situ* sequencing technique used in BARseq (Fig. S2).

To demonstrate the feasibility of resolving both barcodes (for projection mapping) and endogenous gene expression in single neurons, we combined BARseq with fluorescent *in situ* hybridization (FISH) using probes against Slc17a7 [a marker of excitatory neurons; (Tasic et al., 2016)] and Gad2 (a marker of inhibitory neurons) in the same cells (Fig. 4A-C). We sampled 100 barcoded cells with high sequencing quality; of these, 80 were excitatory neurons (Slc17a7+ and Gad2-), 3 were inhibitory neurons (Slc17a7- and Gad2+), and 17 were non-neuronal cells (Slc17a7- and Gad2-). The ratio between inhibitory and excitatory neurons was not significantly different from previous estimates [~11.5% (Meyer et al., 2011), $p = 0.13$ using Fisher's exact test]. Out of these 100 cells, 52% (52/100) projected to at least one of the sampled areas with strengths above the noise floor defined by the olfactory bulb (Table S2). Consistent with the fact that most projection neurons in the cortex are excitatory, all 52 projection neurons identified by BARseq expressed Slc17a7. The non-projecting excitatory neurons likely projected locally or to nearby cortical areas we did not sample. The projection neurons identified by BARseq are thus consistent with their transcriptomic cell types. These experiments demonstrate that projection mapping with BARseq is compatible with *in situ* interrogation of gene expression, and could readily be extended to relate projection patterns with detection of expression of dozens or hundreds of individual genes (Fig. S2).

Projection diversity in the mouse auditory cortex

Having established the sensitivity and specificity of BARseq, we applied it to study the organization of long-range projections from the mouse auditory cortex. We performed BARseq on two brains, and MAPseq on an additional brain. We allowed one mismatch when matching *in situ* barcodes to the projection barcodes (see STAR Methods; Fig. 5A), and found 1806 (55%) of 3237 cells sequenced *in situ* projected to at least one target area (Table S3). We further excluded cells obtained from tissue deformed during processing, and from labeled cells outside of the cortex. This resulted in 1309 neurons sequenced *in situ* and 5082 neurons sequenced *in vitro* using conventional MAPseq (Table S3; Fig. 5B). Because only neurons at the center of the injection sites were sequenced in each BARseq brain, the MAPseq experiment produced more neurons per brain than BARseq, but might include neurons from nearby cortical areas.

We focused on 11 auditory cortex projection target areas, including four ipsilateral cortical areas (orbitofrontal, motor, somatosensory, and visual), two contralateral cortical areas (visual and auditory), three subcortical telencephalic areas (amygdala, rostral striatum, and caudal striatum), the thalamus, and the tectum (see STAR Methods for details of dissected areas). These included all major brain areas to which the auditory cortex projects, as determined by conventional bulk GFP tracing experiments (Oh et al., 2014). We also collected tissue from the olfactory bulb, an area to which the auditory cortex does not project, as a negative control.

Only five out of 6391 neurons had non-zero barcode counts in the olfactory bulb (four neurons with a count of one molecule, and one neuron with a count of four molecules), indicating that the technical false-positive rate for a single target area is extremely low at 5 / 6391 = 0.08%. Using the false-positive rate and the number of barcodes with zero counts in each area, we estimated that a total of 40 out of 18851 projections detected in our dataset might be false positives, corresponding to a false-discovery rate of 0.2% (i.e., for each 1000 detected projections, two are likely false positives). BARseq thus has a very low false discovery rate.

MAPseq has now been validated using several different methods, including single neuron reconstruction, in multiple brain circuits (Han et al., 2018; Kebschull et al., 2016a). The contribution of potential artifacts, including those due to non-uniform barcode transport and variable barcode expression strength, have been carefully quantified in previous work (Han et al., 2018; Kebschull et al., 2016a). Although one potential challenge of MAPseq arises from fibers of passage, in practice these have not represented a major source of artifact (Han et al., 2018; Kebschull et al., 2016a). One reason is that the total amount of axonal material due to fibers of passage is typically small compared with the rich innervation of a fiber at its target. A second reason is that the barcode carrier protein is designed to be enriched at synapses (Kebschull et al., 2016a), which further minimizes the contribution of passing fibers to barcode counts. To determine the extent of contamination from fibers of passage, we examined the striatal projections of putative CT neurons (i.e. neurons that project to the thalamus, but not to any cortical area or the tectum) and putative PT neurons (i.e. neurons that project to the tectum, the thalamus, but not to any cortical area) from the two BARseq brains (the MAPseq brain was excluded because it may include PT neurons from neighboring areas that did not project to the tectum, and may be misidentified as CT neurons). Both types of neurons send descending fibers through the internal capsule adjacent to the caudal striatum, and thus can be used to evaluate fibers of passage contamination in the striatum. Consistent with the fact that PT neurons, but not CT neurons project to the striatum, only 8 / 250 (3.2%) putative CT neurons sequenced *in situ* projected to the caudal striatum, whereas 59 / 245 (24%) putative PT neurons sequenced *in situ* projected to the caudal striatum. The fact that ~97% of putative CT neurons did not show

projection to the caudal striatum suggests that fibers of passage had minimal impact on projections mapped in our experiments.

At the mesoscopic level, the projection patterns of the auditory cortex revealed by BARseq were consistent with anterograde tracing in the Allen Mouse Connectivity Atlas (Oh et al., 2014). In MAPseq and BARseq, the strength of the projection of a neuron to a target is given by the number of individual barcode molecules from that neuron recovered at the target. This is analogous to a conventional fluorophore mapping experiment, in which projection strength is assumed to be proportional to GFP intensity. Consistent with conventional bulk GFP tracing (Oh et al., 2014) (Fig. 5C; Pearson correlation coefficient $r = 0.94$, $p < 0.0001$), projections from the auditory cortex to the thalamus, the tectum, the contralateral auditory cortex, and the caudal striatum were particularly strong (Fig. 5D). BARseq thus provides quantitative measurements of long-range projections in the auditory cortex consistent with those obtained by conventional bulk labeling techniques.

The projection patterns of individual neurons were remarkably diverse. To quantify this diversity, we binarized projection patterns, using a conservative threshold for projection detection (four molecules per barcode per area, given by the OB negative control). We observed 264 distinct patterns, or 13% of the $2^{11} = 2048$ possible patterns. This high diversity was unlikely an artifact of projections missed by BARseq due to false negatives: 237 to 247 unique projection patterns remain even when a false negative rate of 10-15% was assumed (see STAR Methods). This high diversity was also unlikely to be caused by limited sensitivity of MAPseq for neurons with fewer barcodes overall: A significant fraction of these projection patterns (112 / 264 = 42%) can be accounted for solely by strong projections (see STAR Methods), and eliminating neurons with weaker primary projections resulted in only a moderate reduction in the diversity of projection patterns (Fig. 5E, blue line). Furthermore, this reduction was largely due to the reduction in sample size (Fig. 5E, red line). Because of the conservative thresholding we used and the fact that the auditory cortex may project to areas we did not sample, the actual number of distinct projection patterns may be higher.

The majority (85%) of cortical projection neurons projected to two or more target areas (Fig. 5F) after binarization. The ratio of the projection strengths of the two strongest projections follow a log normal distribution (Fig. 5G), and about half of all neurons have a secondary projection that is at least 20% as strong as their primary projections. These estimates were based on neurons whose primary projections were strong enough to observe a secondary projection, and therefore were not limited by the sensitivity of MAPseq/BARseq or by the variability in the expression level of barcodes. This high projection diversity and multiple projections per neuron are consistent with findings in the visual cortex (Han et al., 2018), and may be a general feature of cortical projections.

BARseq recovers known laminar organization of auditory cortical projection neuron classes

Simple hierarchical clustering on the non-binarized projections (see STAR Methods) partitioned neurons into established (Harris and Shepherd, 2015) cortical classes: CT (corticothalamic), PT (pyramidal tract), and IT (intratelencephalic, Fig. 6A). The CT neurons project to the thalamus; the PT neurons project to both the thalamus and the tectum; the IT neurons project to the cortex but not subcortically to the thalamus or tectum. IT neurons were further subdivided into those

that project to both the ipsilateral and contralateral cortex (the ITc class), and those that project ipsilaterally only (the ITi class). These patterns correspond to the classically-defined top-level subdivisions of cortical projection patterns.

Interestingly, although IT neurons projected to both the caudal striatum—an area implicated in auditory decision making (Znamenskiy and Zador, 2013)—and the rostral striatum, PT neurons projected only to the caudal striatum. We confirmed this finding by triple retrograde tracing (Fig. 6B): A significant fraction of neurons (66/296) projecting to the rostral striatum also projected to the caudal striatum, but none (0/296) projected to the tectum (Fig. 6C, D). The results of BARseq were thus consistent with, but went beyond, classical taxonomy.

The classification of cortical excitatory projection patterns into PT, IT and CT classes was supported by the corresponding segregation of these neurons into distinct laminae. As expected, the major classes segregated strongly with laminar position: superficial layers contained predominantly IT neurons, whereas deep layers contained predominantly PT and CT neurons (Harris and Shepherd, 2015) (Fig. 6E; see STAR Methods and Fig. S3A-D for the estimation of layers using layer-specific marker genes). Thus major classes of cortical projection neurons defined by BARseq had distinct laminar distributions, consistent with observations using conventional methods (Custo Greig et al., 2013).

BARseq uncovers diversity within class

Even within the major classes, projections from the auditory cortex to individual targets were not statistically independent (Fig. 7A). For example, a PT neuron with a strong thalamic projection was also likely to have a strong tectal projection, whereas a PT neuron with a strong thalamic projection would likely have a weak amygdalar projection (Fig 7Ac). This lack of independence suggests that the remarkable diversity of projections—264 distinct binary patterns—did not arise by a process in which each neuron selected targets purely at random, and that even within each class there is further statistical structure.

To explore the statistical structure underlying these projection patterns, we performed hierarchal clustering on the non-binarized projection patterns. This clustering uncovered structure well beyond the previously established top-level classes (Fig. 7B; see Fig. S4A-F and STAR Methods for details of clustering methods). Depending on the precise clustering algorithm and statistical criteria used, as many as 25 subclusters were revealed. This structured organization was reminiscent of the transcriptomic landscape of cortical neurons (Saunders et al., 2018; Tasic et al., 2016; Tasic et al., 2017; Zeisel et al., 2015).

We used a probabilistic approach to assess how well each neuron was classified (see STAR methods; Fig. S5A). This approach assigned 5682/6391 (89%) of neurons unambiguously into one of the 25 subclusters (Fig. 7B, C). Based on the fraction of unambiguously classified neurons, we further labeled 19 of the 25 subclusters as “core subclusters” (i.e. containing at least 80% well-classified neurons) and 6 subclusters as “intermediate subclusters” (i.e. containing less than 80% well-classified neurons, Fig. 7D). These 25 subclusters included nine subclusters of ITc neurons, ten subclusters of ITi neurons, five subclusters of PT neurons, and one subclusters of CT neurons (subcluster 14 split from the CT class was likely a PT subcluster based on projection pattern and laminar positions; see STAR Methods). All subclusters of PT and CT neurons, and all but one ITc subclusters, were core subclusters, whereas only half of ITi

subclusters were core. Other clustering methods confirmed similar clusters (Fig. S5B; see STAR Methods), despite fine differences in cluster boundaries and hierarchies. Because this remarkable diversity was obtained using only the mesoscopic projection resolution afforded by MAPseq, finer resolution in mapping projections might reveal even more subclusters; for example, L6 IT neurons mainly project to the deep layers of the cortex (Tasic et al., 2017). These results indicate that BARseq identified organization in projections well beyond the previously described major divisions of projection neuron types.

Laminar organization of projections within class

As noted above, the three major cortical classes—PT, IT and CT—are spatially segregated into laminae. By contrast, the subclusters partitioned according to projection pattern did not result in further refinement in laminar distribution (Fig. 7E; $p = 0.95$ for cluster hierarchy 2-7, Kurskal-Wallis test; Fig. S6A, B; see STAR Methods). Most subclusters spanned many layers (Fig. 7F). Notably, no IT subcluster was restricted to L2/3, despite the existence of L2/3-specific subclasses defined by gene expression (Tasic et al., 2017). The laminar distribution of subclusters was thus not generally more restricted than the classes from which they were derived.

Our failure to observe layer-specific projection subclusters could accurately reflect true independence of laminar position and projection pattern among neurons in different subclusters. Alternatively, it could be that laminar position was correlated with projection pattern, but these correlations were not strong enough to drive the formation of cluster boundaries. To distinguish these possibilities, we tested whether we could predict laminar position from projection pattern; if subclusters were truly independent of laminar position, knowledge of the projection pattern would have no predictive power for estimating laminar position. For both ITc and PT classes, a random forest classifier could consistently predict the laminar location above chance level using the raw projection data (Fig. 7G; $p < 5\times10^{-70}$ after Bonferroni correction; Fig. S6C, D). However, significant overlap in the distributions of strengths among layers (Fig. S7A-C) suggest a minor effect of laminae on projection strengths. We further analyzed PT neurons in layer 5 and layer 6. The PT neurons of the auditory cortex, unlike those in other cortical areas, reside in both layer 5 and layer 6, with distinct morphology and physiological properties (Slater et al., 2013). Consistent with these differences in other neuronal properties, the corticotectal projections were slightly stronger in L5 than those in L6, but with considerable overlap [Fig. S7A (PT-Tect); 11.7 ± 2.2 (mean ± stdev), $N = 80$ for L6 neurons and 13.8 ± 1.6, $N = 185$ for L5 neurons, $p < 0.0001$ using bootstrap K-S test]. The strengths of other major projections were indistinguishable between the two groups (Fig. S7A-PT, $p = 0.5$ for both corticothalamic projections and corticostriatal projections using bootstrap K-S test). Therefore, the predictive power of projection patterns, combined with the strong overlap in strengths across layers, is consistent with the latter hypothesis that laminar position weakly correlates with projection pattern, but is not the driving determinant of projection strengths within classes. The fact that projection pattern within a class can be used to predict laminar position hints at the molecular and cellular processes underlying these projection differences.

Discussion

We have described BARseq, a sequencing-based neuroanatomical method that combines the high throughput of MAPseq with cellular and potentially subcellular spatial resolution. To achieve this, we developed an efficient method for *in situ* sequencing that allows efficient recovery of short diverse barcode sequences with single nucleotide resolution. The sensitivity of BARseq was comparable to that of conventional anterograde and retrograde methods. As a proof of principle, we applied BARseq to the projections in the auditory cortex to reveal diverse projection patterns at cellular resolution. Clustering analysis recovered previously described major classes of projection neurons, each with its expected laminar distribution. We also showed that BARseq can be combined with multiplexed *in situ* interrogation of gene expression, allowing us to relate gene expression with both projection patterns and soma location. Because neurons in the cortex and many other structures are spatially organized, BARseq can be used to probe the logic governing neuronal projections.

BARseq achieves multiplexed projection mapping by matching barcode sequences found at the soma with sequences recovered at projection targets. This strategy fundamentally differs from that used by traditional single-neuron tracing experiments based on one or a few distinct tracers, which confers both unique advantages and limitations upon BARseq.

Multiplexed projection mapping using cellular barcodes

The main advantages of BARseq stem from the fact that sequencing-based projection mapping does not require tracing neurites. In conventional optical approaches to mapping projections, axons are reconstructed by observing continuity between successive optical or physical sections. Any lost or distorted section may result in an error, and in general the probability of error increases with the length of the axonal projection and the number of neurites multiplexed. By contrast, because BARseq relies on matching sequences to reconstruct projections, errors do not accumulate for long axons: A distant subcortical projection is just as reliably matched to its source as the projection to a nearby cortical area. This feature allows BARseq to determine projection patterns for orders of magnitude higher density of neurons (hundreds to thousands of neurons per cortical area, and tens of thousands neurons per brain) than even the most advanced state-of-the-art multiplexed optical tracing methods [~50 neurons per cortical area (Guo et al., 2017)].

An additional advantage of BARseq is that it does not require a specialized high-speed microscope needed for anatomical tracing. BARseq requires only a standard epifluorescence or confocal microscope and access to high-throughput sequencing, both of which are widely available. BARseq thereby renders routine single-neuron projection mapping accessible even to laboratories that do not specialize in high-throughput neuroanatomy. BARseq thus complements conventional optical neuroanatomical methods, and broadens the use of single-neuron projection mapping in neuroscience.

A third advantage of BARseq is that it is compatible with many other methods. Many state-of-the-art high-throughput optical tracing methods (Gong et al., 2016; Guo et al., 2017) rely on destructive processes during imaging, and thus cannot be readily combined with other techniques, such as *in situ* sequencing, FISH, or antibody staining. This prevents relating neuronal morphology with other single-neuron properties, such as gene and protein expression. By contrast, BARseq can be readily combined with additional information from the same tissue. As an example, we have demonstrated FISH in combination with BARseq, but it is also

compatible with more highly multiplexed methods for mRNA detection (Chen et al., 2015; Shah et al., 2016; Wang et al., 2018). Combining these highly multiplexed *in situ* gene detection methods with BARseq could provide a basis for understanding how the myriad neuronal types defined by transcriptomics relate to their projection patterns. Moreover, following strategies analogous to those developed in similar contexts (Ko et al., 2011), BARseq can be used to determine the projections and gene expression of neurons whose *in vivo* functional properties have been determined with two-photon imaging. BARseq thus allows multiple neuronal properties to be queried simultaneously in individual samples.

The spatial resolution of BARseq

In its current form, an important limitation of BARseq is the relatively low spatial precision with which projections are resolved. Although BARseq localizes neuronal somata with cellular resolution, the spatial resolution for axonal projections is determined by the size of the cubelets dissected. There is thus a tradeoff between the resolution of projection mapping, the number of targets sampled, and throughput.

In this study, we chose to dissect brains relatively coarsely—only sufficient to distinguish among brain areas. Nonetheless, even at this scale we were able to recover known differences among neuronal classes, as well as to discover novel projection classes. With careful manual dissection or laser capture microdissection (Huang et al, *in preparation*), the higher resolution needed to isolate brain nuclei or laminae can be achieved. However, as the number of cubelets increases to more than a few hundred, sample processing becomes a limiting factor.

In the present work, we focused on using *in situ* sequencing to determine the sequence of barcodes expressed at high levels in the soma. However, *in situ* sequencing can readily achieve single molecule resolution (Fig. S2) (Ke et al., 2013; Lee et al., 2014; Wang et al., 2018). With high-resolution *in situ* single molecule sequencing, BARseq can be used to determine axonal projections or dendritic morphology (Yuan et al, *unpublished*), and potentially even synaptic connectivity. Such an approach would yield an "infinite color Brainbow," allowing reconstruction of densely labeled neurons.

Laminar organization of projection neurons

BARseq uncovered a highly diverse ensemble of projection patterns—more than an eighth (264) of the 2^{11} = 2048 possible patterns. Hierarchical clustering of these patterns recovered the three classical classes—IT, PT and CT—segregated into laminae as expected (Harris and Shepherd, 2015). Further clustering revealed as many as 25 projection subclusters (Fig. 7B), but unlike the major classes, these subclusters were not restricted to laminae or sublaminae. However, projection patterns were not wholly independent of laminar position (Fig. 7G). Thus, it appears that projection patterns correlate with laminar position, but that the dominant differences in projection driving the partitioning into subclusters are not aligned with laminae.

The lack of one-to-one correspondence between subclusters and laminae raises the question of how to interpret the projection clusters we observed. The top-level classes (IT, PT and CT) not only possess distinctive projection pattern and laminar position, but also differ in other ways including in morphology, gene expression and developmental timing. Although it would be tempting to interpret subclusters as subclasses, it is not yet clear whether they represent such distinct entities. One possibility is that subclusters can predict some properties (e.g. gene expression) beyond the connection patterns on which they are based (Economo et al., 2017;

Klingler et al., 2018), but that they correlate poorly with laminae because laminar position may be driven by other neuronal properties, such as the inputs (e.g. thalamus, contralateral cortex, etc.). In this case, the subclusters serve as a partial description for distinct subclasses of neurons. This hypothesis would be consistent with previous findings of mixed subcortical projection patterns in a single class of PT neurons in the somatosensory cortex (Rojas-Piloni et al., 2017). Alternatively, it is possible that the laminar position of the soma correlates better with the precise lamina—rather than area—in which an axon terminates. For example, Tasic et al. (2017) reported that L6 IT neurons project to the deep layers of the cortex, whereas L5 and L2/3 IT neurons project to all layers. Finally, it could be that clustering is not the most useful statistical approach for understanding projection diversity, and the subclusters we identified represent unnatural consequences of our Procrustean analyses. However, both the correlation between projection patterns (Fig. 7G) and laminae and the strong overlap in projection patterns across laminae (Fig. S7A-C) were observed independent of clustering beyond the major classes. Therefore, whether the subclusters represent biologically distinct groups of neurons does not affect the conclusion that neuronal projections correlate with laminae, but are largely intermingled.

The lack of laminar segregation of subclusters stands in contrast with the lamina-specific neuronal subclasses identified based on gene expression. For example, our analysis of auditory cortical projection patterns did not reveal any L2/3-specific IT cluster, whereas several L2/3-specific subclasses were identified in the visual cortex based on gene expression (Tasic et al., 2017). Given that the expression of genes driving axonal pathfinding, fasciculation, and targeting may have subsided in adult animals (Krishna et al., 2009; Redies, 1997), it is perhaps not surprising that some genes are specific to particular laminae whereas axonal projection patterns are not.

Rosetta Brain

Neuronal classes are defined by the combination of multiple features (Cadwell et al., 2016; Economo et al., 2017; Paul et al., 2017), including anatomical characteristics such as morphology and connectivity (Economo et al., 2016; Gerfen et al., 2016), molecular characteristics such as gene (Tasic et al., 2016; Tasic et al., 2017; Zeisel et al., 2015) or protein expression, and functional characteristics such as behaviorally-evoked activity (Cadwell et al., 2016). Defining biologically relevant neuronal classes thus requires a multi-faceted approach that combines measurements of several neuronal properties together at cellular resolution. BARseq offers a solution to incorporate high-throughput single-neuron projection mapping into such approaches by preserving the soma locations of neurons.

The ability to relate diverse characteristic of many single neurons simultaneously, in a co-registered fashion, within single brains, is an important challenge in neuroscience (Marblestone, 2014). Even simultaneous co-registration of two characteristics can be challenging, but the few successes in this arena have led to insights about the functional organization of neural circuits (Bock et al., 2011; Sorensen et al., 2015). BARseq represents a significant step toward this goal by allowing the incorporation of sequencing-based neuro-anatomy into a multi-modal description of neurons using the soma locations with high throughput. Adapting *in situ* methods in addition to those detecting gene expression, such as *in vivo* two-photon imaging, to combine with BARseq 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 (Marblestone, 2014).

Acknowledgements

The authors would like to acknowledge Kay Tye, 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.

Author contributions

X.C. and A.M.Z conceived the study. X.C., H.Z. and Y.S. optimized and performed BARseq. X.C. and Y.S. performed BARseq in combination with FISH and *in situ* sequencing of endogenous genes. 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.

Competing interests

A.M.Z is a founder and equity owner of MapNeuro.

Materials & Correspondence

All correspondence and material requests should be made to A.M.Z

Figure legends

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 (NGS). 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. (B) 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 template. 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) Comparison of *in situ* sequencing and hybridization techniques. Methods shown on the left use multiple rounds of hybridization to probe and read out multiple mRNAs. Methods shown in the middle use sequencing to multiplex read out of *in situ* hybridization signals. Methods shown on the right copy target sequences from the mRNA into the rolonies, allowing true sequencing of mRNA *in situ*. (a) In merFISH, multiple probes with barcode sequences on both ends are hybridized to each target mRNA. The probes are then read out using multiple rounds of hybridization using fluorescently labeled secondary probes. (b) In seqFISH, multiple rounds of HCR-amplified FISH are carried out on the same target mRNAs, and the mRNA identities are read out using the color sequence. (c) In STARmap, a linear probe and a padlock probe are hybridized to adjacent sequences on the target. When bound, the linear probe acts as a splint to allow circularization of the padlock probe. The circularized padlock probe is then used as template for rolling circle amplification (RCA). (d) In the non-gapfilling padlock method, the target mRNA is reverse transcribed to cDNA. A padlock probe then hybridizes to the cDNA and circularized. The circularized padlock probe is then used as template for RCA. (e) In the Circligase-based FISSEQ, the cDNA is circularized using an ssDNA ligase. This circularized cDNA is then used as template for RCA. (f) The gapfilling padlock method is similar to the non-gapfilling version, except that a gap between the two arms of the padlock is filled by copying from the cDNA, thus allowing the actual cDNA to be sequenced. (D) 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. (E) Images of the first four sequencing cycles of the same neurons shown in (D). The bases corresponding to the four colors and the sequences of the three neurons circled in (E) are indicated to the right. 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*) and rolonies (*right*) generated in the same brain slice. GFP intensity does not correlate perfectly with rolony intensity due to differences in protein and RNA expression. Insets: GFP and rolonies of a non-barcoded brain slice. Scale bars = 50 µm. (B) Representative high-resolution sequencing image of a barcoded brain slice. The color code is shown in Fig. 1E. Inset shows sequencing image of a non-barcoded brain slice. Scale bar = 50 µm. (C) Representative low-resolution images of the indicated cycles of barcode sequencing in a brain slice. The sequences of the three cells base-called are indicated below the images. Only bases corresponding to the images shown are

capitalized and color-coded. Scale bars = 100 µm. (D) 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. (E) 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). (F) 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. See also Fig. S1.

Fig. 3. Validation of BARseq using retrograde tracing. (A) A brain is injected with CTB in the contralateral auditory cortex, and barcoded in the ipsilateral auditory cortex. BARseq results of the barcoded neurons are then compared to retrograde labeling by CTB. (B) A representative image of a brain slice double labeled with barcodes (cyan) and CTB (magenta) from the contralateral auditory cortex. Scale bar = 100 µm. (C) 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). (D) Histograms of the number of barcode molecules for each neuron recovered in the olfactory bulb (OB, blue) and in the contralateral auditory cortex (AudC, red). The x-axis is drawn on a log scale. The vertical dashed line indicates the noise threshold used in the experiment (5 molecules). See also Table S1.

Fig. 4. Correlating gene expression and projections using BARseq. (A) The brain slices are processed for FISH and imaged at high resolution, followed by barcode sequencing. The barcode sequences are then matched to barcodes recovered at target sites to determine projections. (B)(C) Representative high-resolution images of FISH (left) against Slc17a7 (magenta) and Gad2 (cyan), and low-resolution sequencing images of cycle 1, 8, and 15 of the same sample are shown in (B). The three cells circled in (B) are base-called in (C) with their expression profile of Slc17a7 and Gad2 shown. Scale bars = 50 µm. See also Fig. S2 and Table S2.

Fig. 5. Mapping projections of the auditory cortex using BARseq. (A) Histogram of the minimal pairwise hamming distance of the first 15 bases of barcodes recovered from brain XC9. (B) Single-cell projection patterns sorted by cluster identities in Fig. 7. Each row represent a barcode and each column represent projection strengths to the indicated brain 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. (C) Conventional bulk GFP tracing intensities were plotted against the bulk projection strength obtained from BARseq. Error bars indicate SEM. N = 5 for GFP tracing and N = 3 for BARseq. 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) The numbers of binarized projection patterns (y-axis) after filtering for primary projection strength (x-axis). Blue line indicates filtering at the indicated thresholds, and red line indicates subsampling to the same sample size. Black lines and error bars indicate 95% confidence interval for subsampling. (F) Histogram of the number of projections per neuron. (G) The fractions of multi-projection neurons (y-axis) are plotted against the ratio between the secondary and primary projections (x-axis). Blue line indicates actual distribution and red line indicates fitting with a log normal distribution. See also Table S3.

Fig. 6. The organization of major classes of projection neurons in the auditory cortex. (A) The mean projection patterns of clusters corresponding to the indicated 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. (B) Triple retrograde tracing of neurons projecting to the rostral striatum (CTB-647), the caudal striatum (CTB-488), and the tectum (RetroBeads). (C) Representative image of the triple retrograde labeling in the auditory cortex showing neurons projecting to the rostral striatum (magenta), the caudal striatum (cyan), and the tectum (yellow). Scale bar = 100 µm. (D) Venn diagram showing the number of neurons projecting to each of the three areas. (E) 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 laminae and their boundaries are marked. Scale bar = 100 µm. Inset: histograms of each class of projection neurons in the pooled BARseq dataset. The y-axis indicates cortical depth and the x-axis indicates the number of neurons at that depth in a particular class. The laminae are indicated on the right and the boundaries between two layers are marked by dashed lines.

Fig. 7. Subcluster organization of auditory projection neurons. (A) Positive (green) or negative (red) Spearman correlation coefficients among projections to the indicated areas in the indicated classes. (B) 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. (C) t-SNE plot of the projection neurons. The neurons are color-coded by their first level subcluster identities post-hoc. (D) The fraction of well-classified neurons in each core (blue) and intermediate (red) subcluster. The subcluster labels correspond to those in (B) and the class they belong to are labeled below. The dotted line indicates 80% well-classified neurons. (E) The normalized entropy of nodes/leaves (y-axis) in the indicated clustering hierarchy (x-axis). The branch of the subcluster nodes/leaves were color coded as indicated. Grey bars indicate mean ± stdev of all nodes/leaves of a specific hierarchy. Hierarchies 1-3 correspond to class divisions and hierarchies 4-7 correspond to subcluster divisions. (F) The laminar distribution of all subclusters. Individual well-classified (red) or ambiguously classified (grey) neurons are superimposed on top of the distribution plots (light grey). The class to which each subcluster belongs is indicated below. Neurons of subcluster 14 were likely misidentified PT neurons, and were labeled with a dashed line underneath. (G) The fractions of correct prediction (x-axis) of laminar location above chance level for ITc (blue) and PT (red) neurons given the combinatorial projection patterns (stars) or single projections (dots) are plotted against the logarithms of the p values (y-axis). The dashed line indicates significance level after Bonferroni correction. See also Fig. S4-S7.

Supplementary Figure and Table legends:

Fig. S1, related to Fig. 2 and STAR Methods. Optimization of BaristaSeq for brain slices. (A) Amplification of barcodes in brain slices in the indicated reaction chambers. Scale bars = 100 µm. (B) Merged images of rolonies (yellow) generated in barcoded brain slices and the residual GFP signals (cyan) with the indicated time of pepsin treatment. Scale bars = 100 µm. (C) Comparison of barcode amplicons generated using BaristaSeq (a), the original padlock method (b), and FISSEQ (c). Scale bars = 50 µm. (D) Sequencing images of cycles 2, 4, and 6 of barcoded brain slices sequenced using SOLiD sequencing chemistry (*top*) and using Illumina sequencing chemistry (*bottom*). Imaging conditions were kept constant throughout each sequencing run. Scale bars = 100 µm. (E) Average signal-to-noise ratio of Illumina (red) and SOLiD (blue) sequencing *in situ* over cycles. Error bars indicate the standard errors for the SNR for pixels. (F-H) Sequencing quality and signal intensity of individual base calls (F), mean signal intensity over cycles (G), and the fraction of the bases over cycles (H) are plotted.

Fig. S2, related to Fig. 4. Multiplexed targeted *in situ* sequencing of mRNA. (A) A representative image of sequencing cycle 2 in the visual cortex. 55 genes are targeted using barcoded non-gapfilling padlock, and the barcodes are sequenced for six cycles to identify genes. (B) Individual mRNA rolonies in (A) are color-coded by barcode identities. Only rolonies with barcodes matching the gene panel are shown. In both (A) and (B), scale bars = 100 µm.

Fig. S3, related to STAR Methods. Laminar distribution of projection neurons. (A)(B) Representative images of FISH against Cux2 (A) and Fezf2 (B) in two adjacent slices. Scale bars = 50 µm. (C) Normalized layer boundaries determined in three pairs of slices across the auditory cortex. In each slice, the thickness of the cortex was normalized to that in the BARseq brains, and the boundary positions were scaled accordingly. (D) Violin plots of the laminar distribution of all BARseq neurons (All) and those with (Proj) or without (Non-proj) detected projections. Individual neurons (red) are plotted on top of the distribution plots.

Fig. S4, related to Fig. 7 and STAR Methods. Hierarchical clustering of projection neurons in the mouse auditory cortex. (A) The workflow of the hierarchical clustering. (B) The projections of neurons (center) are filtered by decomposing into six projection modules (Basis 1-6). (C) The projection patterns of two example neurons (left) are filtered using projection modules (right). The weight for each module is labeled to the left of the modules. (D) Comparison of the original projection strengths (blue) and the filtered projection strengths (red) for the two neurons shown in (C). (E) The fraction of variance explained (y-axis) using NMF (blue), individual projections (red), and PCA (black). (F) The fraction of neurons that remain in the same class-level clusters (y-axis) when filtering the projection data with the indicated number of projection modules (x-axis) compared to the clusters without filtering.

Fig. S5, related to Fig. 7 and STAR Methods. Hierarchical clustering of projection neurons in the mouse auditory cortex. (A) The distribution of the maximum cluster probability for individual neurons when classified using all 11 projection areas (a) or 10 projection areas (b-l). For classification using 10 projection areas, the unused projection area is labeled on top of each graph. (B) Comparison of clusters obtained using k-means (upper row), spectral clustering (middle row), and Louvain community detection (lower row) at the indicated hierarchies. All clusters were color-coded onto the same t-SNE plot. The colors are randomly assigned to individual clusters.

Fig. S6, related to Fig. 7. Laminar distribution of projection neuron subclusters. (A) Differences in normalized entropy (x-axis) of individual subclusters between the two brains are plotted against the negative logarithm of the p values (y-axis). The subclusters were color-coded according to their class-level divisions as indicated. The p values were shown without multiple testing correction. The red vertical dashed line indicate no difference in entropy, and the black horizontal dashed line indicate significance level after Bonferroni correction. (B) Differences in mean laminar locations (x-axis) of individual subclusters between the two brains are plotted against the negative logarithm of the p values (y-axis). The subclusters were color-coded according to their class-level divisions as indicated. The p values were shown without multiple testing correction. The red vertical dashed line indicate no difference in the mean laminar locations, and the black horizontal dashed line indicate significance level after Bonferroni correction. (C)(D) The fractions of correct prediction (x-axis) of laminar location above chance level for ITc (blue) and PT (red) neurons given the combinatorial projection patterns (stars) or single projections (dots) are plotted against the logarithms of the p values (y-axis) for brain XC9 (C) and XC28 (D). The dashed line indicates significance level after Bonferroni correction.

Fig. S7, related to Fig. 7. Laminar distribution of projections. (A) The strengths of the indicated projection (y-axes) of individual neurons of the indicated classes are plotted against their laminar position (x-axes). The mean projection strengths are marked by black line. The layer boundaries are marked by dotted vertical lines. (B) The first four principle components of the major classes. In each graph, the thickness of the arrows indicate relative projection strengths. Blue arrows indicate projections to the contralateral side, and red arrows indicate negative values. (C) The loadings of the first four principle components of neurons of the indicated classes (y-axes) are plotted against their laminar positions (x-axes). Black lines indicate mean loadings. Dotted lines indicate layer boundaries.

Table S1, related to Fig. 3. Comparison between MAPseq and retrograde tracing. Each row represents a single neuron recovered from BaristaSeq with visible GFP signal from the barcodes and good sequencing quality (quality score > 0.75, column 4). The first four columns indicate the raw barcode counts in the olfactory bulb (OB), contralateral auditory cortex (c1), the cortical area surrounding the contralateral auditory cortex with CTB signals visible to the naked eyes (c2), and an even larger cortical area surrounding the tracer area with CTB signals visible under the microscope (c3). The rest of the columns indicate whether the cell projects contralaterally based on CTB and/or BARseq

Table S2, related to Fig. 4. Correlation of gene expression and projections using BARseq. Each row represents a single barcoded neuron with a minimum quality score over 0.8. Columns 2-7 indicate barcode molecule counts in each collected area, and column 8 indicates total molecule counts. Columns 9 and 10 indicate whether the cells expressed Slc17a7 and/or Gad2.

Table S3, related to Fig. 5. The number of barcodes sequenced per brain from the projection sites, the number of cells sequenced per brain from the auditory cortex, and the number of *in situ* sequenced barcodes matching barcodes at the projection sites with or without quality filtering are indicated for each brain. *In XC14, the auditory cortex was sequenced *in vitro*.

Table S4, related to STAR Methods. List of oligos for multiplexed *in situ* sequencing of mRNA.

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 (Chen et al., 2018). RT primers and padlock probes for gene detection are listed in Table S4.

For validation of BaristaSeq in brain slices, we used a barcode library previously described by Kebschull et al. (2016a). 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 (Han et al., 2018). This library was not fully sequenced *in vitro*.

Animals and tissue processing:

All animal procedures were carried out in accordance with Institutional Animal Care and Use Committee protocol 16-13-10-07-03-00-4 at Cold Spring Harbor Laboratory. Eight to nine week old male C57BL/6 mice were injected in the left auditory cortex at -4.3 mm ML, -2.6 mm AP from bregma, with 140nL 1:3 diluted 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 (Kebschull et al., 2016a). The post-fixed auditory cortex was cryo-protected, embedded, and cryo-sectioned as described above.

For combined BARseq/FISH experiments, we injected 8-week old C57BL/6 male mouse at -4 mm ML, -2.5 mm AP from bregma, with 140nL 1:3 diluted Sindbis virus at depths 200 µm, 400 µm, 600 µm, and 800 µm straight down. After 24 hours, the animal was processed as described above for BARseq experiments.

To compare BARseq to retrograde tracing, we injected 140 nL Alexa 647 labeled cholera toxin subunit B (CTB) into the right auditory cortex at 4.3 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 1:3 diluted JK100L2 virus into the left auditory cortex at -4.3 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 another 44 hrs, the animals were then processed as for conventional BARseq samples.

BaristaSeq

BaristaSeq on cultured neurons was performed as described (Chen et al., 2018). 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 hr 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, we tested three commercially available reaction chambers that were physically compatible with our samples (Fig. S1Aa-c), and found that the HybriWell-FL sealing system was the only system that did not inhibit rolony formation (Fig. S1Ab). The ImmEdge hydrophobic barrier pen also produced good amplification (Fig. S1Ad), but the HybriWell-FL system offered better control of liquid evaporation during heating steps and easier handling. All slides with brain slices were thus 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 pepsin digestion step. This step was necessary to increase accessibility of fixed RNAs (Fig. S1B) and to reduce the GFP signal from the cells (cyan in Fig. S1B), which may interfere with sequencing signals. We found that 3 mins of 0.2% pepsin digestion in 0.1 M HCl at room temperature greatly increased rolony formation (Fig. S1Bb) compared to no pepsin treatment (Fig. S1Ba), whereas 5 mins of pepsin digestion caused excessive tissue loss (Fig. S1Bc). We therefore used 3 mins of pepsin digestion for BaristaSeq in most brain slices, but the optimal timing could vary with different fixation conditions. We then proceeded with ethanol dehydration, followed by reverse transcription, padlock gap-filling and ligation, and RCA as described above for cultured neurons.

To sequence the barcodes using Illumina sequencing chemistry, we based our sequencing protocol on the HiSeq recipe files (Chen et al., 2018) and reduced the CRM incubation time to two minutes each due to efficient heat transfer of the reaction chambers. We also increased the PBST washes to four to eight times after the IRM reactions to counteract the increased background staining in tissue slices. This optimized Illumina sequencing protocol resulted in an average signal-to-noise ratio 39 ± 4 in the first six sequencing cycles (Fig. S1D, E), ~10-fold higher than that of sequencing by ligation (SOLiD; 4 ± 1) used by other sequencing methods (Ke et al., 2013; Lee et al., 2014).

Imaging

All imaging except for the experiments in Fig. 2C-F was performed on an UltraView VoX spinning disk confocal microscope (Perkin Elmer) with Volocity 6.3 software as previously described (Chen et al., 2018). The sequencing channels and their calibrations were described previously (Chen et al., 2018). The sequencing channel filters were as previously described

(Chen et al., 2018). The experiments in Fig. 2C-F were produced on a Zeiss LSM 710 laser scanning confocal microscope with Zeiss Zen software as previously described (Chen et al., 2018).

All images in figures were maximum projections of z-stacks shown after rolling ball background subtraction except for images in Fig. S1D, which were shown with only the camera blackpoint subtracted from the images. For Fig. 2B we further adjusted gamma to 0.5 to show both cell bodies and single rolonies in dendrites and axons.

Base-calling

Base-calling was performed as described (Chen et al., 2018), 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 (Ke et al., 2013).

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 (Chen et al., 2018; Lee et al., 2014). 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 reduce 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 included 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. The auditory cortex dissections also included the neighboring temporal association area, and the visual cortex collected included the posterior parietal cortex. Slice images, dissected areas, and their correspondence to the Allen Reference Atlas are available at Dryad (see Data Availability). The projection sites were sequenced as described for MAPseq (Kebschull et al., 2016a).

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 (Table S3). 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 with more than 20 molecules per barcode at the injection site and were used for the subsequent analyses.

In BARseq experiments, the injection sites were sequence *in situ* to 15 bases. The 15 bases read length *in situ* was sufficient to distinguish unambiguously all infected barcodes allowing one mismatch. For the XC9 brain, barcodes recovered through MAPseq had a mean hamming distance of 4.5 ± 0.7 (mean ± stdev; Fig. 5A). Only one pair (0.04%) out of 4841 barcodes had a hamming distance of 1 and 10 pairs (0.4%) out of 4841 had a hamming distance of 2. Because the sequencing experiment in Fig. 2 showed only a single error for 51 barcodes, each sequenced 25 bases, our sequencing error rate was approximately $1/(51 \times 25) = 0.08\%$. Therefore, assuming that sequencing errors have no bias toward a particular base, the probability of matching an *in situ* barcode to the wrong MAPseq barcode, while allowing one mismatch, is $\frac{2}{4841} \times 0.08\% \div 3 = 1e-7$. The probability of an *in situ* barcode matching to two MAPseq barcodes is $\frac{20}{4841} \times 0.08\% \times 2 \div 3 + \frac{2}{4841} = 4e-4$. Although we cannot detect false positive matches, an ambiguous match could be detected. In the XC9 data, however, no ambiguous match between the *in situ* barcodes and the MAPseq barcodes has occurred.

In addition, XC9 had three pairs of barcodes whose first 15 bases were the same. These appeared to have arisen from amplification errors in homopolymer stretches of the same barcode rather than different barcodes, because each pair had a single in-del and had almost identical projection patterns. These three pairs were not recovered *in situ* and thus did not affect the analyses.

Similarly, out of 13581 total sequences, XC28 had 5 pairs of barcodes within one mismatch and 106 pairs of barcodes within two mismatches for the first 15 bases. No XC28 barcodes had identical sequences in the first 15 bases. The probability of a wrong match in XC28 is $\frac{10}{13581} \times 0.08\% \div 3 = 2e-7$, and the probability of an ambiguous match in XC28 is $\frac{212}{13581} \times 0.08\% \times 2 \div 3 + \frac{10}{13581} = 7e-4$. No actual ambiguous match was seen in XC28. Therefore, allowing one mismatch for a 15-base sequence is sufficient to match barcodes in the somas to those at the projection sites unambiguously for both brains.

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 callosal commissure) and neurons in highly distorted slices (as judged by an abnormal cortical thickness). After filtering, 1309 neurons were used in the analyses.

Comparison of BARseq to retrograde tracer

The animals were injected and processed as described above. We 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 was visible to the naked eye, and the surrounding areas where CTB was visible under a fluorescent microscope. The last three samples thus formed concentric rings around the CTB injection site. All three samples gave consistent results regarding contralateral projections (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. We then only counted neurons clearly expressing GFP, which was essential to properly judge colocalization with CTB. To find the fraction of CTB labeled neurons that were 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 was 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 from bregma 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).

Multiplexed targeted *in situ* detection of endogenous mRNA

An 8-week old C57BL/6 male mouse was perfused and its brain was post-fixed in 4% PFA for 96 hours, followed by cryo-protection and cryo-sectioning to 14 µm brain slices. To produce rolonies against endogenous genes, we followed the BaristaSeq protocol for brain slices as described above with two modifications. First, the pepsin treatment was performed at pH 4 for one hour instead of at pH 2 for 3 mins. Second, no Phusion DNA polymerase was added during the padlock ligation step. We used RT primers and padlock probes designed against a panel of 55 genes with differential expression in the primary visual cortex. Each padlock probe carried a 6-nt barcode corresponding to the gene identity. The sequences of the RT primers and padlock probes are provided in Table S4.

The sample was then sequenced for five cycles as described above, except that the images were collected using a 20x 0.75NA objective without binning. The sequencing images were then corrected for shifting among channels, followed by median filtering, spectral correction, and background subtraction. The preprocessed images were then roughly aligned by maximizing

enhanced correlation coefficients (Evangelidis and Psarakis, 2008). We then identified rolonies using the "Find Maxima" function in ImageJ, and aligned the rolonies across sequencing cycles using non-rigid Iterative Closet Point (Amberg et al., 2007). We then matched the rolonies in sequencing cycles 2-5 to rolonies identified in cycle 1, allowing a maximum shift of 2 pixels (0.65 µm), and base-called. The resulting barcode sequences were then matched to possible barcodes in the padlock probes.

Correlating projections to gene expression using BARseq

We injected and processed animals as described above. For target projection sites, we collected the olfactory bulb, ipsilateral cortical areas (mainly the visual cortex), contralateral cortical areas (the auditory and visual cortex), the striatum, the thalamus, and the tectum. These areas were processed for MAPSeq as described above and sequenced on an Illumina MiSeq. The injection site was cut out from the brain, post-fixed and cryo-protected for cryo-sectioning.

We cryo-sectioned the injection site to 14 µm slices, which were adhered to glass slides using a UV-activated optical glue (NOA 81, Norland Products) and a home-made tape transfer system (Pinskiy et al., 2015). The brain slices were then processed for BARseq to the first cross-link step after reverse transcription. After neutralization of additional crosslinkers with Tris, we proceeded with ViewRNA ISH (Thermo Fisher Scientific) using a class 1 probeset for Slc17a7 and a class 6 probeset for Gad2 according to the manufacturer's protocol. We then imaged each area around the injection site (identified by the presence of GFP) using a 20x 0.75NA objective on a spinning disk confocal for both FISH channels (RFP and Cy5), the GFP channel, and the DIC channel. We then stripped away the FISH probes using 80% formamide for 5 mins twice, followed by three washes in 10% formamide in 2x SSC, and two washes with PBST. We then proceeded to the padlock ligation step in BARseq and produced rolonies as described above.

The sequencing of rolonies was done as described above except that the first sequencing cycle was imaged first using a 20x 0.75NA objective without binning for all four sequencing channels and the DIC channel. This was followed by imaging using a 10x 0.45NA objective with 2x binning. All subsequent sequencing cycles were imaged at the lower resolution to increase throughput.

We then registered the high-resolution FISH images to the high-resolution first cycle sequencing images using the DIC channel. To match the high-resolution first cycle sequencing images to the low-resolution sequencing images, we down-sampled the first-cycle sequencing images by four folds and applied a Gaussian filter to mimic the lower optical resolution. The two images were then roughly aligned manually. We then extracted cell-body locations using the "Find Maxima" function in ImageJ, and further aligned the cell-body locations using Iterative Closest Points (Besl and McKay, 1992) in MATLAB (https://www.mathworks.com/matlabcentral/fileexchange/27804-iterative-closest-point). The barcodes were called from the low-res images using the cell-body locations extracted in the previous step, and the expression of Slc17a7 and Gad2 were determined manually by examining the overlap between FISH signals and barcode rolonies. We then filtered out cells with a minimum sequencing quality score of less than 0.8, and further removed debris based on morphology and homopolymer barcode sequences. The remaining 100 cells were analyzed for projection pattern and gene expression.

Estimation of laminar boundaries using fluorescent *in situ* hybridization (FISH)

To estimate the boundaries of cortical layers, we performed FISH against two known layer-specific marker genes, Cux2 (Custo Greig et al., 2013) (Fig. S3A), and Fezf2 (Fig. S3B). Cux2 was strongly expressed in L2/3 and only sporadically in other layers; Fezf2 was strongly expressed in L5 and weakly in L6. Because L4 is poorly defined in the auditory cortex(Linden and Schreiner, 2003), we omitted L4 and defined only the remaining two borders. We defined the L2/3 and L5 border as below the Cux2 band and above the strong Fezf2 band, and defined the L5 and L6 border as between the strong and weak bands of Fezf2. To account for variation in cortex thickness and sample preparation, we examined three slices spanning 800 µm in the auditory cortex, normalized all cortical thickness to 1200 µm (i.e. the same cortical thickness as the BARseq brains), and calculated the mean positions of layer boundaries (Fig. S3C), The L2/3 and L5 border defined by Fezf2 agreed with that defined by Cux2. Based on these measurements, we defined the L2/3 and L5 border to be at 590 µm and the L5 and L6 border to be at 830 µm. These borders were used for the BARseq analyses when layer identities were involved.

We saw few projection neurons in superficial L2/3 in our dataset. This is partially due to smaller number of neurons labeled near the cortical surface, and partially due to an enrichment of neurons without detectable projections in superficial L2/3 (Fig. S3D). Neurons in superficial L2/3 (or L2) of the auditory cortex are known to project locally and not contralaterally (Oviedo et al., 2010). Because we did not sample neighboring cortical areas, these locally projecting ITi neurons would have shown as non-projecting neurons in BARseq.

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 (Oh et al., 2014) (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. Several areas we collected only corresponded to part of the areas of the same name in the Allen database, including 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 in the Allen Mouse Brain Connectivity Atlas; these area labels were different from labels in the Allen Reference Atlas), and the striatum (the rostral striatum and the caudal striatum samples were separated by two brain slices, or 600 µm). The auditory cortex area we collected also included the temporal association area. For both BARseq/MAPseq bulk projections and GFP bulk tracing data, we normalized the strengths of projection 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 Pearson correlation coefficient and the associated p value between the GFP tracing and BARseq/MAPseq bulk projection strengths of the corresponding brain areas.

Population analysis of projections

All population level analyses were carried out in MATLAB.

We calculated the false positive rate $FPR = N_{OB}/N_{total}$, where N_{OB} is the number of neurons with barcodes detected in the OB, and N_{total} is the total number of neurons. We then calculated the false discovery rate $FDR = (FPR \times n_0)/n_p$, where n_p indicates the total number of

projections detected, and n_0 indicates the total number of possible projections that were not detected.

To estimate the number of projection patterns, we binarized the projections using a threshold set by the maximum number of molecules recovered in the olfactory bulb negative control, and counted unique patterns. Because only one neuron out of 6391 had four molecules in the OB, and four out of 6391 had a single molecule each in the OB, this thresholding likely resulted in a conservative estimate of the unique projection patterns. To estimate the contribution of strong projections, we defined a strong secondary projection as one whose normalized count for a particular barcode was at least 10% of the strongest projection for that barcode.

We then estimated whether each projection pattern could have resulted from missing projections due to false negatives. For each "target" projection pattern, we identified all "parent" projection patterns with one more projection. Neurons with the parent projection patterns might have been misidentified as the target projection pattern of interest if a projection was missed. If the ratio of neurons with the target projection pattern to the ratio of neurons with parent projection patterns was smaller than the false negative rate, then we eliminated the target projection pattern. We repeated this process for all projection patterns, and the counted the remaining projection patterns.

To test if the estimated number of projection patterns was affected by the sensitivity of BARseq/MAPseq, we filtered neurons with a varying threshold for the primary projections (i.e. the strongest projections; Fig. 5E), which resulted in a reduction in projection patterns. To test if such reduction was caused by smaller sample sizes due to filtering, we randomly subsampled the neurons to the same sample size as that caused by thresholding the primary projection. This subsampling was repeated 100 times to estimate the 95% confidence interval.

To estimate the distribution of the strengths of secondary projections in multi-projection neurons (i.e. neurons with two or more projections), we filtered out neurons whose primary projections were too weak to allow more than one barcode molecules in the secondary projections. For example, when calculating the fraction of neurons with secondary to primary projection ratio of 0.1, we only included neurons whose primary projection was at least 20 barcodes. This would allow at least 2 barcodes to be observed in the secondary projection. We then separated the neurons into 100 bins according to their ratio of strengths between secondary and primary projections, and calculated the probability density function from the cumulative density function. We then fitted with a log normal distribution using gradient descent.

Hierarchical clustering

All clustering analyses (Fig. S4A) were done using the logarithm of the spike-in-normalized projection strength. We first filtered the projection data using non-negative matrix factorization (NMF)(Lee and Seung, 1999) in Matlab (Fig. S4B). 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. The projection patterns of individual neurons can thus be approximated by a weighted sum of the projection modules (Fig. S4C-E). We chose a k value that captured most of the variance in the data (Fig. S4E) and resulted in similar classification of neurons in the first two hierarchies (Fig. S4F). 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 (Fig. S4A), we split each node into two groups using k-means clustering on the squared Euclidean distance of the projection patterns. We then evaluated whether the split was significant using a Matlab implementation of SigClust (available from 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 used a probabilistic approach (Tasic et al., 2016) to assign neurons to these clusters. 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. The majority of neurons (5968/6391, or 93% of all neurons) were uniquely assigned to a single cluster with high probability (>98% probability), but a small set of neurons (421/6391, or 7% of all neurons) were assigned to two clusters at around 50% probability each (Fig. S5Aa).

Although BARseq has a low false-negative rate (~10%), such a false-negative rate may accumulate for subclusters with multiple projections, leading to higher rate in misclassification. To examine the extent of such misclassification, we classified neurons probabilistically using 10 out of the 11 projections, thus simulating the effect of a projection being uninformative (Fig. S5Ab-l). This analysis revealed similar distribution of neurons with the majority being well-classified and a smaller fraction being ambiguously assigned to two or more clusters. We then considered a neuron ambiguously classified if it was assigned to two or more clusters using all 11 projections, or assigned to the wrong cluster using any of the 10 projections if the resulting cluster was consistent with a dropped projection rather than a false-positive projection. This resulted in 5682/6391 (89%) well-classified neurons and 709/6391 (11%) ambiguous neurons. These estimates represent an upper bound on the number of ambiguous neurons, because it did not take into consideration the actual low false-negative rate of BARseq. The ambiguous neurons were concentrated in a few ITi subclusters (Fig. 7D), indicating that these subclusters may have resulted from misclassification of other subclusters due to false negative projections. Therefore the majority of neurons were unambiguously assigned to a single cluster even when missed projections are taken into consideration.

We also compared our clustering to graph based clustering using Louvain community detection (Blondel et al., 2008) and spectral clustering (Ng et al., 2002) (Fig. S5B). Louvain community detection identified 2-4 clusters at each hierarchy, and therefore did not fully correspond to the clusters obtained by bifurcation only at any hierarchical level. However, the resulting clusters from both methods, especially high-level nodes, were similar to those obtained using k-means. We chose to base all further analyses on clustering using k-means, because the major classes were better separated than using spectral clustering and the imposed bifurcation was easier to interpret than the clusters produced by Louvain community detection.

For Louvain community detection, we used 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).

For normalized spectral clustering (Ng et al., 2002), we used 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(van der Maaten and Hinton, 2008) was performed using a MATLAB implementation of the standard t-SNE (https://lvdmaaten.github.io/tsne/) using the log projection data as inputs.

We identified a cluster of neurons (Leaf 14, Fig. 7B) that appeared to be CT neurons projecting to the striatum, but not to the tectum. Layer 6 CT neurons, however, usually do not project to the striatum (Shepherd, 2013). The apparent striatal projections could be caused by contamination by fibers passing through the striatum. Alternatively, these neurons could be PT neurons in layer 5 and deep layer 6 whose subcerebral projections were missed due to either weak projections in the tectum or projections to other targets rather than the tectum. Our analyses of projection neurons *in situ* support the latter hypothesis. We observed a total of 590 corticothalamic neurons that did not project to the tectum or the striatum, and 505 that did project to the striatum. However, only 10 of the latter group was obtained in the *in situ* sequenced brains, compared to 229 of the former group. Therefore, most of the striatum and thalamus projecting neurons were from the conventional MAPseq brain (XC14). Because we only sequenced neurons at the center of each injection site *in situ* (XC9 and XC28), but collected a much larger injection site that may have included neighboring cortical areas in the conventional MAPseq brain (XC14), these neurons were rare in the auditory cortex, and were more likely in neighboring cortical areas, where PT neurons could project to other targets. Furthermore, the 10 neurons that projected to the striatum had a laminar profile similar to that of PT neurons, but different from those of the layer 6 CT neurons (Fig. 7F). These results suggest that most of the striatum projecting "corticothalamic" neurons were likely PT neurons in layer 5 and deep layer 6 in neighboring cortical areas.

Normalized entropy for the laminar distribution of a group of neurons

We examined whether clustering resulted in subclusters that were more restrictive in laminar locations. One natural measure of spatial compactness is the standard deviation of the spatial distribution of neurons within a class, but such a measure yields spuriously high values for multimodal distributions. We therefore examined the entropy (normalized to fall between 0 and 1) of the laminar distribution of all nodes and leaves in the clustering, a measure which is insensitive to the shape of the distribution.

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.

We did not see significant difference in the entropy of laminar distribution of subclusters across brains (Fig. S6A; $p > 0.05$ for all subclusters after Bonferroni correction). Furthermore, the distribution of neurons was also consistent across brains: the average laminar locations of neurons from each brain were similar for all but one subcluster (Fig. S6B; $p > 0.05$ after Bonferroni correction, Mann-Whitney U test). The one exception was CT neurons, which were on average 50 µm deeper in the XC28 than in XC9 (Fig. S6B; $p < 0.0005$ after Bonferroni correction), probably because more deep L6 neurons were labeled in XC28 than in XC9 (XC9 and XC28 had 165 and 132 L6 labeled neurons with laminar depths < 1000 µm, respectively, compared to 56 and 129 L6 neurons with laminar depths > 1000 µm; $p < 10^{-7}$, fisher's exact test). Therefore, this observed lack of laminar specificity of subclusters was consistent across samples.

Predicting laminar location using projections

We trained a random forest classifier with an in-bag fraction of 0.5 and 50 trees to predict whether the neurons were in L2/3, L5, or L6 given the combinatorial projection patterns or the strengths of a single projection. We used 5-fold cross-validation to evaluate the performance of the classifier, each time training on 80% of neurons and evaluating using the remaining 20%. To compare to the chance performance, we trained the same classifier using neurons with randomized layer labels.

Quantification and statistical analysis

All *p* values reported were after Bonferroni correction except in volcano plots (Fig. 7G and Fig. S6), where the original p values were plotted with the significance level threshold corrected for multiple testing. The statistical tests used for individual experiments were noted in the main text. All statistical tests were two-sided.

Data availability

All *in vitro* high throughput sequencing datasets (including the two BaristaSeq brains, XC9 and XC28, one MAPseq brain, XC14, one brain for the comparison between BARseq and retrograde tracing, XC54, and one brain for the combination of BARseq with FISH, XC66) are deposited to SRA (accession number SRP137078l; https://submit.ncbi.nlm.nih.gov/subs/sra/SUB3866273/overview). *In situ* sequencing data, images for the MAPseq dissections, and a list of Allen Reference Atlas slices corresponding to the dissection slices will be available from Dryad upon publication.

Code availability

Analysis code and data needed to reproduce all analyses will be available from Dryad upon publication.

References

Amberg, B., Romdhani, S., and Vetter, T. (2007). Optimal step nonrigid ICP algorithms for surface registration. Proc Cvpr Ieee, 1491-+.

Besl, P., and McKay, N. (1992). A Method for Registration of 3-D Shapes. IEEE Trans Pattern Anal Mach Intell *14*, 239-256.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment *2008*, P10008.

Bock, D.D., Lee, W.C., Kerlin, A.M., Andermann, M.L., Hood, G., Wetzel, A.W., Yurgenson, S., Soucy, E.R., Kim, H.S., and Reid, R.C. (2011). Network anatomy and in vivo physiology of visual cortical neurons. Nature *471*, 177-182.

Cadwell, C.R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K.F., *et al.* (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. Nat Biotechnol *34*, 199-203.

Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*, aaa6090.

Chen, X., Sun, Y.C., Church, G.M., Lee, J.H., and Zador, A.M. (2018). Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. Nucleic Acids Res *46*, e22.

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

Economo, M., Viswanathan, S., Tasic, B., Bas, E., Winnubst, J., Menon, V., Graybuck, L.T., Nguyen, T.N., Wang, L., Gerfen, C.R., *et al.* (2017). Distinct descending motor cortex pathways and their roles in movement. bioRxiv.

Economo, M.N., Clack, N.G., Lavis, L.D., Gerfen, C.R., Svoboda, K., Myers, E.W., and Chandrashekar, J. (2016). A platform for brain-wide imaging and reconstruction of individual neurons. Elife *5*, e10566.

Evangelidis, G.D., and Psarakis, E.Z. (2008). Parametric image alignment using enhanced correlation coefficient maximization. IEEE Trans Pattern Anal Mach Intell *30*, 1858-1865.

Gerfen, C.R., Economo, M.N., and Chandrashekar, J. (2016). Long distance projections of cortical pyramidal neurons. J Neurosci Res.

Gong, H., Xu, D., Yuan, J., Li, X., Guo, C., Peng, J., Li, Y., Schwarz, L.A., Li, A., Hu, B., *et al.* (2016). High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. Nat Commun *7*, 12142.

Guo, C., Peng, J., Zhang, Y., Li, A., Li, Y., Yuan, J., Xu, X., Ren, M., Gong, H., and Chen, S. (2017). Single-axon level morphological analysis of corticofugal projection neurons in mouse barrel field. Sci Rep *7*, 2846.

Han, Y., Kebschull, J.M., Campbell, R.A.A., Cowan, D., Imhof, F., Zador, A.M., and Mrsic-Flogel, T.D. (2018). The logic of single-cell projections from visual cortex. Nature *556*, 51-56.

Harris, K.D., and Shepherd, G.M. (2015). The neocortical circuit: themes and variations. Nat Neurosci *18*, 170-181.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wahlby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. Nat Methods *10*, 857-860.

Kebschull, J.M., Garcia da Silva, P., Reid, A.P., Peikon, I.D., Albeanu, D.F., and Zador, A.M. (2016a). High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. Neuron *91*, 975-987.

Kebschull, J.M., Garcia da Silva, P., and Zador, A.M. (2016b). A New Defective Helper RNA to Produce Recombinant Sindbis Virus that Infects Neurons but does not Propagate. Front Neuroanat *10*, 56.

Klingler, E., Prados, J., Kebschull, J., Dayer, A., Zador, A., and Jabaudon, D. (2018). Single-cell molecular connectomics of intracortically-projecting neurons. Biorxiv.

Ko, H., Hofer, S.B., Pichler, B., Buchanan, K.A., Sjostrom, P.J., and Mrsic-Flogel, T.D. (2011). Functional specificity of local synaptic connections in neocortical networks. Nature *473*, 87-91.

Krishna, K., Nuernberger, M., Weth, F., and Redies, C. (2009). Layer-specific expression of multiple cadherins in the developing visual cortex (V1) of the ferret. Cereb Cortex *19*, 388-401.

Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature *401*, 788-791.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S., Li, C., Amamoto, R.*, et al.* (2014). Highly multiplexed subcellular RNA sequencing in situ. Science *343*, 1360-1363.

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

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. (2014). Rosetta Brains: A Strategy for Molecularly-Annotated Connectomics. *arXiv*.

Meyer, H.S., Schwarz, D., Wimmer, V.C., Schmitt, A.C., Kerr, J.N., Sakmann, B., and Helmstaedter, M. (2011). Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5A. Proc Natl Acad Sci U S A *108*, 16807-16812.

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

Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L., Henry, A.M.*, et al.* (2014). A mesoscale connectome of the mouse brain. Nature *508*, 207-214.

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

Paul, A., Crow, M., Raudales, R., He, M., Gillis, J., and Huang, Z.J. (2017). Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity. Cell *171*, 522-539 e520.

Pinskiy, V., Jones, J., Tolpygo, A.S., Franciotti, N., Weber, K., and Mitra, P.P. (2015). High-Throughput Method of Whole-Brain Sectioning, Using the Tape-Transfer Technique. PLoS One *10*, e0102363.

Redies, C. (1997). Cadherins and the formation of neural circuitry in the vertebrate CNS. Cell Tissue Res *290*, 405-413.

Rojas-Piloni, G., Guest, J.M., Egger, R., Johnson, A.S., Sakmann, B., and Oberlaender, M. (2017). Relationships between structure, in vivo function and long-range axonal target of cortical pyramidal tract neurons. Nat Commun *8*, 870.

Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., *et al.* (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell *174*, 1015-1030 e1016.

Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. Neuron *92*, 342-357.

Shepherd, G.M. (2013). Corticostriatal connectivity and its role in disease. Nat Rev Neurosci *14*, 278-291.

Slater, B.J., Willis, A.M., and Llano, D.A. (2013). Evidence for layer-specific differences in auditory corticocollicular neurons. Neuroscience *229*, 144-154.

Sorensen, S.A., Bernard, A., Menon, V., Royall, J.J., Glattfelder, K.J., Desta, T., Hirokawa, K., Mortrud, M., Miller, J.A., Zeng, H., *et al.* (2015). Correlated gene expression and target specificity demonstrate excitatory projection neuron diversity. Cereb Cortex *25*, 433-449.

Stahl, P.L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., *et al.* (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78-82.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., *et al.* (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci *19*, 335-346.

Tasic, B., Yao, Z., Smith, K.A., Graybuck, L., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., *et al.* (2017). Shared and distinct transcriptomic cell types across neocortical areas. bioRxiv.

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

Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., *et al.* (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science.

Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., *et al.* (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138-1142.

Znamenskiy, P., and Zador, A.M. (2013). Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. Nature *497*, 482-485.

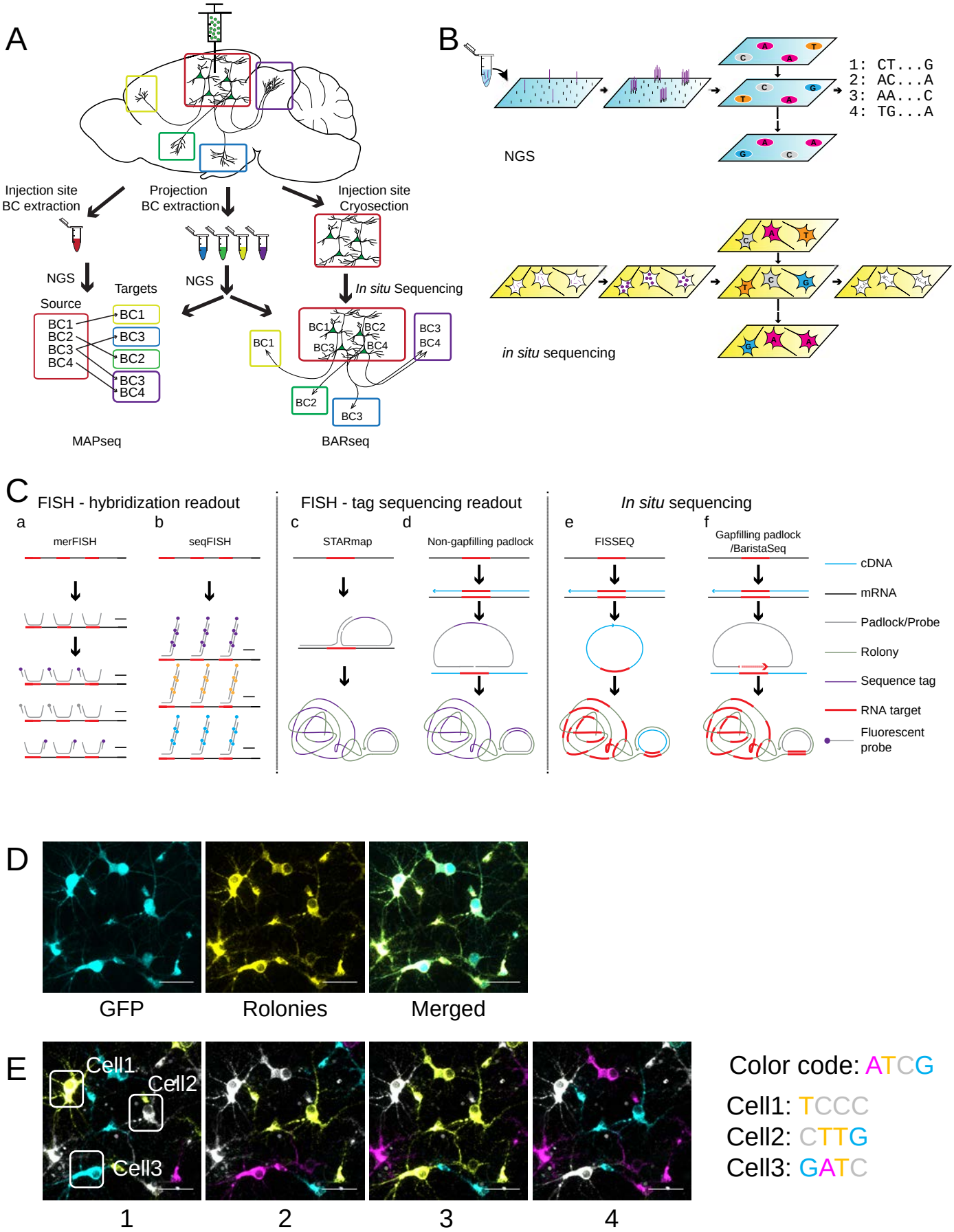

A
Injection site
BC extraction
Projection
BC extraction
Injection site
Cryosection
NGS
Targets
Source
BC1
BC2
BC3
BC4
NGS
In situ Sequencing
MAPseq
BARseq
B
NGS
1: CT...G
2: AC...A
3: AA...C
4: TG...A
in situ sequencing
C
FISH - hybridization readout
FISH - tag sequencing readout
In situ sequencing
a
b
c
d
e
f
merFISH
seqFISH
STARmap
Non-gapfilling padlock
FISSEQ
Gapfilling padlock
/BaristaSeq
cDNA
mRNA
Padlock/Probe
Rolony
Sequence tag
RNA target
Fluorescent
probe
D
GFP
Rolonies
Merged
E
Cell1
Cell2
Cell3
1
2
3
4
Color code: ATCG
Cell1: TCCC
Cell2: CTTG
Cell3: GATC

Figure 1

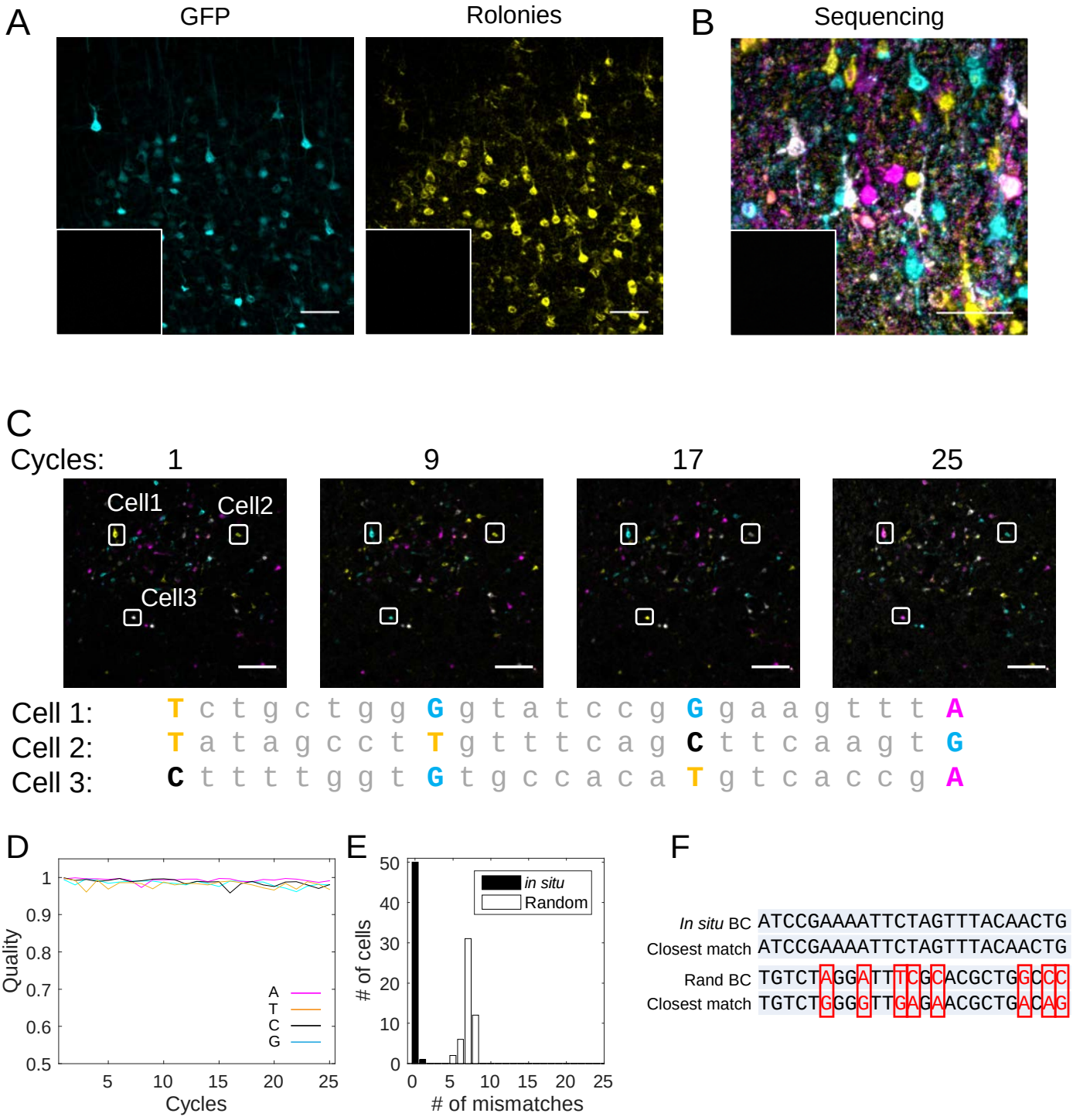


Figure 2, see also Fig. S1

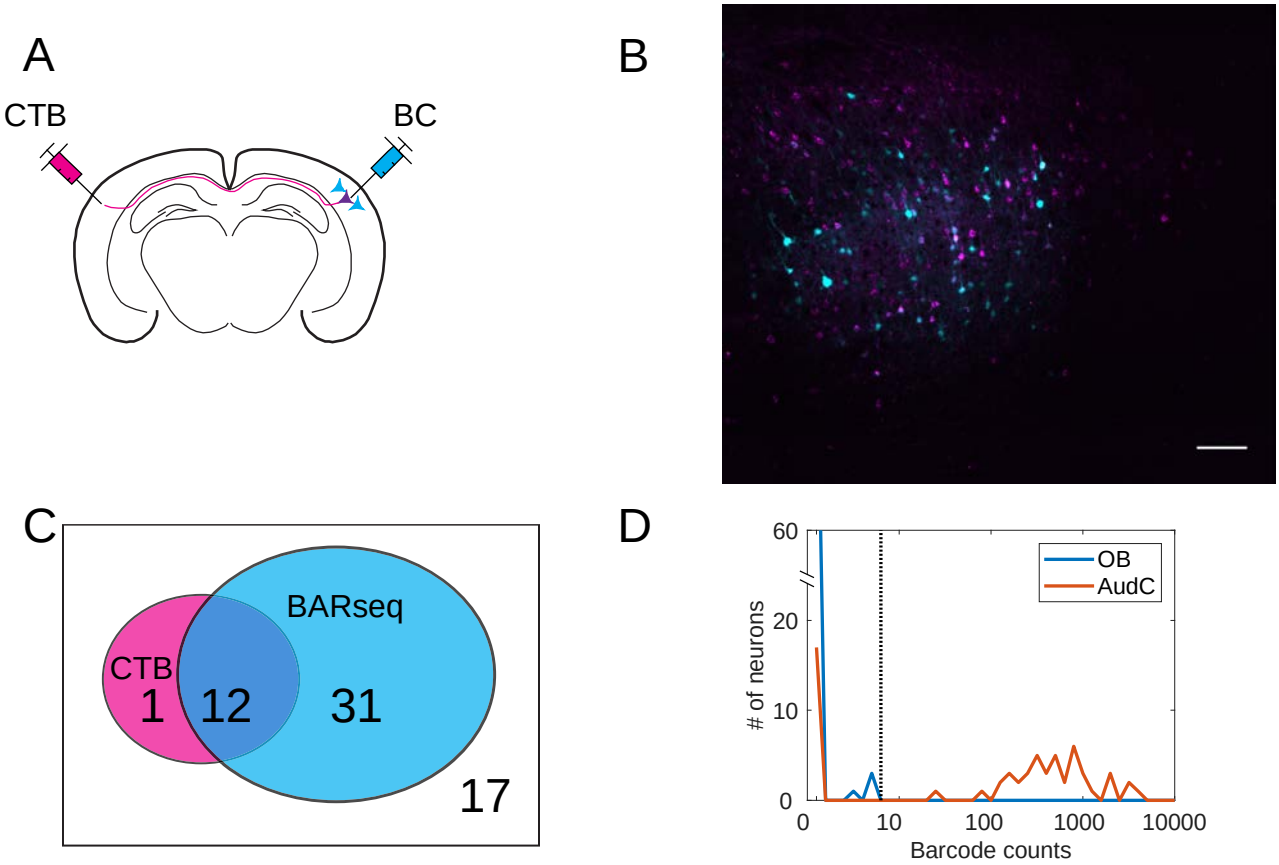
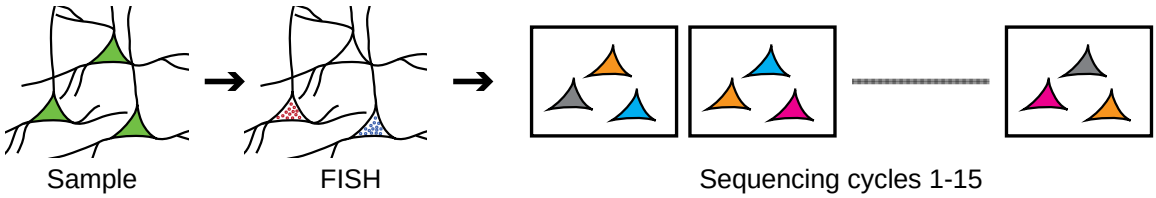
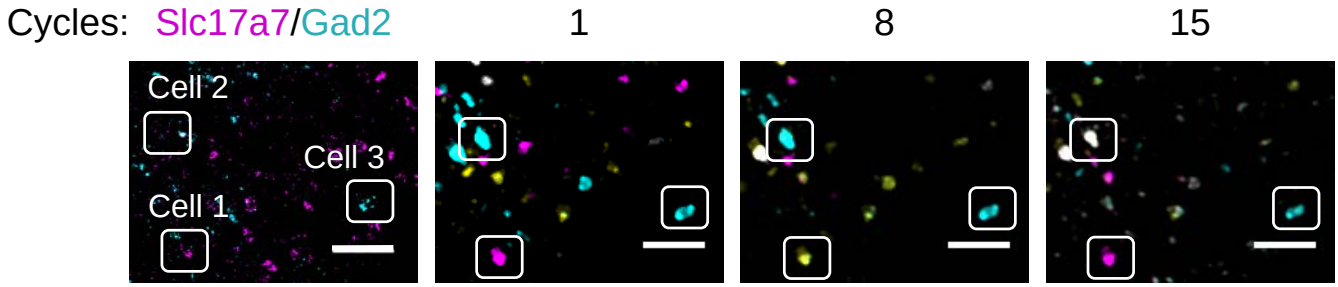


Figure 3, see also Table S1

A



B



C

	Slc17a7	Gad2	Barcode sequence
Cell 1:	+	-	A a t g a t a T c t a a t a A
Cell 2:	-	-	G t c g a a c G a t g t c t C
Cell 3:	-	+	G a a c t c g G t a t g t g G

Figure 4, see also Fig. S2 and Table S2

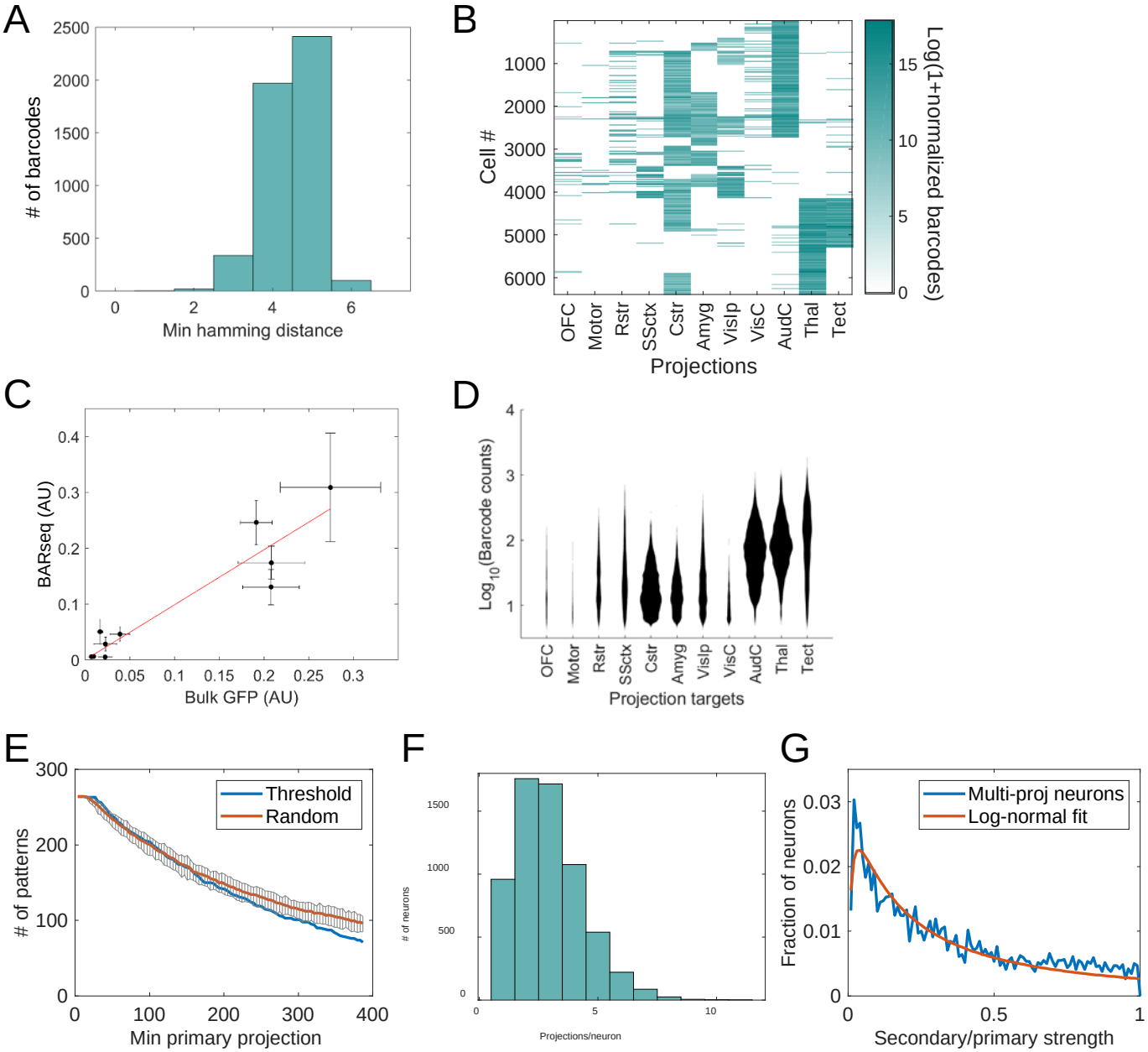


Figure 5, see also Table S3

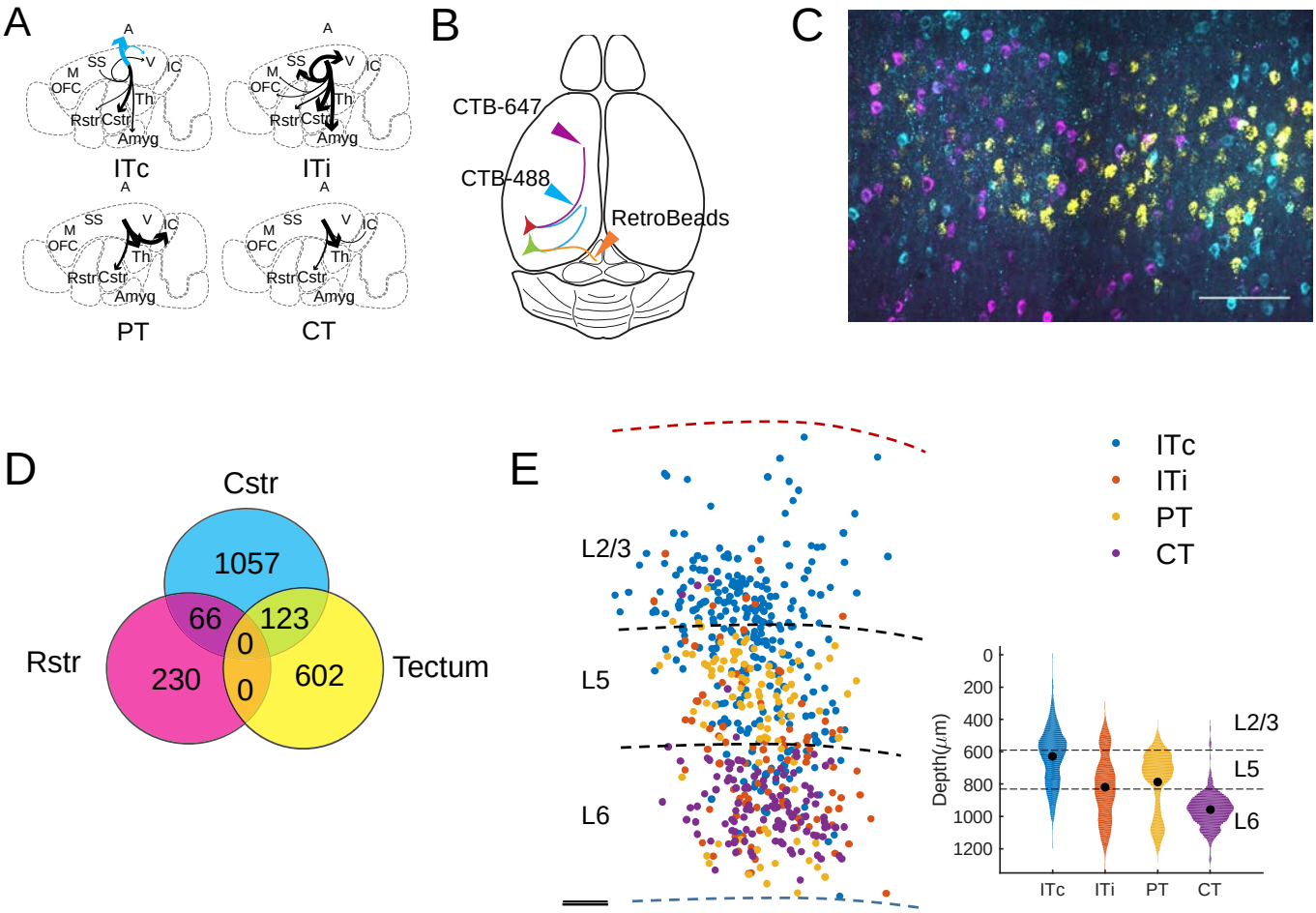

A
A
SS
V
IC
M
OFC
Th
Rstr
Cstr
Amyg
ITc
ITi
PT
CT
B
CTB-647
CTB-488
RetroBeads
C
D
Cstr
1057
66
123
0
Rstr
230
0
602
Tectum
E
L2/3
L5
L6
ITc
ITi
PT
CT
Depth(μm)
0
200
400
600
800
1000
1200
L2/3
L5
L6
ITc
ITi
PT
CT

Figure 6

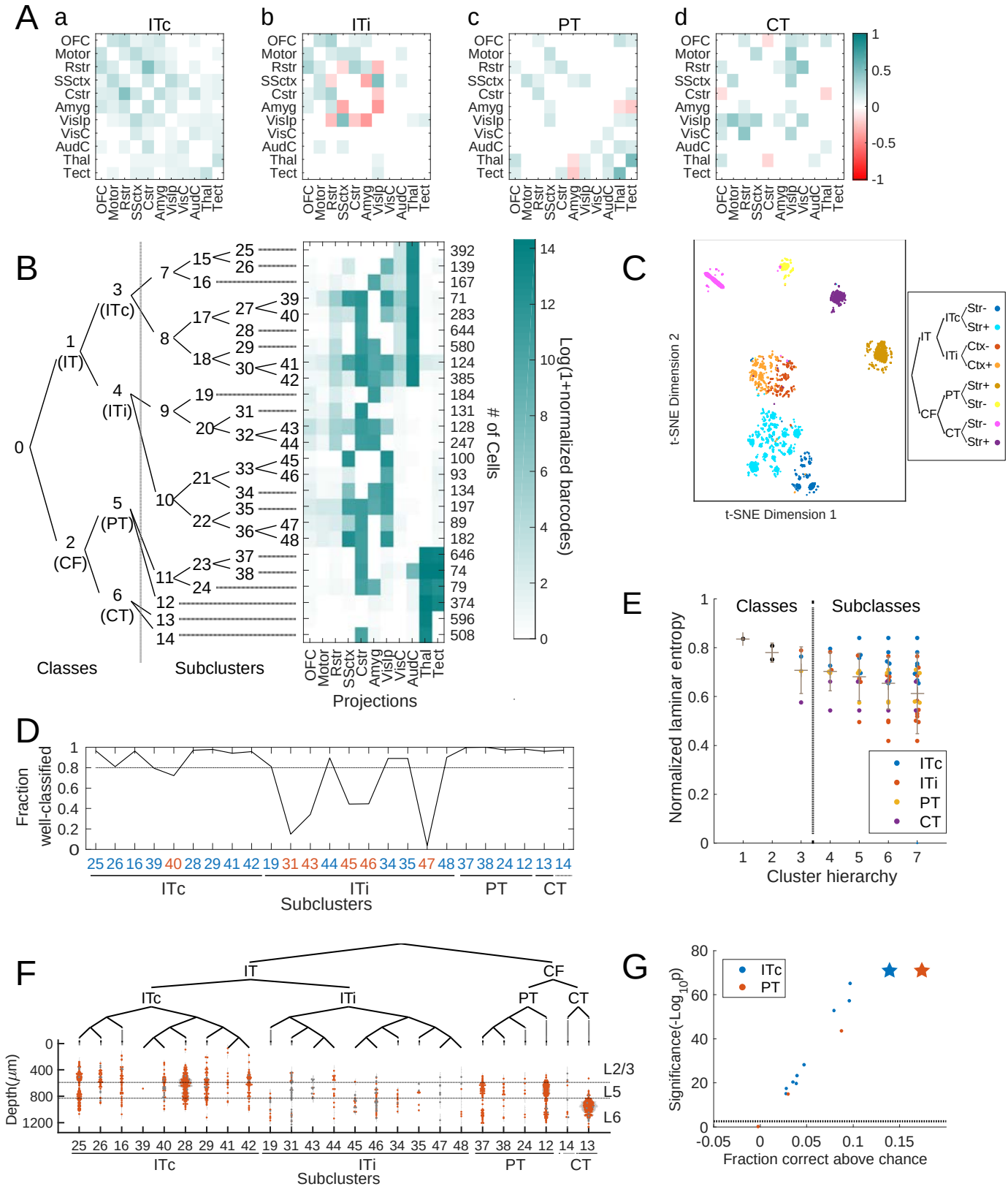


Figure 7, see also Fig. S4-S7

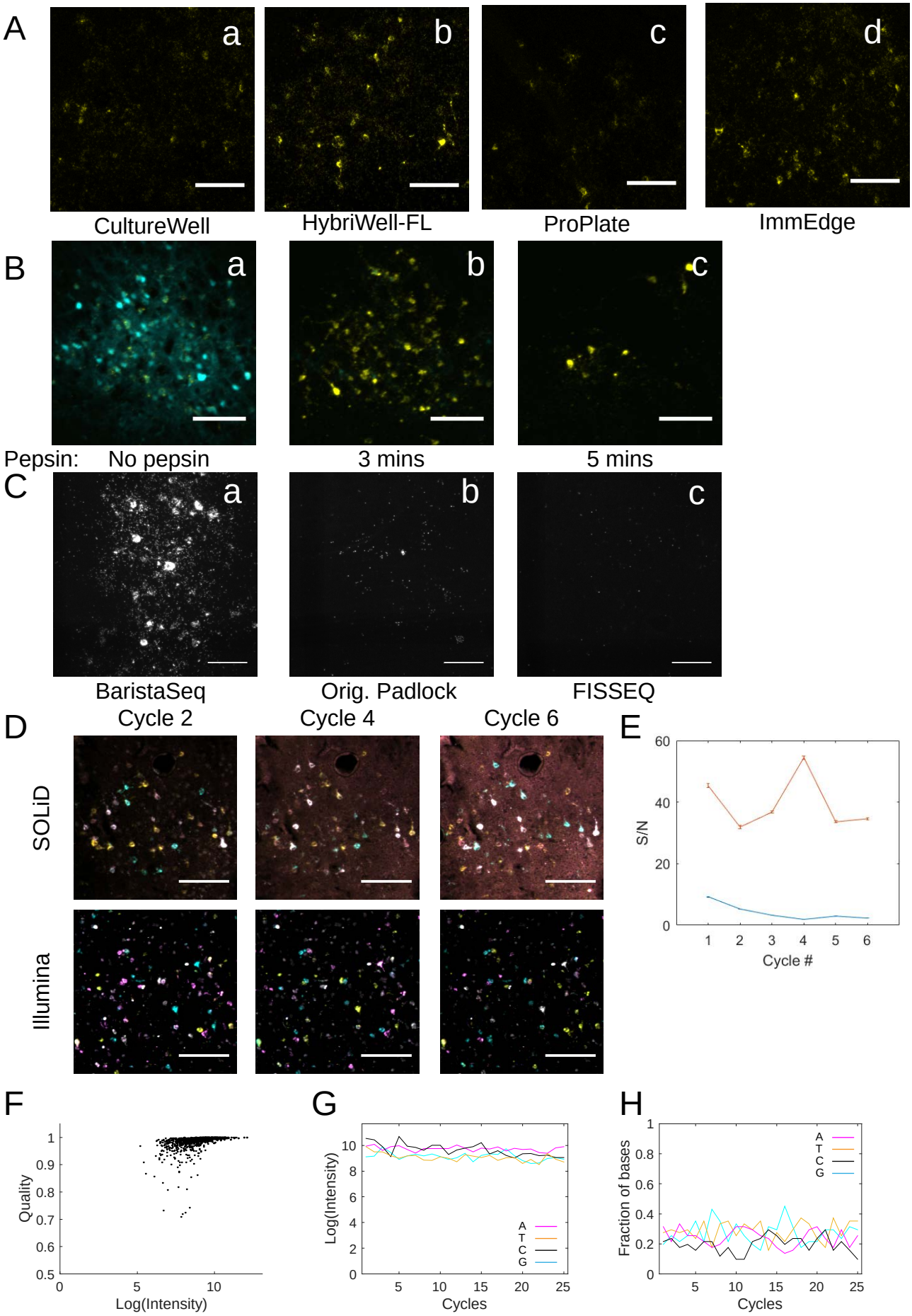


Fig. S1, related to Fig. 2 and STAR Methods

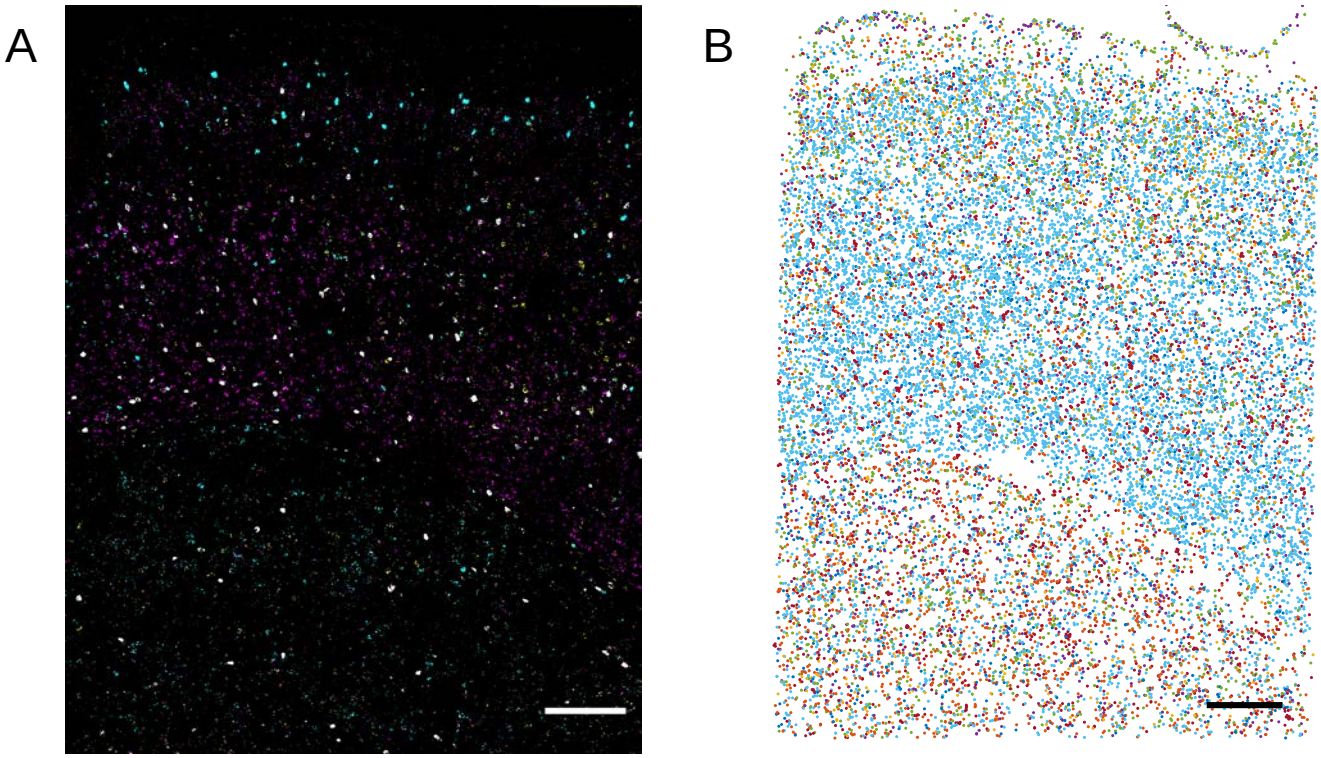
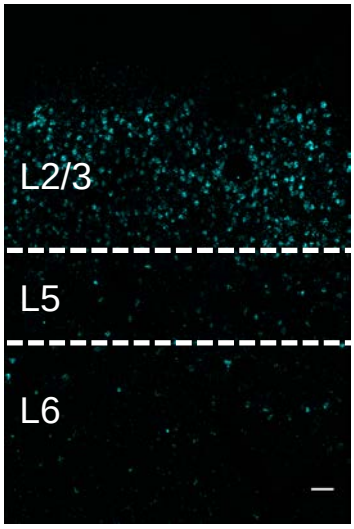
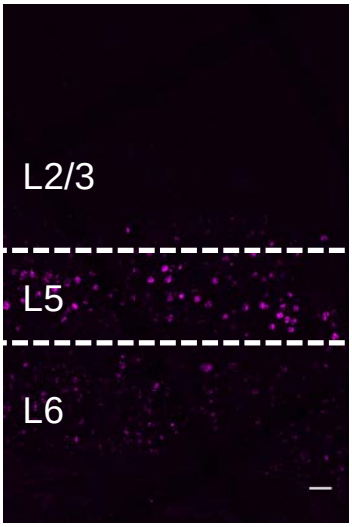


Fig. S2, related to Fig. 4

A



B



C

Slice#	1	2	3	Mean ± sem
L2/3 – L5 (Cux2)	640 µm	590 µm	550 µm	600 ± 20 µm
L2/3 – L5 (Fezf2)	550 µm	600 µm	590 µm	580 ± 10 µm
L5 – L6 (Fezf2)	800 µm	830 µm	870 µm	830 ± 20 µm

D

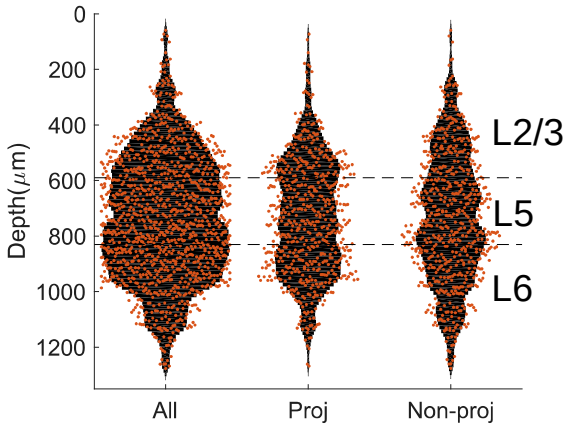


Figure S3, related to STAR Methods

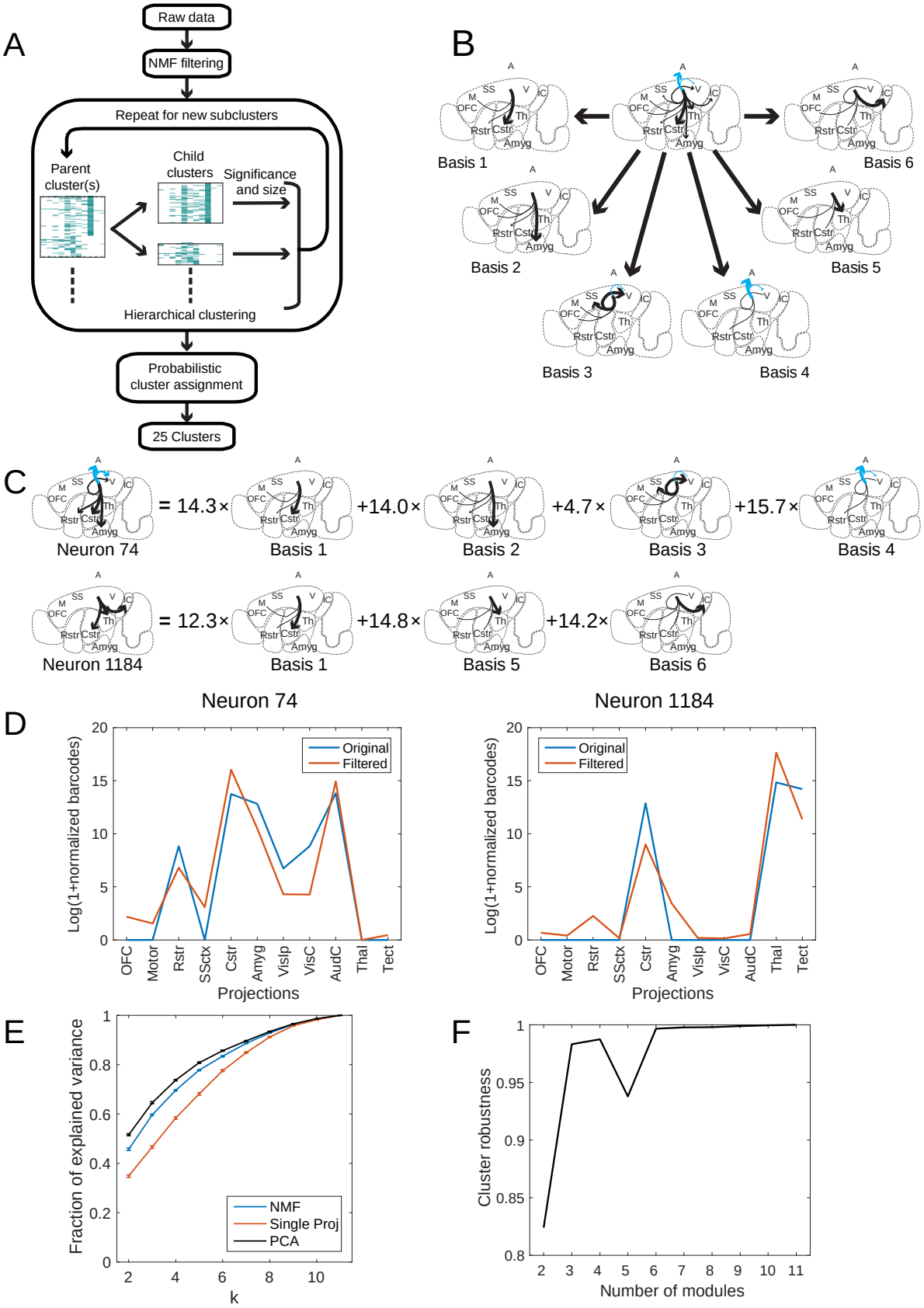


Figure S4, related to Fig. 7 and STAR Methods

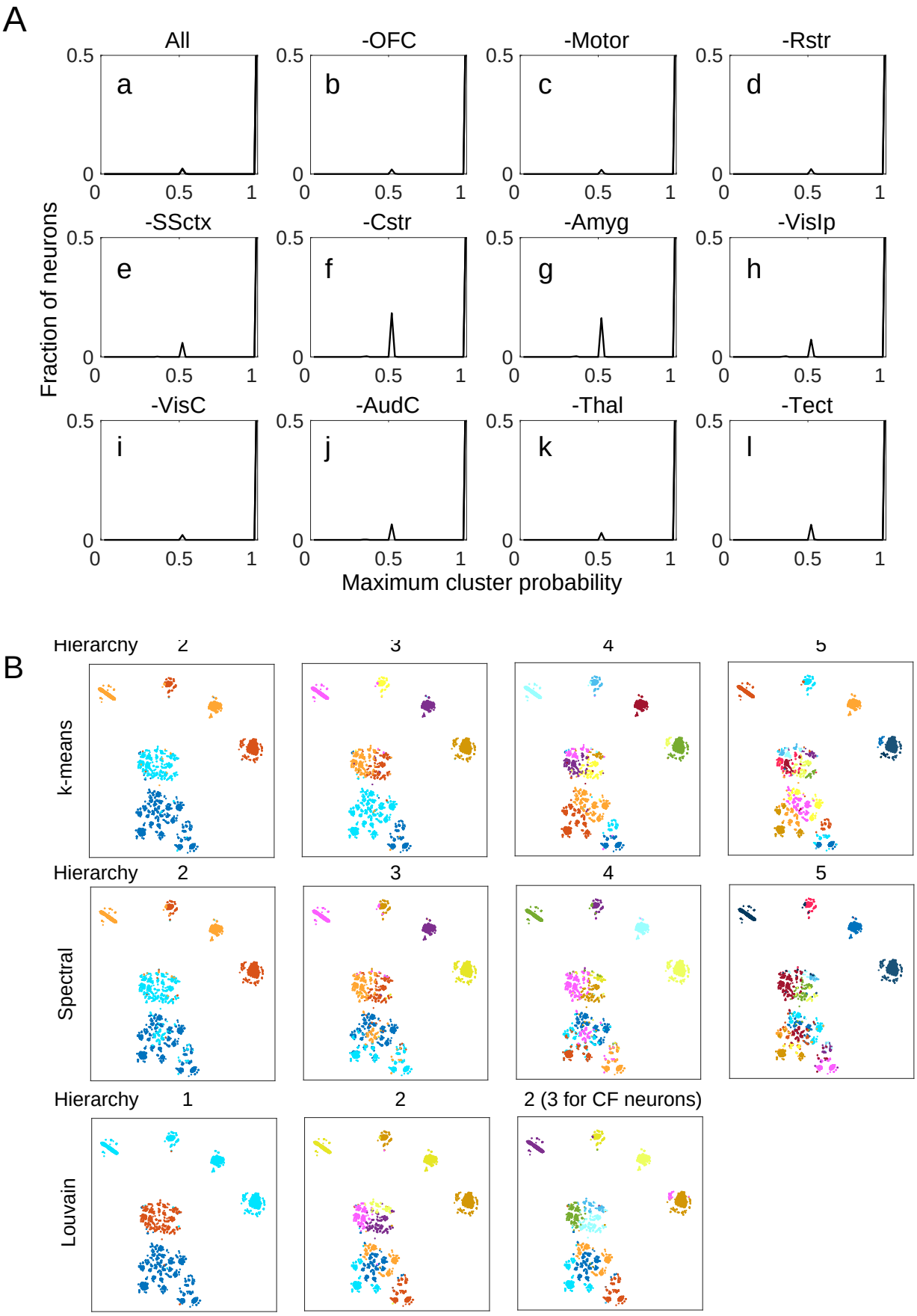


Figure S5, related to Fig. 7 and STAR Methods

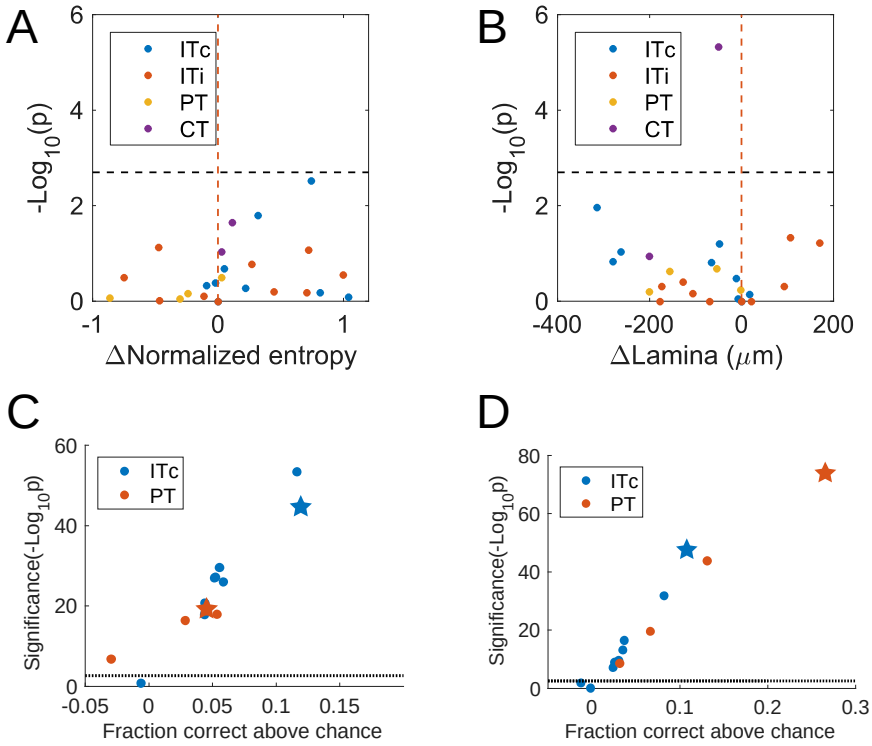


Figure S6, related to Fig. 7

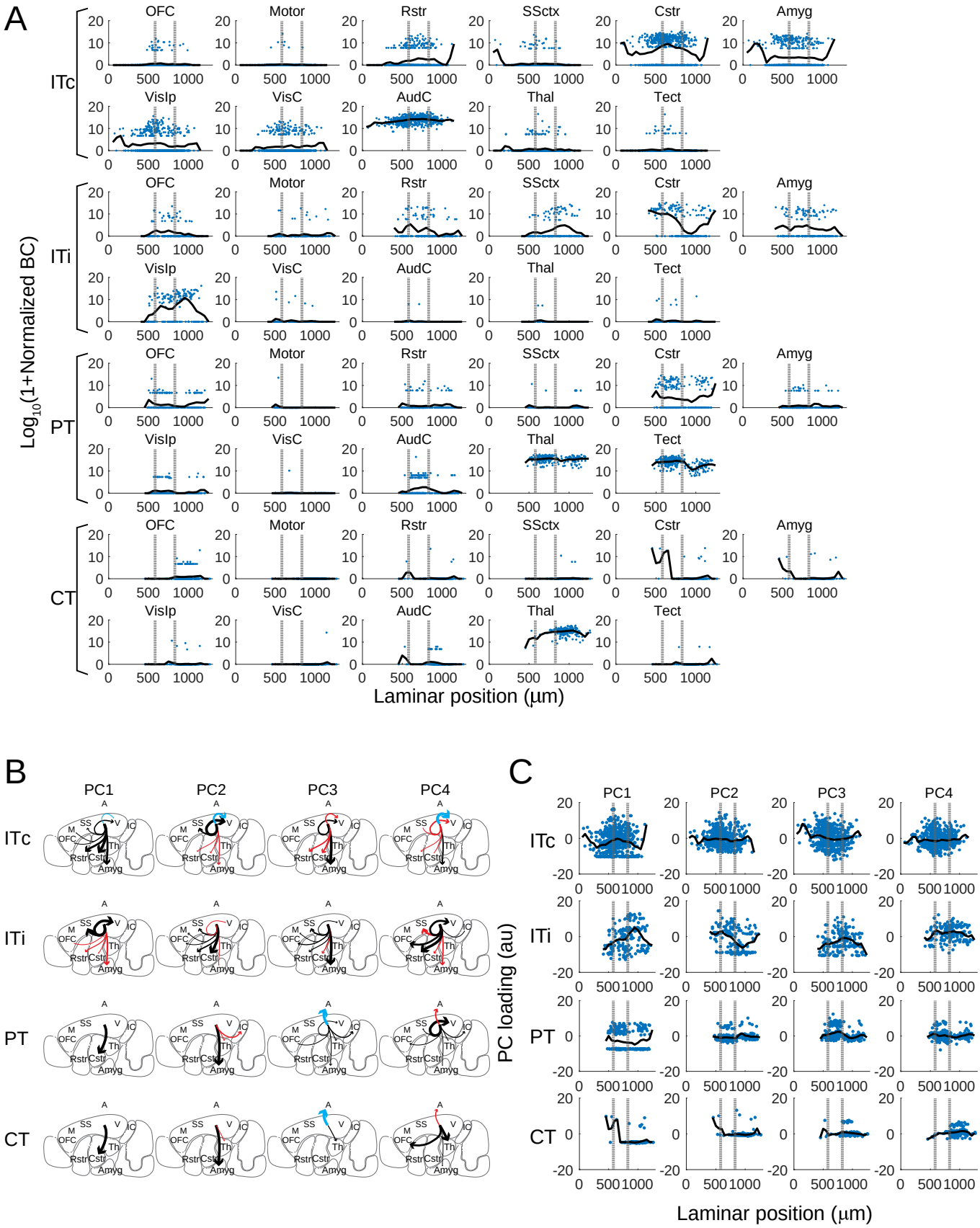


Figure S7, related to Fig. 7

OB	c1	c2	c3	min qual (>0.75)	CTB+	Barseq+
0	151	186	4	0.95	0	1
0	1433	995	87	0.96	0	1
0	155	70	11	0.91	0	1
0	177	197	17	0.89	0	1
0	17	102	1	0.92	1	1
0	132	135	10	0.88	0	1
0	96	157	12	0.88	0	1
0	102	58	6	0.90	0	1
0	665	58	14	0.82	0	1
0	14	100	5	0.87	1	1
0	0	0	0	0.91	0	0
0	0	0	0	0.97	0	0
0	126	112	7	0.84	1	1
3	281	359	28	0.88	0	1
0	419	313	34	0.94	1	1
0	106	139	3	0.79	0	1
0	72	89	15	0.88	1	1
0	342	159	4	0.88	0	1
2	1571	1061	100	0.92	0	1
0	0	0	0	0.79	0	0
0	0	0	0	0.86	0	0
0	0	0	0	0.83	0	0
0	232	51	5	0.89	0	1
0	0	0	0	0.92	0	0
0	372	344	19	0.87	1	1
0	0	0	0	0.88	0	0
0	44	20	1	0.83	0	1
0	234	209	20	0.95	1	1
0	109	10	6	0.90	0	1
3	281	359	28	0.91	0	1
0	61	66	4	0.95	0	1
0	0	0	0	0.82	0	0
0	0	0	0	0.95	0	0
0	335	79	15	0.92	0	1
0	114	186	9	0.94	0	1
0	0	0	0	0.90	1	0
0	271	211	21	0.91	0	1
0	3159	415	28	0.78	1	1
0	0	0	0	0.79	0	0
0	305	148	18	0.94	0	1
0	0	0	0	0.94	0	0
0	637	293	31	0.85	1	1
0	471	413	15	0.93	0	1
0	1062	528	94	0.92	1	1
0	335	79	15	0.92	0	1
0	59	73	3	0.86	0	1
0	565	359	27	0.93	0	1
0	460	208	82	0.94	0	1
0	0	0	0	0.83	0	0
0	0	0	0	0.84	0	0
0	0	0	0	0.96	0	0
0	208	69	5	0.91	1	1
0	0	0	0	0.91	0	0
0	1148	481	30	0.97	0	1
0	200	102	14	0.80	1	1
0	238	154	25	0.94	0	1
3	609	575	62	0.88	0	1
0	18	5	0	0.79	0	1
0	0	0	0	0.95	0	0
0	1748	163	31	0.92	0	1
0	0	0	0	0.96	0	0

Table S1, related to Figure 3

Cell Index	OB	Ipsi-Ctx	Contra-Ctx	Str	Thal	Tect	Total	Slc17a7	Gad2
1012	0	14	87	0	0	0	101	1	0
1020	0	115	119	0	0	0	234	0	0
1024	0	119	0	0	0	0	119	1	0
1029	0	0	0	0	0	0	0	1	0
1074	0	21	18	25	0	0	64	1	0
1075	0	0	0	0	0	0	0	0	0
1128	0	30	44	10	0	0	84	1	0
1135	0	28	3	0	0	0	31	1	0
1155	0	0	0	0	0	0	0	0	1
1167	0	0	0	0	0	0	0	1	0
1179	0	0	170	1	0	0	171	1	0
1192	0	0	0	0	0	0	0	0	0
1199	0	0	0	0	0	0	0	0	0
1226	0	0	0	0	0	0	0	0	0
1245	0	0	0	0	0	0	0	1	0
1267	0	0	0	0	0	0	0	0	0
1274	0	0	16	0	0	0	16	1	0
2007	0	4	0	0	0	0	4	1	0
2051	0	0	0	0	0	0	0	0	0
2054	0	0	0	0	0	1	1	1	0
3009	0	0	0	0	0	0	0	0	0
3012	0	41	0	0	0	0	41	1	0
3076	0	68	48	0	0	0	116	1	0
3082	0	9	182	0	0	0	191	1	0
3090	0	0	1	0	0	0	1	1	0
3130	0	0	74	0	0	0	74	1	0
4003	0	34	0	0	0	0	34	1	0
4005	0	0	0	0	0	0	0	1	0
4008	0	0	0	0	0	0	0	0	0
4010	0	253	10	5	0	0	268	1	0
4018	0	0	0	0	0	0	0	0	1
4022	0	51	51	16	0	0	118	1	0
4062	0	0	12	0	0	0	12	1	0
4066	0	0	64	1	1	0	66	1	0
4067	0	0	0	0	151	0	151	1	0
4068	0	11	24	0	0	0	35	1	0
4082	0	39	0	0	0	0	39	1	0
4088	0	0	0	0	0	0	0	0	0
4101	0	10	0	0	0	0	10	1	0
4112	0	0	0	0	0	0	0	0	0
4114	0	80	57	0	0	0	137	1	0
4129	0	0	0	0	0	0	0	1	0
4150	0	0	0	0	0	0	0	1	0
4163	0	20	4	0	0	0	24	1	0
4164	0	0	0	0	0	0	0	1	0
4165	0	0	0	0	0	0	0	1	0
4169	0	0	0	0	0	0	0	0	0
4204	0	0	90	0	0	0	90	1	0
4217	0	0	0	0	0	0	0	1	0
4224	0	0	47	0	0	0	47	1	0
4232	0	2	6	22	0	0	30	1	0

Cell Index	OB	Ipsi-Ctx	Contra-Ctx	Str	Thal	Tect	Total	Slc17a7	Gad2
4238	0	0	0	0	0	0	0	0	0
4241	0	0	105	0	0	0	105	1	0
4264	0	14	149	18	0	0	181	1	0
4269	0	0	5	0	0	0	5	1	0
5001	0	8	106	1	0	0	115	1	0
5035	0	4	114	0	0	0	118	1	0
5038	0	0	0	0	0	0	0	1	0
5076	0	0	80	0	0	0	80	1	0
5092	0	0	63	2	0	0	65	1	0
5096	0	0	0	0	0	0	0	1	0
5111	0	0	0	0	0	0	0	0	0
5112	0	0	0	0	0	0	0	0	0
5114	0	0	13	0	0	0	13	1	0
5144	0	0	3	0	0	0	3	1	0
5165	0	7	34	0	0	0	41	1	0
5166	0	0	0	0	0	0	0	1	0
5178	0	0	0	0	0	0	0	0	1
6003	0	0	0	0	0	0	0	1	0
6005	0	2	37	1	0	0	40	1	0
6030	0	0	22	19	0	0	41	1	0
6044	0	42	77	14	1	0	134	1	0
6051	0	0	0	0	0	0	0	1	0
6077	0	0	0	0	0	0	0	1	0
6143	0	0	68	0	0	0	68	1	0
6149	0	0	0	0	0	0	0	1	0
7006	0	0	0	0	0	0	0	0	0
7047	0	0	0	0	0	0	0	1	0
7049	0	0	0	0	1	0	1	1	0
7063	0	0	0	0	0	0	0	1	0
7070	0	0	4	0	0	0	4	1	0
7076	0	0	0	0	0	0	0	1	0
7083	0	0	0	0	0	0	0	1	0
7126	0	0	0	0	0	0	0	1	0
7136	0	0	1	0	0	0	1	1	0
7140	0	0	24	4	0	0	28	1	0
8025	0	36	19	4	0	0	59	1	0
9013	0	0	0	0	0	0	0	1	0
9024	0	0	0	0	0	0	0	1	0
9057	0	116	0	1	2	0	119	1	0
9058	0	14	21	0	0	0	35	1	0
9063	0	0	0	0	0	0	0	1	0
9067	0	12	9	15	0	0	36	1	0
9073	0	8	0	0	0	0	8	1	0
9078	0	106	0	0	0	0	106	1	0
9083	0	0	0	0	0	0	0	0	0
9085	0	0	0	0	0	0	0	1	0
11012	0	20	10	0	0	0	30	1	0
11025	0	0	0	0	0	0	0	1	0
11032	0	6	117	0	2	6	131	1	0

Table S2, related to Fig. 4

	BARseq		MAPseq
Brain ID	XC9	XC28	XC14*
Proj barcode #	4841	13581	8418
# of cells in ACx	1575	1662	13998
# of cells projecting	895	911	5082
# of filtered cells	605	704	5082

Table S3, related to Figure 5

Table S4, related to STAR Methods

See TableS4.xls