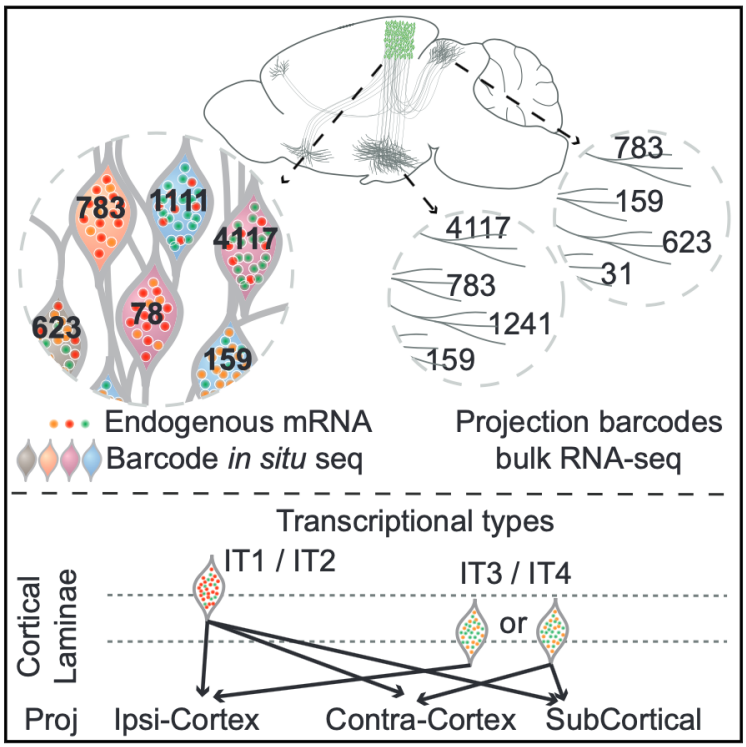
Resource

Cell

High-Throughput Mapping of Long-Range Neuronal Projection Using *In Situ* Sequencing

Graphical Abstract



Highlights

- BARseq uses *in situ* sequencing to map neuronal projections with high throughput
- BARseq correlates neuronal projections to gene expression and *Cre*-labeling
- BARseq recapitulates known organization of projections in auditory cortex
- BARseq reveals distinct projections of transcriptionally defined IT subtypes

Authors

Xiaoyin Chen, Yu-Chi Sun, Huiqing Zhan, ..., Z. Josh Huang, Jesse Gillis, Anthony M. Zador

Correspondence

zador@cshl.edu

In Brief

BARseq is a high-throughput, multiplexed method based on RNA barcoding that helps bridge anatomical and transcriptomic approaches at cellular resolution with the potential to discover organizing principles of neural circuits as exemplified by the uncovering of distinct, transcriptionally defined subtype projections in the mouse auditory cortex.

Chen et al., 2019, Cell *179*, 772–786

Cell

Resource

High-Throughput Mapping of Long-Range Neuronal Projection Using *In Situ* Sequencing

Xiaoyin Chen,[1] Yu-Chi Sun,[1] Huiqing Zhan,[1] Justus M. Kebschull,[1,2] Stephan Fischer,[1] Katherine Matho,[1] Z. Josh Huang,[1] Jesse Gillis,[1] and Anthony M. Zador[1,3,*]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[2]Department of Biology, Stanford University, Stanford, CA 94305, USA
[3]Lead Contact
*Correspondence: zador@cshl.edu
https://doi.org/10.1016/j.cell.2019.09.023

SUMMARY

Understanding neural circuits requires deciphering interactions among myriad cell types defined by spatial organization, connectivity, gene expression, and other properties. Resolving these cell types requires both single-neuron resolution and high throughput, a challenging combination with conventional methods. Here, we introduce barcoded anatomy resolved by sequencing (BARseq), a multiplexed method based on RNA barcoding for mapping projections of thousands of spatially resolved neurons in a single brain and relating those projections to other properties such as gene or *Cre* expression. Mapping the projections to 11 areas of 3,579 neurons in mouse auditory cortex using BARseq confirmed the laminar organization of the three top classes (intratelencephalic [IT], pyramidal tract-like [PT-like], and corticothalamic [CT]) of projection neurons. In depth analysis uncovered a projection type restricted almost exclusively to transcriptionally defined subtypes of IT neurons. By bridging anatomical and transcriptomic approaches at cellular resolution with high throughput, BARseq can potentially uncover the organizing principles underlying the structure and formation of neural circuits.

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 physiological properties. Understanding how diverse neurons are organized thus requires methods that can map various neuronal characteristics at cellular resolution, with high-throughput, in single brains.

A particular challenge is to map the long-range axonal projection patterns that form the physical basis for neuronal circuits across brain areas. Traditional neuroanatomical methods based on [illegible] 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; Lin et al., 2018; Zheng et al., 2013). Furthermore, these methods usually rely on dedicated imaging platforms specifically designed for neuronal tracing, which are not widely accessible to many laboratories.

To allow high-throughput projection mapping, we previously developed multiplexed analysis of projections by sequencing (MAPseq) (Han et al., 2018; Kebschull et al., 2016a) (Figure 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 the precise location of each soma is lost. Thus, with MAPseq, it is difficult to combine information about projections with other types of information, such as gene expression and neuronal activity.

Here, we describe BARseq (barcoded anatomy resolved by sequencing) (Figure 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 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-like (PT-l) (we call these neurons pyramidal tract-like because, unlike pyramidal tract neurons in the motor cortex, these neurons in the auditory cortex do not project to the spinal cord), and intratelencephalic (IT) neurons (Harris and Shepherd, 2015; Shepherd, 2013)—in the mouse auditory cortex and can additionally uncover previously unknown organization beyond these major classes. BARseq thereby complements traditional neuro-anatomical approaches by providing a high-throughput method for linking axonal projections, spatial organization, gene expression, and potentially other neuronal properties that define neuronal cell types and circuits.

RESULTS

In what follows, we first demonstrate robust *in situ* sequencing of [illegible]

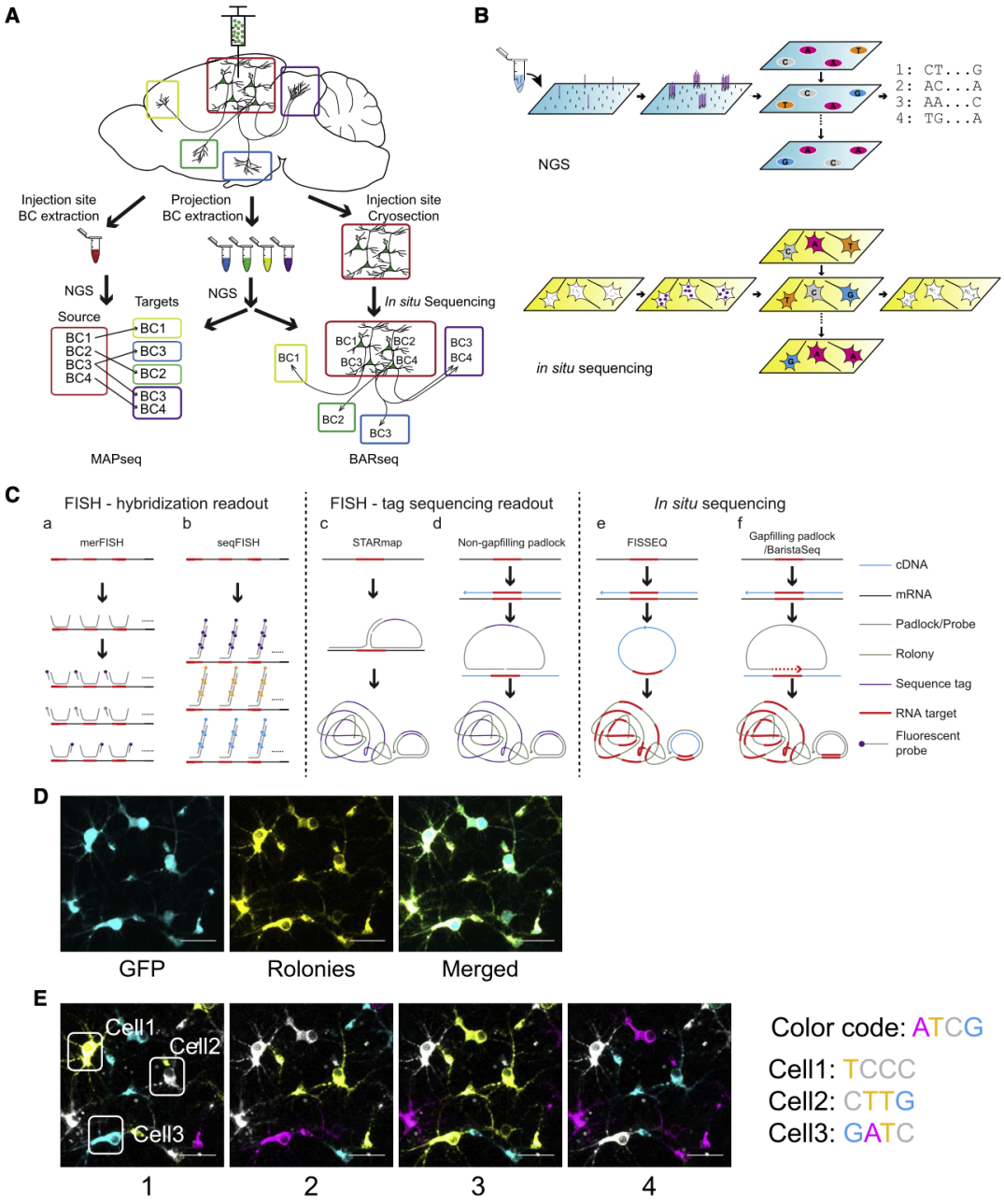


Figure 1. Multiplexed Projection Mapping Using *In Situ* Sequencing

(A) Workflow of MAPseq (left) and BARseq (right).

(B) NGS versus *in situ* sequencing.

(C) Comparison of *in situ* sequencing and hybridization techniques. These techniques may use multiple rounds of hybridization to probe and read out multiple mRNAs (a and b), use sequencing to multiplex read out of hybridization signals (c and d), or copy target sequences from the mRNA into the rolonies to allow true sequencing (e and f).

(D) Representative images of barcode rolonies generated in primary hippocampal neuronal culture coexpressing GFP. All GFP positive neurons were filled with barcode amplicons, indicating efficient barcode amplification in neuronal somata. Scale bars, 50 μm.

(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 are indicated to the right. In all images, scale bars, 50 μm.

Cell

(MAPseq). We next combine BARseq with fluorescence *in situ* hybridization (FISH) and subpopulation-targeted *Cre* labeling to correlate neuronal projections with gene expression and fluorescent labeling. Finally, we apply BARseq to understand the organization of projection neurons in the auditory cortex and identify a structured organization of projections beyond previously defined 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., multiplexed error-robust fluorescence *in situ* hybridization [merFISH] [Chen et al., 2015)], sequential fluorescence *in situ* hybridization [seqFISH] [Eng et al., 2019; Shah et al., 2016], STARmap [Wang et al., 2018], *in situ* sequencing [Ke et al., 2013], and cyclic single molecule fluorescent *in situ* hybridization [osmFISH] [Codeluppi et al., 2018]), these methods lack single-nucleotide specificity, and can distinguish only up to 10^2–10^4 different transcript sequences, each with a length of 10^2–10^3 nt. By contrast, cellular barcodes used for projection mapping are much shorter (30-nt), more diverse ($\sim10^6$ to 10^7) and may differ from each other by just a few nucleotides. Cellular barcodes thus cannot be distinguished using currently available multiplexed FISH approaches. Spatial transcriptomics (Rodriques et al., 2019; Ståhl et al., 2016) can read out diverse short barcodes, but at present, the spatial resolution is insufficient to resolve single neurons. Laser micro-dissection combined with RNA sequencing (RNA-seq) 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 (Figure 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 (Figures 1Ca–1Cf). Although some FISH 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 (Figures 1Cc and 1Cd). Circligase-based fluorescent *in situ* RNA sequencing (*FISSEQ*) (Figure 1Ce) can readout mRNA sequences but is not sufficiently sensitive for barcode sequencing. We therefore used a targeted approach, which padlock probe hybridization followed by gap-filling and ligation was used to form a circular template for rolling circle amplification (Figure 1Cf). The gap-filling padlock approach allowed for efficient amplification of barcodes in neurons co-expressing barcodes and GFP (Figure 1D). We also adapted Illumina sequencing by synthesis (SBS) chemistry for *in situ* reactions. In combination, these strategies formed the basis for a protocol that allowed efficient and robust sequencing of barcoded cultured neurons (Figure 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 that enabled convenient handling and was compatible with the amplification reactions (Figure S1A; STAR Methods). Because barcoded neurons express high levels of GFP (that interferes with imaging during sequencing), we developed a protocol for tissue digestion to reduce the GFP signal from barcoded cells (Figure S1B; STAR Methods) and to increase RNA accessibility. Together these optimizations resulted in highly efficient amplification of barcodes (Figure 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) (Figures S1D and S1E; STAR Methods). These modifications resulted in highly efficient and specific barcode amplification (Figure 2A) and sequencing (Figure 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, 2016b) injected in the auditory cortex (Figure 2C). Basecall quality (Figures 2D and S1F; STAR Methods) and signal intensity (Figure S1G) remained high through all 25 cycles. We observed no bias toward any particular base (Figure 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 (Figures 2E and 2F). 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 ($\sim10^{-9}$) 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. In MAPseq (Figure 1A, left), neurons are barcoded using a Sindbis virus library. 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

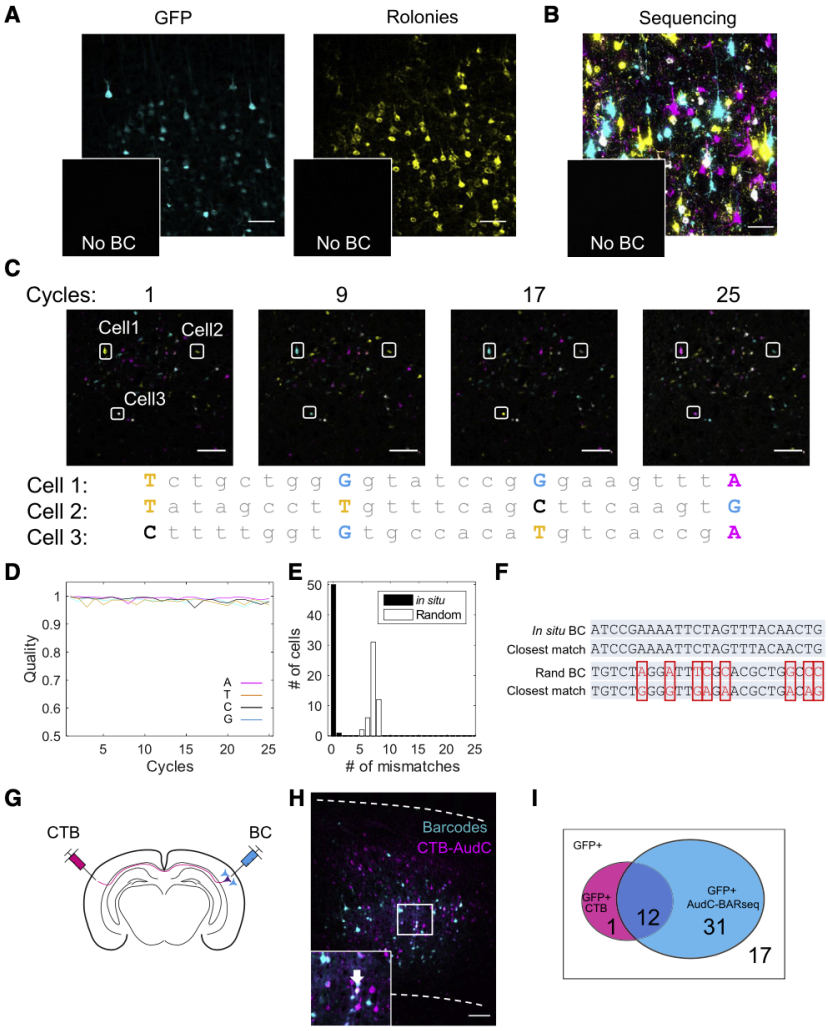


Figure 2. Validation of BARseq

(A) Representative low-magnification images of GFP (left) co-expressed with barcodes and rolonies (right) in a brain slice. GFP intensity does not correlate perfectly with rolony intensity due to differences in protein and RNA expression. Insets: Negative control images of GFP and rolonies of a non-barcoded brain slice taken with the same exposure settings. No GFP or rolonies are visible in these images. Scale bars, 50 μm.

(B) Representative high-resolution sequencing image of a barcoded brain slice. Scale bar, 50 μm.

(C) Low-resolution images of the indicated cycles of barcode sequencing in a brain slice. The sequences of the three cells are indicated below. Scale bars, 100 μm.

(D) The quality of the base calls on the barcoded brain slice.

(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.

(G) A brain was injected with CTB in the contralateral auditory cortex and barcoded in the ipsilateral auditory cortex. BARseq results of the barcoded neurons were then compared to retrograde labeling by CTB.

(H) A representative image of a brain slice double labeled with barcodes (cyan) and CTB (magenta) from the contralateral auditory cortex. Dashed lines indicate the top and the bottom of the cortex. Scale bar, 100 μm. A magnified view of the squared area is shown in the inset. The arrow in the inset indicates a GFP^+ CTB^+ double-labeled neuron.

(I) 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).

See also Figure S1 and Table S1.

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 (Figure 1A, right). BARseq thus inherits the throughput and cubelet resolution of projections of MAPseq but allows the precise somatic origin of each axonal projection to be determined with cellular 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 ($\sim 10^7$ barcodes) viral library into the auditory cortex and CTB in the contralateral auditory cortex (Figures 2G and 2H). We used BARseq to identify contralaterally projecting neurons and the precise locations of their somata. We then identified barcoded somata that neurons were detected by BARseq, whereas only 28% (12/43) of contralaterally projecting neurons identified by BARseq were labeled with CTB (Figure 2I; 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; Figure S1I), 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 relate pro-

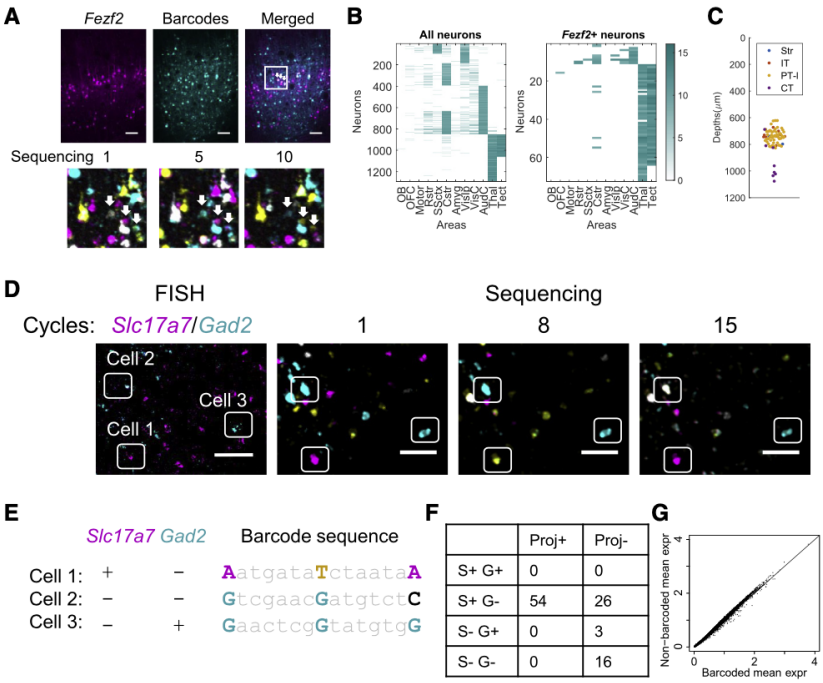


Figure 3. Correlating Gene Expression and Projections Using BARseq

(A) Barcode sequencing in *Cre*-labeled animals. Top: images of *Fezf2*$^+$ tdTomato expressing cells (left), GFP expressing barcoded cells (middle), and merged image of the two (right). Scale bar, 100 μm. Bottom: sequencing images of the boxed area in the merged image. Arrows indicate tdTomato-expressing barcoded neurons.

(B) Projection patterns of all neurons (left) and *Fezf2*$^+$ neurons (right). Rows indicate neurons and columns indicate projection areas.

(C) The cortical depth distribution of barcoded *Fezf2*$^+$ somata color-coded by projections.

(D) Representative images of FISH (left) and sequencing images (right) of the same sample. Scale bars, 50 μm.

(E) The expression profile of *Slc17a7* and *Gad2* and barcode sequences of the neurons boxed in (D).

(F) The number of neurons with or without projections (Proj) in cells that did or did not express *Slc17a7* (S) and/or *Gad2* (G).

(G) Mean log normalized expression of each gene averaged over all barcoded (x axis) and non-barcoded (y axis) neurons. The gene expression is regressed with a Poisson model to remove the effect of both the percentage of mitochondrial genes and endogenous UMI counts.

See also Figure S2 and Table S2.

To demonstrate BARseq in Cre-labeled neuronal subpopulations, we performed BARseq in the auditory cortex in a *Fezf2*-2A-*CreER*::Rosa-CAG-LSL-tdTomato (Ai14) mouse induced at age 6 and 7 postnatal weeks (Figure 3A). *Fezf2* is predominantly expressed in layer 5 PT-l, but is also expressed at lower levels in some layer 5 IT neurons and layer 6 CT neurons (Tasic et al., 2018). We sequenced 1,291 projection neurons with good sequencing quality *in situ*, including 72 neurons co-labeled by tdTomato, and assayed their projection patterns to 11 target areas (STAR Methods). The *Fezf2*$^+$ population included 53 (74%) PT-l neurons projecting to the tectum and the thalamus, eight (11%) CT neurons projecting to the thalamus only, nine (13%) IT neurons in layer 5 projecting to the contralateral cortex, and two (3%) neurons in layer 5 projecting to the striatum (Figures 3B and 3C). The lack of corticotectal projection in the callosal projecting neurons is unlikely an artifact caused by limited detection sensitivity of corticotectal projections, because these neurons also project to other telencephalic areas that PT-l neurons do not. Neurons labeled with *Fezf2*-2A-*CreER* thus consisted mainly of PT-l neurons, along with a small fraction of callosally projecting neurons in layer 5 and CT neurons in layer 6. These results demonstrate the combination of BARseq with Cre-dependent labeling of neuronal subpopulations and further validate the results of BARseq.

To correlate projections with the expression of multiple genes, BARseq can be combined with *in situ* detection of genes in the same sample. To demonstrate the feasibility of resolving both barcodes (for projection mapping) and endogenous gene expression *in situ*, we combined BARseq with FISH against *Slc17a7* (a marker of excitatory neurons) (Tasic et al., 2016) and *Gad2* (a marker of inhibitory neurons) in the same cells (Figure

neurons in the cortex are excitatory, we identified 54 neurons with long-range projections, all of which expressed *Slc17a7*, but not *Gad2* (Figure 3F; Table S2; see STAR Methods for details). The projection neurons identified by BARseq were thus consistent with their transcriptional cell types. These experiments demonstrate that projection mapping with BARseq is compatible with *in situ* interrogation of gene expression.

We next assessed whether barcoding disrupted the transcriptomic landscape of single neurons. We performed single-cell RNA-seq in 398 barcoded and 1,088 non-barcoded cells dissociated from the mouse auditory cortex. Although fewer endogenous mRNA molecules were recovered from barcoded compared to non-barcoded neurons, the relative levels of endogenous gene expression remained largely unchanged in barcoded cells (Figures 3G and S2A; see STAR Methods for details). The fact that relative levels were preserved is compatible with a model in which infection with the barcoded viral vector affects the expression of most endogenous genes about equally. Importantly, top principal components of genes identified in the non-barcoded neurons were equally effective in describing gene expression in barcoded neurons (Figure S2B). These principal components contained many known neuronal subtype markers (Figure S2C). These results indicate that the expression pattern of genes, especially those that contribute to cell typing, remain largely intact in barcoded cells.

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

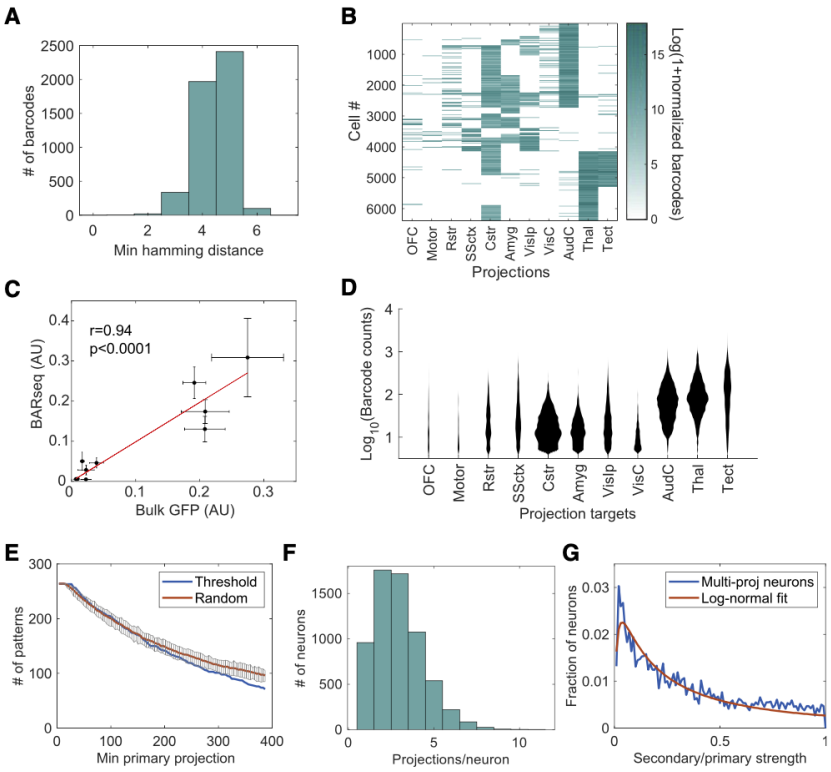


Figure 4. Mapping Projections of the Auditory Cortex Using BARseq

(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 clusters.

(C) Conventional bulk GFP tracing intensities (x axis) were plotted against the bulk projection strength obtained from BARseq (y axis). 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 indicates the number of cells.

(E) The numbers of binarized projection patterns (y axis) after filtering for primary projection strength (x axis, blue line) or after random subsampling to the same sample size (red line). 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 Figure S3 and Table S3.

barcodes to the projection barcodes (Figures 4A and S3A; STAR Methods),and found 1,806 (55%) of 3,237 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 1,309 neurons sequenced *in situ* and 5,082 neurons sequenced *in vitro* using conventional MAPseq (Figure 4B; Table S3). 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 most 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 6,391 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 5/6,391 = 0.08%. Using the false-positive rate in our dataset might be false positives, corresponding to a false discovery rate of 0.2% (i.e., for every 1,000 detected projections, two were 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-l 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-l neurons from neighboring areas that did not project to the tectum, and may be misidentified as

Cell

and thus can be used to evaluate fibers of passage contamination in the striatum. Consistent with the fact that PT-l 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-l 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 reinforce previous observations that fibers of passage have minimal impact on projections mapped in these 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) (Figure 4C; Pearson correlation coefficient $r = 0.94$, $p < 0.0001$; STAR Methods), projections from the auditory cortex to the thalamus, the tectum, the contralateral auditory cortex, and the caudal striatum were particularly strong (Figure 4D). 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} = 2,048 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 (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 (STAR Methods), and eliminating neurons with weaker primary projections resulted in only a moderate reduction in the diversity of projection patterns (Figure 4E, blue line). Furthermore, this reduction was largely due to the reduction in sample size (Figure 4E, 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 (Figure 4F) after binarization. The ratio of the projection strengths of the two strongest projections follow a log normal distribution (Figure 4G), and approximately 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 sparse and separate axonal arborizations in the auditory cortex revealed by single-neuron reconstruction (Han et al., 2018) and high-throughput mapping in the visual cortex (Han et al., 2018), and may be a general feature of cortical projections.

BARseq Recovers Structured Projections across Major Classes of Projection Neurons

Simple hierarchical clustering on the non-binarized projections (STAR Methods) partitioned neurons into established (Harris and Shepherd, 2015) cortical classes: CT, PT-l, and IT (Figure 5A). CT neurons project to the thalamus; PT-l neurons project to both the thalamus and the tectum; and 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.

The classification of cortical excitatory projection patterns into PT-l, 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-l and CT neurons (Harris and Shepherd, 2015) (Figure 5B; see STAR Methods and Figures S3B–S3E 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 previous observations using conventional methods (Harris and Shepherd, 2015).

Our data revealed two interesting features of the PT-l neurons in the auditory cortex. First, 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-l neurons projected only to the caudal striatum. We confirmed this finding by triple retrograde tracing (Figure S4A): 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 (Figures S4B and S4C). Second, we identified a minor population of PT-l neurons in layer 6 of the auditory cortex in addition to the more common layer 5 PT-l neurons (Figure 5B). This population of layer 6 PT-l neurons has previously been reported to differ from the layer 5 PT-l neurons based on physiological and other properties (Slater et al., 2013). Similar to the differences observed between the layer 5 and layer 6 PT-l neurons, the corticotectal projections were slightly stronger in layer 5 than those in layer 6, but with considerable overlap (Figure S4D [PT-Tect]; 11.7 ± 2.2 [mean ± SD], $n = 80$ for layer 6 neurons and 13.8 ± 1.6, $n = 185$ for layer 5 neurons, $p < 0.0001$ using bootstrap Kolmogorov-Smirnov [K-S] test). The strengths of other major projections were indistinguishable between the two groups (Figure S4D [PT], $p = 0.5$ for both corticothalamic projections and corticostriatal projections using bootstrap K-S test). The results of BARseq were thus consistent with, but went beyond, classical taxonomy.

Projection Subclusters Do Not Segregate by Laminae

Even within the major classes, projections from the auditory cor-

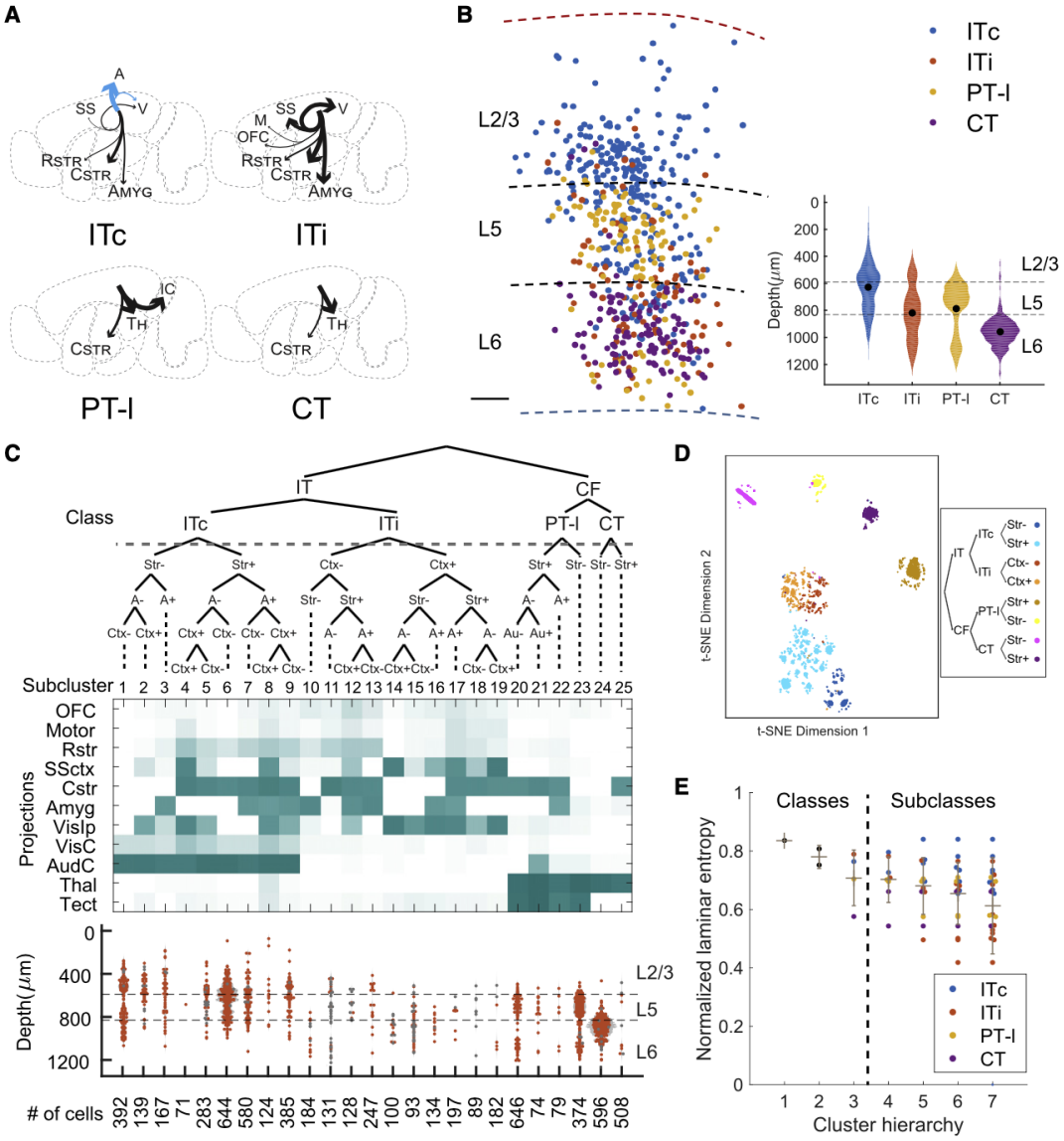


Figure 5. The Laminar Organization of Projection Neurons in the Auditory Cortex

(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) 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 the laminar depths of each class of projection neurons in the pooled BARseq dataset.

(C) Hierarchical clustering of single-cell projection data. Top: dendrogram of the hierarchical structure of the clusters. Middle: the mean projection patterns of the corresponding leaf clusters. Bottom: The laminar distribution of the corresponding leaf clusters. Individual neurons are superimposed on top of the distribution plots (light gray). Neurons whose cluster identity were less confident were marked in gray. The number of cells that belong to each leaf cluster is indicated below. Neurons of subcluster 25 were likely misidentified PT-I neurons (STAR Methods).

(D) T-distributed stochastic neighbor embedding (t-SNE) plot of the projection neurons. The neurons are color-coded by their first level subcluster identities post hoc.

(E) The normalized entropy of nodes/leaves (y axis) in the indicated clustering hierarchy (x axis). Grey bars indicate mean ± SD of all nodes/leaves of a specific hierarchy.

See also Figures S4, S5, and S6.

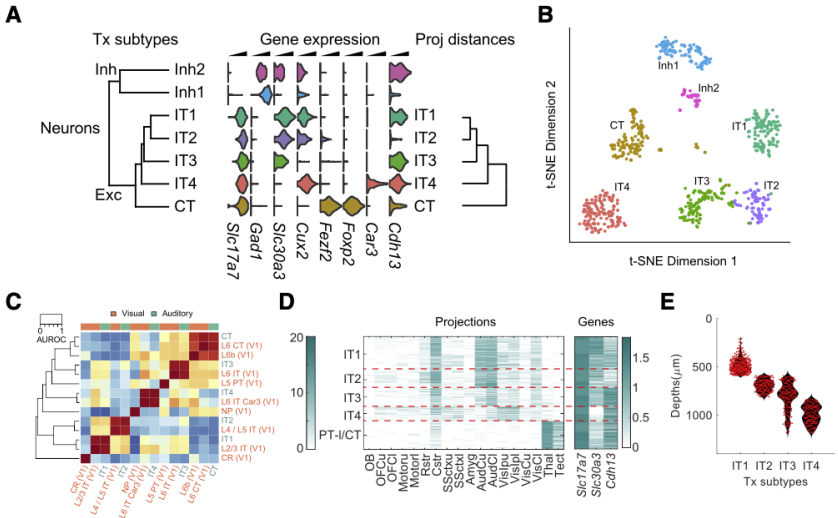


Figure 6. Subtypes of IT Neurons Defined by Gene Expression in the Auditory Cortex

(A) Histograms of the log normalized expression of the indicated marker genes in the indicated clusters obtained from single-cell RNA-seq in the auditory cortex. The dendrograms show distances of mean gene expression among transcriptomic clusters (left) and distances of mean projection pattern (right) obtained through BARseq and FISH.

(B) t-SNE plot of the gene expression of neurons color-coded by cluster identity as in (A).

(C) MetaNeighbor comparison of neuronal clusters obtained in the auditory cortex to those in the visual cortex from Tasic et al. (2018).

(D) Projections (left) and the expression of genes (right) of neurons obtained using combination of BARseq and FISH are shown on a log scale. Projection areas are the same as in Figure 4B, except that each cortical area is divided into upper (u) and lower (l) layers.

(E) Distributions of laminar positions of neurons. Individual neurons (red) are superimposed on the smoothed distribution (black).

See also Figure S7.

somatosensory cortex was also likely to have a strong projection to the visual cortex, but a weak striatal projection (Figure S5Ab). Such correlations suggest that the remarkable diversity of projections did not arise purely by chance (e.g., by a process in which each neuron selected targets at random). We therefore exploited the large sample size available from BARseq to look for statistical signatures of the organization of projections from auditory cortex.

To explore the statistical structure underlying these projection patterns, we performed hierarchical clustering on the projection patterns. For this analysis, we did not binarize the projection patterns; we considered not only whether a neuron projected to an area, but also the strength of that projection. This clustering uncovered structure well beyond the previously established top-level classes (Figures 5C and 5D; see Figures S5B–S5E and STAR Methods for details of clustering methods). Depending on the precise clustering algorithm and statistical criteria used (Figure S5F), as many as 25 subclusters were revealed. Most of these subclusters were robust: 5,682/6,391 (89%) of neurons were unambiguously assigned into one of the 25 subclusters using a probabilistic approach (Figures S5G and S5H; STAR Methods). All top-level subclusters were observed in all three brains (Figure S6A; see STAR Methods for likely misclassification of some striatal-projecting PT-I neurons as CT neurons due to choice of projection targets sampled). BARseq thus uncovered organization in projections well beyond the previously described major divisions.

As noted above, the three major cortical classes—PT-I, IT, and CT—are spatially segregated into laminae. By contrast, most subclusters defined by projection pattern, especially those within the IT class, spanned many layers (Figures 5C and 5E; $p = 0.95$ for cluster hierarchy 2–7, Kruskal-Wallis test; STAR Methods). For example, a subcluster consisting of IT neurons that project only to contralateral cortical areas, but not the ipsi- patterns was consistent across brains (Figures S6B and S6C; STAR Methods). Therefore, laminar position could not fully explain projection patterns or diversity.

BARseq Associates Projections with Transcriptionally Defined IT Subtypes

Although projection subclusters did not segregate cleanly into laminae, we hypothesized that the conjunction of laminar position and gene expression might predict projection patterns. Although gene expression patterns of single neurons have been studied in other cortical areas, including visual, somatosensory, and motor cortices (Tasic et al., 2018; Zeisel et al., 2015), gene expression patterns of auditory cortical neurons have mostly been explored in a targeted manner (Chang and Kawai, 2018; Guo et al., 2016; Hackett, 2018; Hackett et al., 2011, 2016) or in bulk (Hackett et al., 2015). We therefore performed single-cell RNA-seq for 735 neurons in the auditory cortex using 10x Genomics Chromium. As expected from work in other cortical regions, the top-level partition among neurons was between excitatory and inhibitory classes (Figure 6A). Within the excitatory neuron class, gene expression further segregated into CT and IT subtypes. As in other studies using the 10x platform (Zeisel et al., 2018), few PT-I neurons were recovered, possibly due to biases in cell survival (STAR Methods). The transcriptional taxonomy of neurons in the auditory cortex were thus consistent with other sensory cortices.

To further explore whether transcriptionally defined subtypes corresponded to subclusters defined by projection pattern, we focused on subdivisions within the IT class. IT neurons were the most diverse class based on projection mapping, with as many as 19 distinct subclusters (Figure 5C). Clustering of gene expression revealed four top-level IT subtypes, which we denote IT1, IT2, IT3, and IT4 (Figures 6A and 6B). These four subtypes corresponded well to the four major subtypes of IT neurons pre-

Cell

projection subclusters were partitioned among the 4 transcriptionally defined subtypes.

To assess the relationship between projection pattern and gene expression among IT neurons at the single neuron level, we combined BARseq with FISH against *Cdh13*, *Slc30a3*, and *Slc17a7* (Figure 6D). The combination of laminar position and expression of these markers allowed us to assign neurons to subtypes: IT1 neurons consisted of all IT neurons in layer 2/3; IT2 neurons consisted of *Cdh13*-negative IT neurons in layer 5; IT3 neurons consisted of both *Cdh13*-expressing neurons in layer 5 and *Slc30a3*-expressing neurons in L6; and IT4 neurons consist of *Slc30a3*-negative IT neurons in layer 6 cortex (Tasic et al., 2018) (Figures 6D and 6E; STAR Methods). We collected the same projection areas as in the previous BARseq experiments, but achieved higher spatial resolution of the projection targets by collecting upper and lower cortical layers separately. We obtained good sequencing and FISH in 979 projection neurons from two brains (Figures 6D and S7A), of which 735 were IT neurons defined by projection (note that this set of 735 IT neurons was independent of the 735 neurons studied using single-cell RNA-seq; the same number of neurons is coincidental). These IT neurons were assigned to one of the four genetically defined subtypes based on laminae and gene expression and to one of the projection subclusters based on projection patterns (Figure S7B). Interestingly, subtypes that were further apart based on mean gene expression were also further apart based on clustering of projections (Figure 6A), suggesting a relationship between gene expression and projection pattern within IT subtypes.

The Organization of Projections across Transcriptionally Defined IT Subtypes

Although the relationship between the IT1–IT4 subtypes defined by gene expression and the 19 subclusters defined by projection patterns was complex— projection patterns were largely mixed across the four IT subtypes—the large sample size enabled us to discern several relationships. We identified both qualitative and quantitative relationships between gene expression and projection pattern, including projection patterns specific to transcriptionally defined subtypes. These findings are detailed below.

Perhaps the most striking observation was the identification of a specific projection subcluster, which we denote ITi-Ctx, found almost exclusively in the two deep-layer subtypes IT3 and IT4, but absent in IT1 or IT2 (Figure 7A). ITi-Ctx neurons are a subset (leaves 14 and 15 in Figure 5C) of ITi neurons that project exclusively to the ipsilateral cortex, but not to the contralateral cortex or the striatum. Whereas ITi-Ctx neurons accounted for 13% (20/155) of IT3 neurons and 34% (45/123) of IT4 neurons, it accounted for only 2% (6/303) of IT1 neurons and fewer than 1% (1/145) of IT2 neurons (Figure 7A; $p < 10^{-4}$ comparing IT3 to either IT1 or IT2, and $p < 5 \times 10^{-15}$ comparing IT4 to either IT1 or IT2, all using Fisher's exact test with Bonferroni correction). These ITi-Ctx neurons would have been difficult to detect using conventional retrograde tracing because they are defined by a combinatorial projection pattern (i.e., projection to ipsilateral but not contralateral cortex).

the deep layers of the contralateral cortex (Figure 7B; $p < 5 \times 10^{-8}$ comparing IT3 to IT1 or IT2, and $p < 5 \times 10^{-7}$ comparing IT4 to IT1 or IT2, all using Fisher's exact test with Bonferroni correction). By contrast, the contralateral projections of IT1 and IT2 neurons terminate throughout all cortical layers. This result is consistent with, and expands upon, previous observations using classical anterograde tracing (Tasic et al., 2018) showing that a subset of IT3 neurons labeled by *Penk-Cre* project preferentially to deep layers.

Third, we found that if an IT3 or IT4 neuron had an ipsilateral projection, then it was unlikely to also have a contralateral projection. By contrast, neurons in IT1 and IT2 often projected both ipsi- and contralaterally (Figure 7C; $p < 5 \times 10^{-4}$ comparing IT3 to either IT1 or IT2, and $p < 5 \times 10^{-16}$ comparing IT4 to either IT1 or IT2 using Fisher's exact test with Bonferroni correction; see STAR Methods). Because the presence of a contralateral projection is the characteristic that distinguishes the ITc from the ITi projection subcluster, most (84%; 377/448) IT1 or IT2 neurons were ITc, whereas fewer than half (43%; 123/287) of IT3 or IT4 neurons were ITc (Figure 7D). These features of projections specific to IT3 and IT4 subtypes reinforce the observation that the rules governing the projections of IT3 and IT4 neurons are distinct from those governing IT1 and IT2 neurons.

Finally, we discovered finer quantitative differences across individual subtypes defined by gene expression. Although we identified two projection patterns (contralateral projections to deep layers and ITi-Ctx projections) shared by both IT3 and IT4, these two projection patterns were differentially enriched between the two subtypes. ITi-Ctx projections were both stronger (Figure S7C, $p = 0.012$ comparing IT3 and IT4 using rank-sum test) and more frequent in IT4 subtype compared to those in IT3 (Figure 7D; 13% in IT3 compared to 34% in IT4 neurons, $p = 3 \times 10^{-5}$ using Fisher's exact test). In contrast, contralateral projections were twice as frequent in IT3 neurons compared to IT4 (Figure 7D; 89/155 IT3 neurons compared to 34/132 IT4 neurons, $p = 6 \times 10^{-8}$ using Fisher's exact test). Within a subtype defined by gene expression, projections further correlate with gene expression: the probability of projecting contralaterally was higher among IT3 neurons with high *Cdh13* expression ($p = 0.001$, Fisher's exact test, Figure 7E), an observation further confirmed using CTB retrograde tracing (Figure S7D, $p < 0.0001$ using rank-sum test).

Although the analyses above support the idea that gene expression and laminar position are correlated with projection pattern, the relationship between IT subtypes defined by gene expression and subclusters defined by projection patterns was in general complex (Figure 7D). Thus, although some projection patterns were enriched in some subtypes (e.g., ITi-Ctx neurons [leaves 14 and 15] and ITc neurons of leaves 4 and 5; Figure 7D), no projection subcluster was specific to a single subtype. Indeed, in many cases, the same projection subcluster could be found in all four IT subtypes; two examples of projection subclusters found in all four transcriptomic subtypes are shown in Figure 7F. Such shared projection patterns across subtypes were common: 54% of IT neurons (400/735) were more similar in projections to a neuron of a different IT subtype than to any

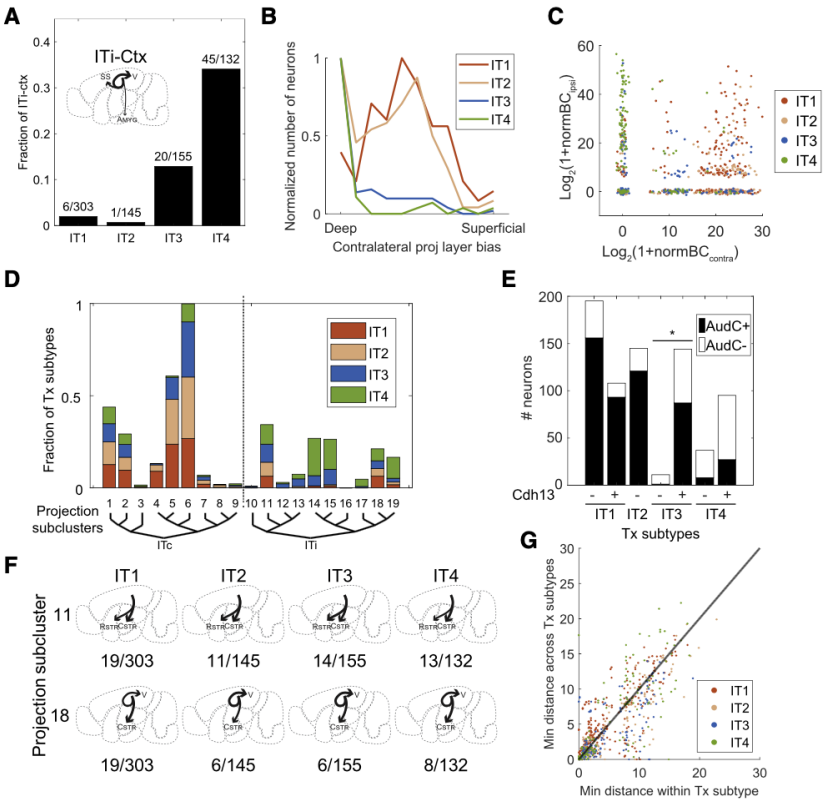


Figure 7. Projections across IT Subtypes Defined by Gene Expression

(A) The fraction of ITi-Ctx neurons in each indicated IT subtype defined by gene expression. The numbers of ITi-Ctx neurons and the total numbers of each subtype are indicated. Inset: the mean projection pattern of ITi-Ctx neurons.

(B) Histograms of the layer bias of contralateral projections of neurons in each IT subtype. The histograms are normalized so that the maximum value for a bin is 1 for each subtype.

(C) The log normalized barcode count of projections to the contralateral auditory cortex (x axis) is plotted against that of projections to the ipsilateral visual and somatosensory cortex (y axis) for each neuron.

(D) The fractions of neurons of IT subtypes defined by gene expression that belong to each indicated IT projection leaf subcluster. All bars belonging to a transcriptionally defined subtype sum to 1 across the whole plot.

(E) The number of neurons with ($AudC^+$) or without ($AudC^-$) projections to the contralateral auditory cortex in each IT subtype defined by gene expression. Neurons in each IT subtype are further divided into those expressing *Cdh13* and those that do not. *$p < 0.005$ using Fisher's exact test after Bonferroni correction.

(F) Two example projection leaf subclusters that were shared across all four IT subtypes defined by gene expression. Projection diagrams indicate example neurons. The numbers of neurons of a transcriptionally defined subtype and of those belonging to a projection subcluster are indicated below.

(G) The minimum distance in projections from a neuron to any neuron within the same subtype defined by gene expression (x axis) or in a different subtype (y axis). See also Figure S7.

projection patterns across the three higher-level classes of neurons (IT/PT-l/CT). Thus, whereas at the top levels of the hierarchies, the classical partitioning into IT/PT-l/CT neurons captures the correlational structure of multiple cellular properties, including gene expression, laminar position, and projection pattern, our results suggest that further subclustering based on one experimental property (such as gene expression) leads to categories that do not map neatly onto clustering based on another property (such as projections).

DISCUSSION

Here, we have described BARseq, a barcoding-based neuroanatomical method that can relate high-throughput projection mapping with other neuronal properties at cellular resolution through *in situ* sequencing. As a proof of principle, we applied BARseq to the projections in the auditory cortex. BARseq of 3,579 neurons recapitulated the organization of cortical projections into the classic IT/PT-l/CT subtypes. Within these classic subtypes, BARseq also revealed the impressive diversity of projection patterns: we observed 264 distinct projection patterns, falling into as many as 25 distinct clusters. We then combined BARseq with single cell analysis of gene expression of an addi-

BARseq thus revealed rich and complex relationships between gene expression and projection patterns that would have been difficult to uncover using conventional transcriptomic and/or anatomical techniques.

Multiplexed Projection Mapping Using Cellular Barcodes

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

BARseq fundamentally differs from conventional optical approaches in that projections are determined by matching barcodes, not tracing. Barcode matching, unlike optical tracing, does not accumulate error over distance. 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 exponentially with the length of the axonal projection and the number of neurites multiplexed. For example, even with a low error rate of

Cell

sequences to reconstruct projections, errors do not accumulate for long axons: A distant subcortical projection is just as reliably matched to its source as a projection to a nearby cortical area. Furthermore, the high diversity of barcodes (tens of millions in our experiments) can uniquely label tens of thousands of neurons in a single experiment. Therefore BARseq can determine projection patterns for orders of magnitude higher densities 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; Lin et al., 2018), without the need of specialized high-speed microscope that is often required for advanced anatomical techniques.

In BARseq, the spatial resolution at which projections are resolved is determined by the size of the cubelets dissected. In this study, we chose to dissect brains relatively coarsely—only sufficient to distinguish among brain areas and between superficial and deep cortical layers—but this resolution was sufficient to resolve the organization of projections across neuronal subtypes. For questions requiring higher spatial resolution, BARseq can be further improved with laser capture microdissection (Huang et al., 2018) or direct *in situ* sequencing of projection barcodes, thus potentially resolving axonal projections and dendritic morphology at subcellular resolution. Such an approach would yield an "infinite color Brainbow," allowing reconstruction of densely labeled neurons.

The Organization of Projections across Neuronal Subtypes

The partitioning of neurons into types and subtypes is useful to the extent that these classes can infer multiple neuronal properties. The utility of these classes and subclasses arises from the fact that these properties co-vary. For example, knowing whether a cortical neuron is excitatory or inhibitory, or the excitatory subtype (IT/PT-l/CT) to which it belongs (Harris and Shepherd, 2015), allows us to infer a great deal about the neuron's pattern of gene expression, morphology, connectivity, and projection pattern. The partitioning of cortical neurons into excitatory versus inhibitory, and of excitatory neurons into IT/PT-l/CT subtypes, is well supported by many lines of data.

Within these major subtypes of excitatory neuron, the predictive value of further subdivisions has been less clear. It has been reported that some PT-l subtypes defined by gene expression correspond to specific projection patterns (Economo et al., 2018), and some CT subtypes show subtype-dependent biases in their projections (Chevée et al., 2018). However, a systematic approach to determining the co-variance of two or more properties would require high-throughput measurements of multiple properties simultaneously, a feat that would be challenging using previous approaches. Indeed, it has only recently been possible to obtain such data on even a single property such as gene expression or projection pattern (Kebschull et al., 2016a).

The combination of BARseq with *in situ* transcriptional mapping has the potential to systematically determine the co-variation of multiple properties. As a proof-of-principle, we combined BARseq and FISH to explore the projections of transcriptionally defined IT subtypes. Using this approach, we identified transcriptionally defined IT subtypes (IT3 and IT4). We also identified several other ways in which patterns of gene expression covaried with projection patterns (Figure 7). However, in marked contrast to the one-to-one relationship between top-level subtypes (IT/CT/PT-l) defined by gene expression and projection pattern, we did not observe a simple mapping of these properties within IT neurons.

The mixed projection patterns observed across IT subtypes defined by gene expression are unlikely to have resulted from errors in classifying neurons to subtypes, because we were able to identify both specific and promiscuous projection patterns in the same IT subtypes. This question can be further tested by correlating projections with more marker genes to better define IT subtypes by gene expression. To achieve this, BARseq can potentially be combined with highly multiplexed FISH (Chen et al., 2015; Codeluppi et al., 2018; Eng et al., 2019; Shah et al., 2016) or *in situ* sequencing methods (Ke et al., 2013; Lee et al., 2014; Wang et al., 2018) to correlate the expression of dozens to hundreds of genes to projections. Such an approach could further dissect the relationship between projections and gene expression beyond IT subtypes and has the potential to generate a comprehensive understanding of long-range projections and gene expression at cellular resolution.

Rosetta Brain

Neuronal types differ by a combination of multiple properties (Cadwell et al., 2016; Economo et al., 2018; Paul et al., 2017), including anatomical characteristics such as morphology and connectivity (Economo et al., 2016; Gerfen et al., 2018), molecular characteristics such as gene (Luo et al., 2017; Tasic et al., 2016, 2018; Zeisel et al., 2015) or protein expression, and functional characteristics such as behaviorally evoked activity (Cadwell et al., 2016). Defining the biological functions of neuronal classes thus requires a multi-faceted approach that combines measurements of several neuronal properties together at cellular resolution (Huang and Paul, 2019; Zeng and Sanes, 2017). The ability to relate diverse characteristics of many single neurons simultaneously, in a co-registered fashion, within single brains, is an important challenge in neuroscience (Marblestone et al., 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; Raj et al., 2018; Sorensen et al., 2015).

Co-registering multiple cellular properties at single-cell resolution is crucial for understanding biological systems with heterogeneous cell types, such as the nervous system. By sequencing barcodes *in situ*, BARseq offers a solution to incorporate high-throughput measurements based on cellular barcoding, such as lineage tracing (Raj et al., 2018), high-throughput screening (Feldman et al., 2018), and projection mapping (Kebschull et al., 2016a) into such integrated approaches. In this study, we focused on combining projections with adult gene expression and the laminar location of neuronal somata. Adapting BARseq for other barcoding techniques and adapting additional *in situ* methods, such as *in vivo* two-photon imaging, to combine

of structure and function in the nervous system (Marblestone et al., 2014).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
- METHOD DETAILS
  - Viruses, constructs, and oligos
  - Animals and tissue processing
  - BaristaSeq
  - Imaging
  - Base-calling
  - Comparison to other *in situ* sequencing methods
  - BARseq
  - Comparison of BARseq to retrograde tracer
  - BARseq in Cre-labeled animals
  - Validation using retrograde tracing
  - Validation of combination of BARseq and FISH
  - Estimation of laminar boundaries using FISH
  - Single-cell dissociation for RNaseq
  - Single-cell RNaseq
  - Projections of IT subtypes defined by genes
  - Considerations for Sindbis toxicity in BARseq
  - Validation of Cdh13 correlation to projections
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - BARseq data processing
  - Comparison of bulk projection patterns
  - Population analysis of projections
  - Hierarchical clustering
  - Laminar distribution of neurons
  - Clustering of single-cell RNaseq
  - Gene expression in barcoded neurons
  - Data processing for BARseq/FISH validation
  - Projections of IT subtypes defined by genes
  - Statistical analyses
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2019.09.023.

ACKNOWLEDGMENTS

The authors would like to acknowledge Kay Tye, Gordon Shephard, and Jessica Tollkuhn for useful discussions, and Barry Burbach, Nour El-Amine, Stephen Hearn, Jon Preall, and Yan Li for technical support. This work was supported by the following funding sources: NIH (5RO1NS073129 and 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 which are funded, in part, by the NIH (Cancer Center support grant 5P30CA045508).

AUTHOR CONTRIBUTIONS

X.C. and A.M.Z. conceived the study. X.C., H.Z., and Y.-C.S. optimized and performed BARseq. X.C. and Y.-C.S. performed BARseq in combination with FISH and in *Cre*-labeled mice. J.M.K. and H.Z. performed MAPseq. X.C. performed single-cell RNA-seq. K.M. and Z.J.H. generated *Cre*-labeled mice. X.C. and A.M.Z. analyzed the BARseq and MAPseq data. X.C., S.F., and J.G. analyzed the single-cell RNA-seq data. X.C. and A.M.Z. wrote the paper.

DECLARATION OF INTERESTS

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

Received: September 4, 2018
Revised: July 30, 2019
Accepted: September 20, 2019
Published: October 17, 2019

REFERENCES

Asokan, M.M., Williamson, R.S., Hancock, K.E., and Polley, D.B. (2018). Sensory overamplification in layer 5 auditory corticofugal projection neurons following cochlear nerve synaptic damage. Nat. Commun. *9*, 2468.

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. J. Stat. Mech. *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.

Chang, M., and Kawai, H.D. (2018). A characterization of laminar architecture in mouse primary auditory cortex. Brain Struct. Funct. *223*, 4187–4209.

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.

Chevée, M., Robertson, J.J., Cannon, G.H., Brown, S.P., and Goff, L.A. (2018). Variation in Activity State, Axonal Projection, and Position Define the Transcriptional Identity of Individual Neocortical Projection Neurons. Cell Rep. *22*, 441–455.

Codeluppi, S., Borm, L.E., Zeisel, A., La Manno, G., van Lunteren, J.A., Svensson, C.I., and Linnarsson, S. (2018). Spatial organization of the somatosensory cortex revealed by osmFISH. Nat. Methods *15*, 932–935.

Crow, M., Paul, A., Ballouz, S., Huang, Z.J., and Gillis, J. (2018). Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. Nat. Commun. *9*, 884.

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.

Economo, M.N., Viswanathan, S., Tasic, B., Bas, E., Winnubst, J., Menon, V., Graybuck, L.T., Nguyen, T.N., Smith, K.A., Yao, Z., et al. (2018). Distinct de-

Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014). Advanced methods of microscope control using μManager software. J. Biol. Methods *1*, e10.

Eng, C.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.C., and Cai, L. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. Nature *568*, 235–239.

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.

Feldman, D., Singh, A., Schmid-Burgk, J.L., Mezger, A., Garrity, A.J., Carlson, R.J., Zhang, F., and Blainey, P.C. (2018). Pooled optical screens in human cells. bioRxiv. https://doi.org/10.1101/383943.

Gerfen, C.R., Economo, M.N., and Chandrashekar, J. (2018). Long distance projections of cortical pyramidal neurons. J. Neurosci. Res. *96*, 1467–1475.

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.

Greig, L.C., 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.

Guo, Y., Zhang, P., Sheng, Q., Zhao, S., and Hackett, T.A. (2016). lncRNA expression in the auditory forebrain during postnatal development. Gene *593*, 201–216.

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.

Hackett, T.A. (2018). Adenosine A_1 Receptor mRNA Expression by Neurons and Glia in the Auditory Forebrain. Anat. Rec. (Hoboken) *301*, 1882–1905.

Hackett, T.A., Takahata, T., and Balaram, P. (2011). VGLUT1 and VGLUT2 mRNA expression in the primate auditory pathway. Hear. Res. *274*, 129–141.

Hackett, T.A., Guo, Y., Clause, A., Hackett, N.J., Garbett, K., Zhang, P., Polley, D.B., and Mirnics, K. (2015). Transcriptional maturation of the mouse auditory forebrain. BMC Genomics *16*, 606.

Hackett, T.A., Clause, A.R., Takahata, T., Hackett, N.J., and Polley, D.B. (2016). Differential maturation of vesicular glutamate and GABA transporter expression in the mouse auditory forebrain during the first weeks of hearing. Brain Struct. Funct. *221*, 2619–2673.

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.

Huang, Z.J., and Paul, A. (2019). The diversity of GABAergic neurons and neural communication elements. Nat. Rev. Neurosci. *20*, 563–572.

Huang, L., Kebschull, J.M., Furth, D., Musall, S., Kaufman, M.T., Churchland, A.K., and Zador, A.M. (2018). High-throughput mapping of mesoscale connectomes in individual mice. bioRxiv. https://doi.org/10.1101/422477.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, 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. https://doi.org/10.1101/378760.

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.

Lin, H.M., Kuang, J.X., Sun, P., Li, N., Lv, X., and Zhang, Y.H. (2018). Reconstruction of Intratelencephalic Neurons in the Mouse Secondary Motor Cortex Reveals the Diverse Projection Patterns of Single Neurons. Front. Neuroanat. *12*, 86.

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.

Luo, C., Keown, C.L., Kurihara, L., Zhou, J., He, Y., Li, J., Castanon, R., Lucero, J., Nery, J.R., Sandoval, J.P., et al. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science *357*, 600–604.

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., and Church, G.M. (2014). Rosetta Brains: A Strategy for Molecularly-Annotated Connectomics. arXiv, arXiv:1404.5103.

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. USA *108*, 16807–16812.

Ng, A.Y., Jordan, M.I., and Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In Advances in Neural Information Processing Systems: Proceedings of the First 12 Conferences (The MIT Press), pp. 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.

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.

Raj, B., Wagner, D.E., McKenna, A., Pandey, S., Klein, A.M., Shendure, J., Gagnon, J.A., and Schier, A.F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat. Biotechnol. *36*, 442–450.

Rock, C., Zurita, H., Wilson, C., and Apicella, A.J. (2016). An inhibitory corticostriatal pathway. eLife *5*, e15890.

Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science *363*, 1463–1467.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

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

Ståhl, P.L., Salmén, 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.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.

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., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. Nature *563*, 72–78.

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 *361*, eaat5691.

Williamson, R.S., and Polley, D.B. (2019). Parallel pathways for sound processing and functional connectivity among layer 5 and 6 auditory corticofugal neurons. eLife *8*, e42974.

Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, 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.

Zeisel, A., Hochgerner, H., Lonnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Haring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse Nervous System. Cell *174*, 999–1014.

Zeng, H., and Sanes, J.R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. Nat. Rev. Neurosci. *18*, 530–546.

Zheng, T., Yang, Z., Li, A., Lv, X., Zhou, Z., Wang, X., Qi, X., Li, S., Luo, Q., Gong, H., and Zeng, S. (2013). Visualization of brain circuits using two-photon fluorescence micro-optical sectioning tomography. Opt. Express *21*, 9839–9850.

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

Zurita, H., Feyen, P.L.C., and Apicella, A.J. (2018). Layer 5 Callosal Parvalbumin-Expressing Neurons: A Distinct Functional Group of GABAergic Neurons. Front. Cell. Neurosci. *12*, 53.

Cell

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
JK100L2 Sindbis virus	This paper; Kebschull et al., 2016a	N/A
Chemicals, Peptides, and Recombinant Proteins		
Pierce MMTS	Thermo Fisher Scientific	Cat#23011
RevertAid H-minus reverse transcriptase	Thermo Fisher Scientific	Cat#EP0452
Ampligase	Lucigen	Cat#A0110K
BS(PEG)9	Thermo Fisher Scientific	Cat#21582
Phusion High-Fidelity DNA polymerase	Thermo Fisher Scientific	Cat#F530L
RiboLock RNase inhibitor	Thermo Fisher Scientific	Cat#EO0384
Phi29 DNA polymerase	Thermo Fisher Scientific	Cat#EP0094
RNase H	QIAGEN	Cat#9220L
Affymetrix/USB Pepsin from Porcine stomach	Fisher Scientific	Cat#AAJ20015A1
Cholera Toxin Subunit B, Alexa 488 conjugates	Thermo Fisher Scientific	Cat#C22841
Cholera Toxin Subunit B, Alexa 647 conjugates	Thermo Fisher Scientific	Cat#C34778
Red Retrobeads	LumaFluor	Cat#R170
Critical Commercial Assays		
Illumina HiSeq SBS kit v4 50 cycles	Illumina	Cat#FC-401-4002
RNAscope fluorescent multiplex reagent kit	ACDBio	Cat#320850
ViewRNA ISH Cell Assay Kit	Thermo Fisher Scientific	Cat#QVC0001
Deposited Data		
MAPseq, BARseq, and single-cell RNaseq high-throughput sequencing data (non-*in situ*)	This paper, SRA	SRA: PRJNA557267 and SRA: PRJNA448728
Bulk GFP tracing in the auditory cortex	Allen Mouse Connectivity Atlas; Oh et al., 2014	Allen Mouse Connectivity Atlas 116903230, 100149109, 120491896, 112881858, and 146858006
In situ sequencing data for BARseq and other raw and processed data	This paper, Mendeley data	doi: 10.17632/mk82s9x82t.1, 10.17632/g7kdxznt6w.1, 10.17632/86wf7xfz5x.1, 10.17632/2w649fccnt.1
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J males, 8-10 wks	Jackson Laboratory	JAX: 000664; RRID: IMSR_JAX:000664
Mouse: *Fezf2*-2A-*CreER*::Rosa-CAG-loxp-STOP-loxp-tdTomato (Ai14)	Laboratory of Z. Josh Huang	N/A
Oligonucleotides		
Primer for reverse transcription: XC1215, /5AmMC12/T+AC+AG+CT+AG+CG+GT+GGTCGTACCTCACGACG	Chen et al., 2018	N/A
Padlock probe/sequencing primer: XC1164, /5phos/gtactgcggccgctacctaaTCCTCTATGATTACTGACTGCGTCTATTTAGTGGAGCCATTGCTATCTTCTTacacgacgctcttccgatct	Chen et al., 2018	N/A
Rolony visualization probe: XC92, /5TYE563/CCTCTATGATTACTGACTGCGTCTA	Chen et al., 2018	N/A
Recombinant DNA		
JK100L2	Kebschull et al., 2016a	Addgene plasmid #79785

(*Continued on next page*)

Cell

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
ImageJ (Fiji)	Schindelin et al., 2012	RRID: SCR_002285; http://fiji.sc
MATLAB	Mathworks	RRID: SCR_001622; https://www.mathworks.com/products/matlab.html
Seurat v2, v3	Stuart et al., 2019	RRID: SCR_016341; https://satijalab.org/seurat/get_started.html
Louvain community detection	Blondel et al., 2008	https://perso.uclouvain.be/vincent.blondel/research/Community_BGLL_Matlab.zip
Normalized spectral clustering	Ng et al., 2002	https://www.mathworks.com/matlabcentral/fileexchange/34412-fast-and-efficient-spectral-clustering
MetaNeighbor	Crow et al., 2018	RRID: SCR_016727; https://github.com/maggiecrow/MetaNeighbor
Rstudio	Rstudio team	RRID: SCR_000432; https://rstudio.com/
Cell Ranger	10x Genomics	RRID: SCR_017344; https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger
Iterative Closest Point	Besl and McKay, 1992	https://www.mathworks.com/matlabcentral/fileexchange/27804-iterative-closest-point
ECC image alignment	Evangelidis and Psarakis, 2008	https://sites.google.com/site/imagealignment/home
Data processing and analysis scripts	This paper, Mendeley data	doi: 10.17632/mk82s9x82t.1, 10.17632/g7kdxznt6w.1, 10.17632/86wf7xfz5x.1, 10.17632/2w649fccnt.1

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, A.M.Z. (zador@cshl.edu). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

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. For BARseq in Cre-labeled neurons, a *Fezf2*-2A-*CreER*::Rosa-CAG-loxP-STOP-loxP-tdTomato (Ai14) female animal was induced at 6 and 7 postnatal weeks and barcoded at three postnatal months. This animal was generated in the Huang lab. For all other BARseq and MAPseq experiments, 8 to 10 week-old male C57BL/6J animals acquired from Jackson Laboratory were used.

METHOD DETAILS

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).

For validation of barcode sequencing 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-

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 tissue for 24 hr. We then cryo-protected the brain in PBS with 10% sucrose for 12 hr, 20% sucrose PBS for 12 hr, and 30% sucrose PBS for 12 hr. 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 trans-cardially perfused the animal with PBS 43-45 hr post-injection. We cut out the left auditory cortex from the brain and post-fixed it in 10% formalin at 4°C for 8 hr, 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 in Figures 3D–3F, we injected 8-week old C57BL/6 male mouse at −4 mm ML, −2.6 mm AP from bregma, with 140 nL 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. For the experiment in Figures 6D, 6E, and 7, 8-week old C57BL/6 male mice were injected at −4.3 mm ML, −2.6 mm AP from bregma with 140nL 1:3 diluted Sindbis virus at depths 300 μm, 600 μm, and 900 μm at 30° angle. After 24 hours, the animal was processed as described above for BARseq experiments, except that the injection site was flash frozen in OCT using isopentane and liquid nitrogen immediately after dissection from the brain. These brains were cryo-sectioned to 14 μm sections using a home-made tape-transfer system (Pinskiy et al., 2015) and glued to slides using NOA 81 (Norland) to reduce tissue distortion.

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 hr, 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 hr, the animals were then processed as for conventional BARseq samples.

For single-cell RNaseq comparing barcoded cells to non-barcoded cells, 8-week old C57BL/6 male mice were injected at −4.3 mm ML, −2.6 mm AP from bregma with 140 nL 1:3 diluted Sindbis virus at depths 300 μm, 600 μm, and 900 μm at 30° angle. After 24 hours, the animal was processed for single-cell dissociation as described below.

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 0.1 M 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 (Epicenter), 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 μM aadUTP, 0.2 μg/μl BSA, 250 μM dNTPs, and 1 U/μl ϕ29 DNA polymerase in 1 × ϕ29 DNA polymerase buffer supplemented with 5% additional glycerol] overnight at room temperature. After crosslinking the rolonies using $BS(PEG)_9$ and neutralizing 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 (Figures S1Aa–S1Ac), and found that the HybriWell-FL sealing system was the only system that did not inhibit rolony formation (Figure S1Ab). The ImmEdge hydrophobic barrier pen also produced good amplification (Figure 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 × 22 mm × 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 (Figure S1B) and to reduce the GFP signal from the cells (cyan in Figure 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 (Figure S1Bb) compared to no pepsin treatment (Figure S1Ba), whereas 5 mins of pepsin digestion caused excessive tissue loss (Figure 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. For slices sectioned using the tape-transfer system (including XC54, XC75, XC91, and XC92), we further incubated the

resulted in an average signal-to-noise ratio 39 ± 4 in the first six sequencing cycles (Figures S1D and S1E), ~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 Figures 2C–2F and 3A–3C 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 were: Channel G, 514 nm laser excitation, 405/440/514/640 quad dichroic, 550/49 emission filter, exposure time 500 ms; Channel Y, 561 nm laser excitation, 405/488/561/640 dichroic, 445/60 and 615/70 dual band emission filter, exposure time 120 ms; Channel A, 640 nm laser excitation, 405/488/561/640 dichroic, 676/29 emission filter, exposure time 250 ms; Channel C, 640 nm laser excitation, 405/488/561/640 dichroic, 775/140 emission filter, exposure time 500 ms. The experiments in Figures 2C–2F were produced on a Zeiss LSM 710 laser scanning confocal microscope with Zeiss Zen software as previously described (Chen et al., 2018). The experiment in Figures 3A–3C was imaged on a Nikon TE2000 with Crest X-light spinning disk, Photometrics Prime 95B camera, and an 89North 7-line LDI laser. The scope was controlled by micro-manager (Edelstein et al., 2014). The four sequencing channels were imaged using the following settings: Channel G, excitation 520 nm laser, ZT443-518rpc-UF2 dual-band dichroic (Chroma), FF01-565/24-25 emission filter (Semrock), exposure time 600 ms; Channel T, excitation 555 nm laser, zt402/468/555/640rpc-uf2 quad-band dichroic (Chroma), FF01-585/11-25 emission filter (Semrock), exposure time 200 ms; Channel A, excitation 640 nm laser, zt402/468/555/640rpc-uf2 quad-band dichroic (Chroma), FF01-676/29-25 emission filter (Semrock), exposure time 300 ms; Channel C, excitation 640 nm laser, zt402/468/555/640rpc-uf2 quad-band dichroic (Chroma), FF01-725/40-25 emission filter (Semrock), exposure time 500 ms. The channels were calibrated as described previously (Chen et al., 2018).

For XC9 and XC28, a 3 × 1 stitched image at 10 × was acquired for each slice, with each tile position including 11 z-positions with 10 μm step size. For XC75 and XC91, a 3 × 3 stitched image at 10 × was acquired for each slice. Imaging time was about one minute per slice for XC9 and XC28, and three minutes per slice for XC75 and XC91. During each sequencing run, 24-36 slices on three slides were imaged together, resulting in a per-cycle imaging time of 0.5-1 hour. For XC75 and XC91, an additional 5 × 5 tiles of 20 × image was acquired for the first sequencing cycle, with each tile position including 17 z-positions with 3 μm step size. This high-resolution imaging typically took 15 minutes per slice.

All images in figures were maximum projections of z stacks shown after rolling ball background subtraction except for images in Figure S1D, which were shown with only the camera blackpoint subtracted from the images.

Base-calling

Max projection images were first processed through a median filter and a rolling ball background subtraction. The processed images were then corrected for channel bleed through. The barcodes were then base-called by picking the channel with the strongest signal. The sequencing 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.

Comparison to other *in situ* sequencing methods

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 (Epicenter) 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, 1 M 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 using ECC image alignment (Evangelidis and Psarakis, 2008). The sequencing images were

Cell

We dissected 12 projection sites from the frozen brains for sequencing. These 12 sites included the olfactory bulb (OB), orbitofrontal cortex (OFC), motor cortex (Motor), rostral striatum (Rstr), somatosensory cortex (SSctx), caudal striatum (Cstr), amygdala (Amyg), ipsilateral visual cortex (VisIp), contralateral visual cortex (VisC), contralateral auditory cortex (AudC), thalamus (Thal), and tectum (Tect). 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 Mendeley (see Data and Code Availability). The projection sites were sequenced as described for MAPseq (Kebschull et al., 2016a).

For experiments using combination of BARseq and FISH (XC75 and XC91), each of the cortical projection sites was separated into upper and lower layers and collected separately, resulting in 18 projection sites.

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.

BARseq in Cre-labeled animals

Fezf2-2A-*CreER*::Rosa-CAG-loxP-STOP-loxP-tdTomato (Ai14) animals were generated and *Cre* recombination was induced at 6 and 7 postnatal weeks by intraperitoneal injection of Tamoxifen (Sigma-Aldrich T-5648) (100 mg/kg for each dose). A *Fezf2*-2A-*CreER*::Ai14 animal was injected at three postnatal months with barcodes in the auditory cortex as described above for BARseq. After 24 hours, the animal was anesthetized with isoflurane and decapitated. The injection site was punched out using a 2 mm diameter biopsy punch, and the rest of the brain was flash frozen for dissection of projection areas. The injection site punch was then post-fixed in 4% PFA for 24 hours, cryo-protected in 10%, 20%, and 30% sucrose in PBS for 12 hours each, mounted in OCT, and frozen. The punch was then cryo-sectioned to 14 um slices using the tape-transfer system and imaged for GFP, RFP, and DIC channels on a Nikon TE2000 microscope with a Crest X-light spinning disk confocal. The slices were then processed for BARseq as described above. Sequencing was done as described above for 14 cycles, except that the imaging was done on a Nikon TE2000 microscope with a Crest X-light spinning disk confocal. In addition, a set of sequencing images was taken after each cleavage step. This background image records the residual tdTomato signals in the sequencing images. After registering and subtracting the background image from sequencing images, we then median filtered the images, performed background subtraction, corrected for channel bleed-through, and registered sequencing images as described above. We also registered the "T" channel of the first sequencing cycle to the tdTomato images taken before library preparation. We then picked cell bodies from the sequencing images using the "Find Maxima" function in ImageJ (Schindelin et al., 2012), base-called for barcode sequences, and read out the intensity from the tdTomato images. We filtered out cells with sequencing quality lower than 0.7. We considered cells with tdTomato signal over 20000 as *Fezf2*+ neurons. These cells were visually inspected to make sure that they had good morphology and were labeled by tdTomato. These cells were then used for analyses.

Validation using retrograde tracing

To validate the striatal projections of PT-l 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 hr, 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).

Validation of combination of BARseq and FISH

To correlate projections to the expression of *Slc17a7* and *Gad2*, 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 using the tape transfer system. The brain slices were then processed for

Cell

manufacturer's protocol. We then imaged each area around the injection site (identified by the presence of GFP) using a 20 × 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 2 × 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 20 × 0.75NA objective without binning for all four sequencing channels and the DIC channel. This was followed by imaging using a 10 × 0.45NA objective with 2 × binning. All subsequent sequencing cycles were imaged at the lower resolution to increase throughput. See QUANTIFICATION AND STATISTICAL ANALYSIS for data processing.

Estimation of laminar boundaries using FISH

To estimate the boundaries of cortical layers, we performed FISH against two known layer-specific marker genes, *Cux2* (Greig et al., 2013) (Figure S3B), and *Fezf2* (Figure S3C). *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 (Figure S3D), 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 (Figure S3E). Neurons in superficial L2/3 of the auditory cortex were 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.

Single-cell dissociation for RNaseq

To dissociate neurons for single-cell RNaseq, the animal was anesthetized with isofluorane and decapitated. We then dissected out the brain and used a 2 mm biopsy punch to remove the auditory cortex. The auditory cortex was immediately dissected in ice cold HABG medium [40 mL Hibernate A (Brainbits), 0.8 mL B27 (Thermo Fisher Scientific), and 0.1 mL Glutamax (Thermo Fisher Scientific)] into small pieces and placed in 3 mL papain solution [3mL Hibernate A-Ca (Brainbits), 6 mg papain (Brainbits), and 7.5 μL Glutamax] pre-warmed to 30°C in a 15 mL tube. The tube was closed and gently rocked at 30°C for 40 mins. The digested tissues were then transferred to a tube containing 2 mL HABG pre-warmed to 30°C and triturated 10 times using a salinized pipette with 500 μm opening. The undissociated tissues were then transferred to a new tube with 2 mL HABG and triturated 10 times using a salinized pipette with 500 μm opening. The remaining undissociated tissues were again transferred to a new tube with 2mL HABG and triturated 5 times. The three tubes of 2 mL HABG each were then combined, and carefully laid on top of a density gradient of 17.3%, 12.4%, 9.9%, and 7.4% (v/v) Optiprep in HABG. This tube was centrifuged at 750 g for 15 mins. The top two fractions were then removed. The next two fractions and half of the bottom fraction, which contained neurons, were collected. The remaining 0.5 mL of the last fraction and the pellet were discarded. The collected neuronal fraction (about 2.5 mL), was diluted in 5 mL HABG and centrifuged at 300 g for 5 mins. The pellet was washed in 5 mL HABG, pelleted again, and resuspended in 100 μL HABG. The cell suspension was then kept on ice and processed for library preparation using 10× Genomics Chromium according to the manufacturer's protocol. The prepared libraries were sequenced using Illumina NextSeq.

Single-cell RNaseq

For non-barcoded cells in the auditory cortex, the library was prepared using Chromium single cell 3′ reagent kit v3 (10× Genomics) and sequenced using one lane of Illumina NextSeq. The raw data was processed using Cell Ranger v3 (10× Genomics) followed by Seurat v3 (Stuart et al., 2019). See QUANTIFICATION AND STATISTICAL ANALYSIS for data processing and analysis.

To compare barcoded and non-barcoded cells in the auditory cortex, we injected the brain in the auditory cortex as described above. After 24 hours, we dissociated neurons from the auditory cortex as described above, and prepared library using Chromium single cell 3′ reagent kit v2 (10× Genomics). The sequencing library was then sequenced using one lane of Illumina NextSeq. See QUANTIFICATION AND STATISTICAL ANALYSIS for data processing and analysis.

Projections of IT subtypes defined by genes

The four IT subtypes defined by gene expression are largely segregated by laminae (Tasic et al., 2018): IT1 neurons are in layer 2/3; IT 2 neurons are in layer 5; IT3 neurons are in layer 5 and 6; and IT4 neurons are in layer 6. However, because layer 5 contains both IT2 and IT3 neurons, and layer 6 contains both IT3 and IT4 neurons, laminar position alone was not sufficient to distinguish between them.

Based on the gene expression (Figure 6A) and the laminar distribution of the four IT subtypes, we could distinguish the four IT subtypes using the laminar position of neurons and the expression of *Cdh13* and *Slc30a3*: IT1 neurons are IT neurons in layer 2/3; IT2 neurons are layer 5 IT neurons without *Cdh13*; IT3 neurons are layer 5 IT neurons with *Cdh13*, and layer 6 IT neurons with *Slc30a3*; IT4 neurons are layer 6 IT neurons without *Slc30a3*. We further included *Slc17a7*, a highly expressed excitatory neuron marker in the cortex, as a control for both cell morphology and FISH quality.

We collected the same 12 projection targets as in the BARseq-only experiments, but each cortical site was further dissected into superficial and deep layers. This resulted in a total of 18 projection sites. The projection sites were processed as in the BARseq-only experiments and sequenced on a Illumina MiSeq.

The injection site (auditory cortex) was cryo-sectioned to 14 μm slices using the home-made tape transfer system. The slides were then fixed and permeated according to RNAscope (ACD Bio) pretreatment procedures for fresh-frozen tissues (cold 4% PFA fixation for 15 mins, followed by dehydration in ethanol and 30 mins of Protease IV treatment at room temperature). We then washed the slide three times in PBST, deactivated residual protease by incubating twice in 40% formamide, 2 × SSC, 0.01% Triton X-100 for 10 mins at 60°C, then washed an additional three times in PBST. We then performed reverse transcription, crosslinking, and neutralization of crosslinkers as in BARseq, except that the reverse transcription was performed for only two hours at 37°C. We then followed procedures for RNAscope according to the manufacturer's protocol.

The FISH signals produced by RNAscope were then imaged on a spinning disk confocal using a 20 × 0.75 NA Plan-Apo objective with 2 × binning. After imaging, the FISH probes were stripped twice in 40% formamide, 2 × SSC, 0.01% Triton X-100 for 10 mins at 60°C. We then continued with BARseq from padlock ligation as described above. Sequencing of barcodes was performed as in BARseq-only experiments, except that the first sequencing cycle was imaged with a 10 × 0.45 NA Plan-Apo objective and a 20 × 0.75NA Plan-Apo objective, both with 2 × binning.

The FISH images were then registered to the first cycle high-resolution sequencing images using the DIC channel using intensity-based registration in MATLAB. The high-resolution first-cycle sequencing images were registered to the low-resolution first-cycle sequencing images using intensity-based registration. We then extracted cell-body locations using the "Find Maxima" function in ImageJ, and base-called the barcoded somata from the low-res images. We then filtered out cells with low sequencing quality (below 0.8). We then manually counted FISH reads for excitatory neurons with high FISH quality. Excitatory neurons with high FISH quality is defined by: (1) The neuronal somata were clearly delineated by barcodes; (2) the *Slc17a7* reads clearly defined the same soma shape; (3) the neuron did not overlap significantly with neighboring neurons. Cells with fewer than 10 copies of *Slc17a7* per cell were further excluded (these might have been cells with low FISH quality, partial cells that were cut during cryo-sectioning, and non-excitatory or non-neuronal cells). The remaining cells, including 781 excitatory neurons from XC75 and 737 excitatory neurons from XC91, were mapped to barcodes at projection sites for further analyses.

Considerations for Sindbis toxicity in BARseq

Long-term Sindbis infection disrupts gene expression, and eventually kills cells. As a result, the timing of infection is critical when using BARseq to correlate gene expression to projections, and additional quality controls are necessary to validate gene expression. In the adult auditory cortex, we have found that both gene expression and projections can be read out at 24 hours. Furthermore, we used the expression of *Slc17a7*, a gene ubiquitously expressed in excitatory neurons, as a quality control to further filter out cells whose gene expression was too low. To use BARseq in other brain areas and at different developmental stages, the best infection time may require additional optimization and a suitable quality-control marker ubiquitously expressed in the cell types of interest is needed.

Validation of Cdh13 correlation to projections

12-week old C57BL/6 animals were injected in the somatosensory cortex with 150 nL CTB-Alexa 488 at 0.6 mm AP, −2.2 mm ML from begma, at 300 and 600 um depths, and in the contralateral auditory cortex with 150 nL CTB-Alexa 647 at 2.3 mm AP, 4.2 mm ML, at 500 and 700 um depths. After 96 hours, the animal was perfused with fresh 4% PFA, post-fixed for 24 hours, and cut to 70 μm coronal sections using a vibratome. Four slices containing the ipsilateral auditory cortex were mounted and imaged for both the green and far red channel on an Ultraview spinning-disk confocal (Perkin Elmer). We then probed for *Cdh13* in these slices using RNAscope Fluorescent Multiplex Assay following the manufacturer's instructions. The *Cdh13* probe was visualized with the red Amp4 in the RNAscope kit. We then imaged the green channel, far red channel, and the *Cdh13* channel (red), and registered the images to the green/far red channel images taken before RNAscope using those two channels. We then manually counted the *Cdh13* expression in ipsilateral projecting or contralateral projecting neurons in layer 6 (defined by a normalized cortical depth of over 830 μm). Rank sum test was used to calculate statistical significance between the distribution of *Cdh13* in ipsilaterally projecting neurons and contralaterally projecting neurons.

QUANTIFICATION AND STATISTICAL ANALYSIS

BARseq data processing

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 sequenced *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; Figure 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 Figure 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 $(2/4841) \times 0.08\% \div 3 = 1e-7$. The probability of an *in situ* barcode matching to two MAPseq barcodes is $(20/4841) \times 0.08\% \times 2 \div 3 + (2/4841) = 4e-4$. 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 $10/13581 \times 0.08\% \div 3 = 2e-7$, and the probability of an ambiguous match in XC28 is $(212/13581) \times 0.08\% \times 2 \div 3 + (10/13581) = 7e-4$. No actual ambiguous match was seen in XC28. Therefore, allowing one mismatch for a 15-base sequence was sufficient to match barcodes in the somata 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 bulk projection patterns

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) (experiment 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 projec-

then we eliminated the target projection pattern. We repeated this process for all projection patterns, and counted the remaining projection patterns. Using a 10 – 15% false negative rate, 237 – 247 of the original 264 patterns remained.

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; Figure 4E), 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 (Figure S5B) 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. 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 (Figure S5C). We chose a k value that captured most of the variance in the data (Figure S5D) and resulted in similar classification of neurons in the first two hierarchies (Figure S5E). 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 (Figure S5B), 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 https://marronwebfiles.sites.oasis.unc.edu/Matlab7Software/BatchAdjust/). 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.

This clustering did not separate all different binary projection patterns into their own clusters, suggesting that it was conservative. However, it captured high-level differences in projections that were likely reproducible, and were sufficient for our downstream analyses. Additional levels of clustering would not likely affect downstream analyses since most of our findings did not rely on the leaf clusters.

We also compared our clustering to graph-based clustering using Louvain community detection (Blondel et al., 2008) and spectral clustering (Ng et al., 2002) (Figure S5F). 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.

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 (Figure S5G).

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 (Figure S5H). 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%)

Cell

into consideration the actual low false-negative rate of BARseq. Therefore the majority of neurons were unambiguously assigned to a single cluster even when potentially missed projections were taken into consideration.

For Louvain community detection, we used a similarity matrix $\boldsymbol{S}$, in which each element $\boldsymbol{S}_{ij}$ was the difference between the Euclidean distance of projections between two neurons and the maximum Euclidean distance of projections between any two neurons 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.

All high-level subclusters were found in significant numbers across all three brains (Figure S6A), with the exception for the striatal projecting neurons in the CT branch (Leaf 25 in Figure 5C). 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-l 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. This is especially likely since we did not sample many potential target areas for PT-l neurons, such as the brain stem. 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 (Figure S6A). 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-l neurons could project to other targets. Furthermore, the 10 neurons that projected to the striatum had a laminar profile similar to that of PT-l neurons, but different from those of the layer 6 CT neurons (Figure 5C). These results suggest that most of the striatum projecting "corticothalamic" neurons were likely PT-l neurons in layer 5 and deep layer 6 in neighboring cortical areas. The lack of such neurons in XC9 and XC28 thus likely reflect differences in the locations of neurons sampled due to the differences between MAPseq and BARseq.

Laminar distribution 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 i^{th} 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 (Figure S6B; $p > 0.05$ 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 (Figure S6C; $p > 0.05$ after Bonferroni correction, Mann-Whitney U test). The one exception was CT neurons, which were on average 50 μm deeper in XC28 than in XC9 (Figure S6C; $p < 0.0005$ 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; $p < 1e\text{-}7$ using Fisher's exact test). Therefore, this observed lack of laminar specificity of subclusters was consistent across samples.

Clustering of single-cell RNaseq

We analyzed 4471 cells with more than 3000 UMI, more than 800 unique RNAs, and less than 15% mitochondria RNA counts. We then linearly regressed out the effect of the percentage of mitochondria RNA counts, and clustered using the top 40 principal components using a resolution of 1 in Seurat. This produced seven clusters (735 cells) that appeared to be neurons with high *Snap25* expression. We then pooled these neuronal clusters together and performed a second round of clustering using the top 40 principal components of this subset of data using a resolution of 0.3. This produced the seven neuronal clusters shown in Figures 6A and 6B.

We then compared the clusters obtained to those obtained in the visual cortex (Tasic et al., 2018) using MetaNeighbor (Crow et al., 2018). MetaNeighbor was performed on a set of 540 highly variable genes. In this comparison, L4 IT and L5 IT clusters from Tasic

contained CT neurons, L6b neurons, and seven NP neurons. Furthermore, four PT-l neurons were also co-clustered with IT3 neurons; the small number of PT-l neurons were likely caused by bias during single-cell dissociation and droplet formation. Because our goal was to identify genes differentially expressed across subtypes of IT neurons, we did not attempt to further segregate these neuronal types by clustering, nor did we attempt to enrich for PT-l neurons.

Gene expression in barcoded neurons

The raw sequencing data of barcoded and non-barcoded cells were processed using Cell Ranger v2 (10× Genomics) to generate the expression matrix. We then further processed the data using Seurat v2. We first filtered out cells with less than 200 unique genes per cell and more than 10% mitochondria RNA per cell) and log normalized the data. We then regressed out either the effect of the percentage of mitochondria genes linearly (Figure S2A) or both the effect of the percentage of mitochondria genes and endogenous UMI counts (i.e., excluding barcode reads) using a poisson model (Figure 3G). We then identified barcoded neurons and non-barcoded neurons by thresholding the level of *Snap25* expression and barcode counts. We observed a general reduction in gene expression across most genes, but the relative level of expression among these genes were preserved in barcoded cells (Figure S2A). Although a minor fraction of genes were overexpressed in barcoded cells, we could further correct for both the under-expression among most genes and overexpression in this subset of genes in the barcoded neurons by simply regressing out the effect of endogenous read counts per cell (Figure 3G), thus restoring the relative gene expression level to the level of the non-barcoded cells. These results indicate that the relative gene expression level is preserved in barcoded neurons.

To further confirm that the relationship among expression of genes were preserved in individual cells, we examined the amount of variance explained by principal components in neurons (Figure S2B). We reasoned that if Sindbis infection perturbs major structures in gene expression, then principal components of gene expression in non-barcoded neurons should fail to capture the variance of gene expression in barcoded neurons. We identified genes with variable expression and exported the expression of these genes for both barcoded and non-barcoded cells to MATLAB. We then scaled and centered the data in MATLAB and calculated the explained variance of top principal components. Our results (Figure S2B) showed that the same top principal components explained similar amount of variance in both barcoded and non-barcoded neurons, thus further indicating that the structures in gene expression is preserved in barcoded neurons. Heatmaps of top genes of principal components (Figure S2C) were generated in Seurat.

Data processing for BARseq/FISH validation

We 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 99 cells were analyzed for projection pattern and gene expression.

Of the 99 barcoded cells with high sequencing quality and good morphology (Figure 3F; Table S2), 80 were excitatory neurons (*Slc17a7*+ and *Gad2*-), 3 were inhibitory neurons (*Slc17a7*- and *Gad2*+), and 16 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 99 cells, 54% (54/99) 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 54 projection neurons identified by BARseq expressed *Slc17a7*. The non-projecting excitatory neurons likely projected locally or to nearby cortical areas we did not sample. Given the small number of inhibitory neurons recovered in this experiment, our data cannot either support or refute recent findings of inhibitory projection neurons in the cortex (Rock et al., 2016; Zurita et al., 2018).

Projections of IT subtypes defined by genes

The projection barcode counts were normalized by spike-ins in the same way as the BARseq-only dataset. The gene expression data were normalized by the mRNA counts of *Slc17a7* in each cell.

To assign neurons to clusters defined by the BARseq-only dataset, we first combined barcode counts for the superficial layer and deep layer of each cortical area. We then scaled the barcode count in each area so that the mean projection barcode count in each area is the same in the BARseq/FISH dataset compared to the BARseq-only dataset. We then used a random forest classifier trained on the BARseq-only projection data to predict clusters of neurons in the BARseq/FISH dataset.

To quantify the laminar bias of projections to cortical areas, we defined laminar bias as $BC_{superficial}/(BC_{superficial} + BC_{deep})$, where $BC_{superficial}$ and BC_{deep} indicates barcode counts in the superficial and deep layers of a cortical area. A bias of 1 thus indicates pro-

Cell

To quantify whether ipsilateral and contralateral projections were more likely to co-occur or were mutually exclusive, we compared the fraction of co-projecting neurons to the fraction of either ipsilateral-only or contralateral-only projection neurons between pairs of IT subtypes defined by gene expression. Among ipsilateral projecting neurons, IT3 neurons were less likely to co-project contralaterally than both IT1 (p = 4e-6) and IT2 neurons (p = 4e-5); IT4 neurons were also less likely to co-project contralaterally than both IT1 (p = 8e-23) and IT2 neurons (p = 4e-17). Among contralaterally projecting neurons, IT3 was more likely to exclude ipsilateral projections than IT1 neurons (p = 0.02). No other significant bias was found among the contralaterally projecting neurons (p = 0.3 comparing IT1 to IT4; p = 1 comparing IT2 to either IT3 or IT4). All statistical significance was obtained using Fisher's exact test after Bonferroni correction. Therefore, IT3 and IT4 neurons that project ipsilaterally were more likely to exclude contralateral projections than those in IT1 and IT2.

Euclidean distance was used when comparing distances between the projections of IT neurons within or across subtypes defined by gene expression (Figure 7G). As a control the same analysis was done on IT neurons within or across high-level projection subclusters (Figure S7E). As expected, most neurons were more similar to neurons within the same projection subcluster compared to neurons in different projection subclusters.

Statistical analyses

All *p* values were produced using the statistical tests as noted in the text. All statistical tests were two-sided. Bonferroni correction was applied to cases involving multiple comparisons, and the corrected p values were reported as indicated in the main text.

DATA AND CODE AVAILABILITY

The accession numbers for all *in vitro* high throughput sequencing datasets reported in this paper (including two BARseq brains, XC9 and XC28, one MAPseq brain, XC14, one brain for the comparison between BARseq and retrograde tracing, XC54, three brains for the combination of BARseq with FISH, XC66, XC75, and XC91, and one brain for BARseq in *Cre*-labeled neurons, XC92) are SRA: PRJNA557267 and SRA: PRJNA448728. *In situ* sequencing data, images for the MAPseq dissections, and a list of Allen Reference Atlas slices corresponding to the dissection slices are available from Mendeley Data (doi: 10.17632/mk82s9x82t.1, 10.17632/g7kdxznt6w.1, 10.17632/86wf7xfz5x.1, 10.17632/2w649fccnt.1). Analysis code and data needed to reproduce all analyses are available from Mendeley Data (doi: 10.17632/mk82s9x82t.1, 10.17632/g7kdxznt6w.1, 10.17632/86wf7xfz5x.1, 10.17632/2w649fccnt.1).

Supplemental Figures

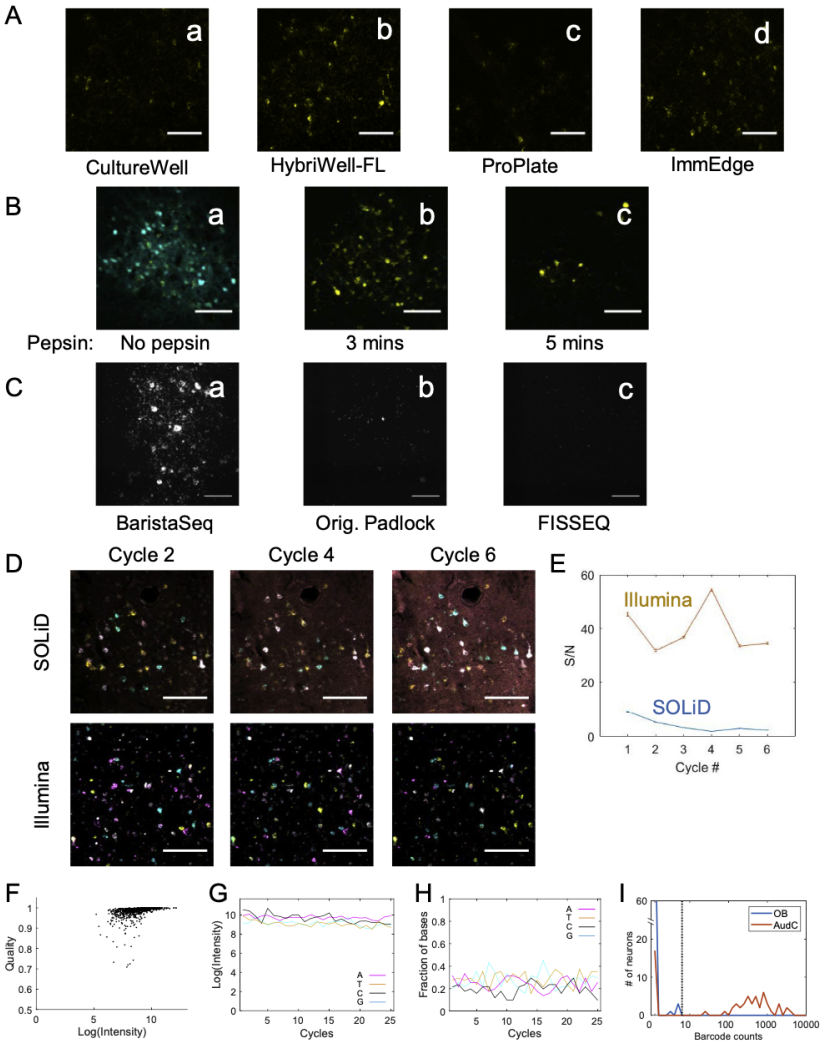


Figure S1. Optimization of *In Situ* Barcode Sequencing for Brain Slices, Related to Figure 2 and STAR Methods

(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. (I) The distribution of barcode counts in contralateral auditory cortex compared to those in the olfactory bulb (negative control). Dashed line indicates threshold above which a projection is considered real.

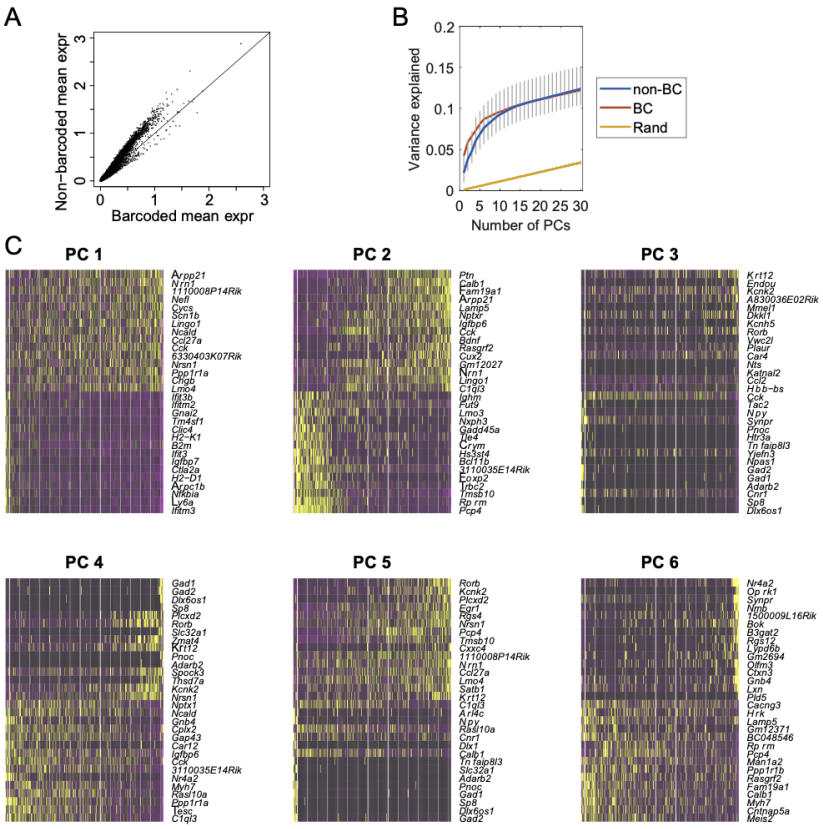
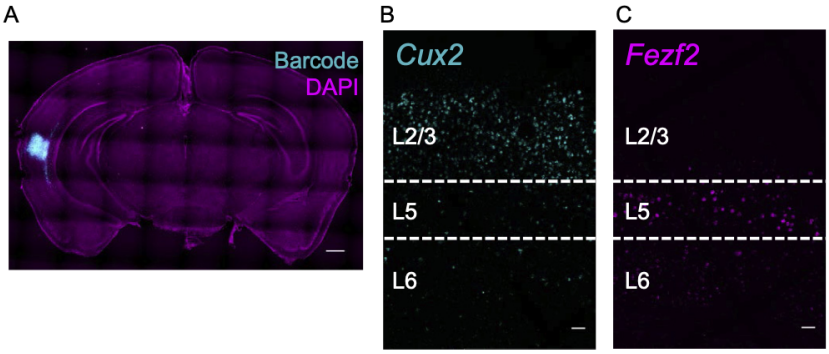


Figure S2. Comparison of Gene Expression in Barcoded Neurons to that in Non-barcoded Neurons, Related to Figure 3

(A) Mean log normalized expression of each gene averaged over all barcoded (x axis) and non-barcoded (y axis) neurons. The gene expression is regressed linearly using the percentage of mitochondrial genes. The diagonal line indicates equal expression in barcoded versus non-barcoded cells. (B) Fraction of variance explained by the same top principal components (x axis) in non-barcoded neurons (blue), barcoded neurons (red), and randomized non-barcoded neurons (yellow). Vertical bars along the non-barcoded line indicate 95% confidence interval. (C) Heatmaps of the expression of top genes in the first six principal components from non-barcoded neuros.



D

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

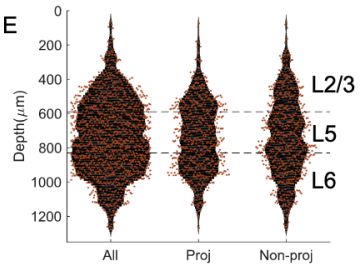


Figure S3. Locations of Neurons in the Auditory Cortex Mapped by BARseq, Related to Figure 4 and STAR Methods

(A) Representative low-resolution image of a brain slice showing barcoded GFP+ cells (cyan) and DAPI staining in a typical injection site in the auditory cortex for BARseq and MAPseq experiments. This image was not acquired in any BARseq or MAPseq brains because the workflow of these experiments does not preserve whole brain slices. Scale bar = 500 μm. (B)(C) Representative images of FISH against *Cux2* (B) and *Fezf2* (C) in two adjacent slices. Scale bars = 50 μm. (D) 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. (E) Violin plots of the laminar distribution of all BARseq neurons in XC9 and XC28 (All) and those with (Proj) or without (Non-proj) detected projections. Individual neurons (red) are plotted on top of smoothed distribution plots.

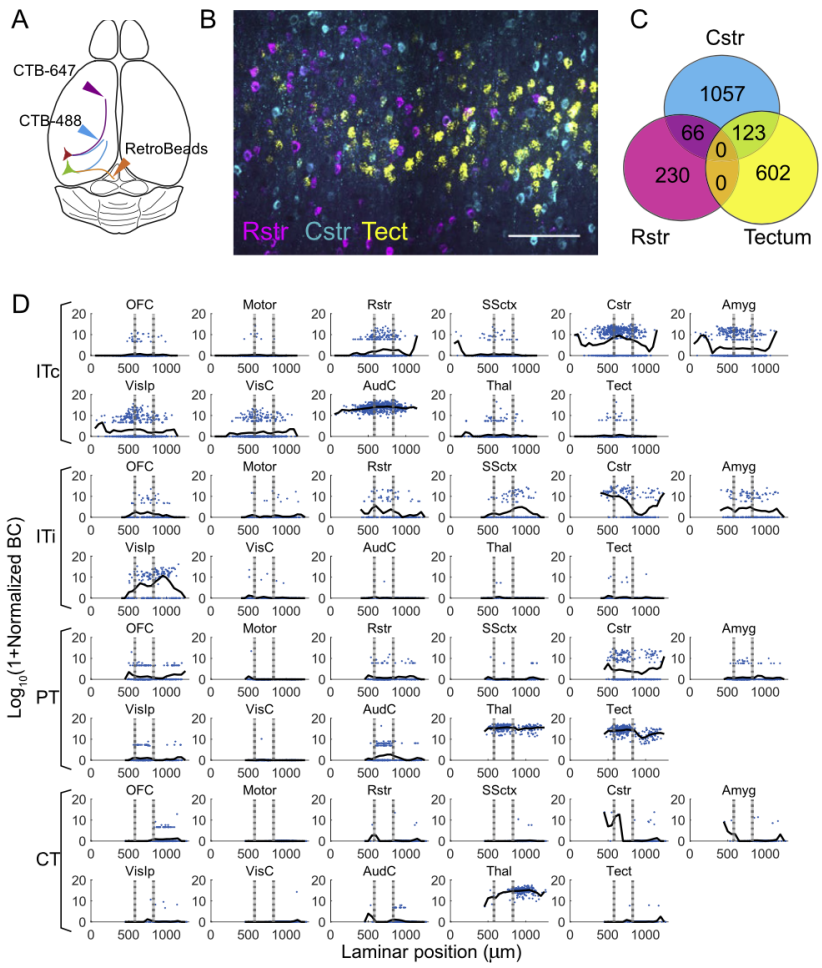


Figure S4. Projections of the Auditory Cortex Revealed by BARseq, Related to Figure 5

(A) Triple retrograde tracing of neurons projecting to the rostral striatum (CTB-647), the caudal striatum (CTB-488), and the tectum (RetroBeads). (B) A representative image of 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. (C) Venn diagram showing the number of neurons projecting to each of the three areas. (D) 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 lines. The layer boundaries are marked by dotted vertical lines.

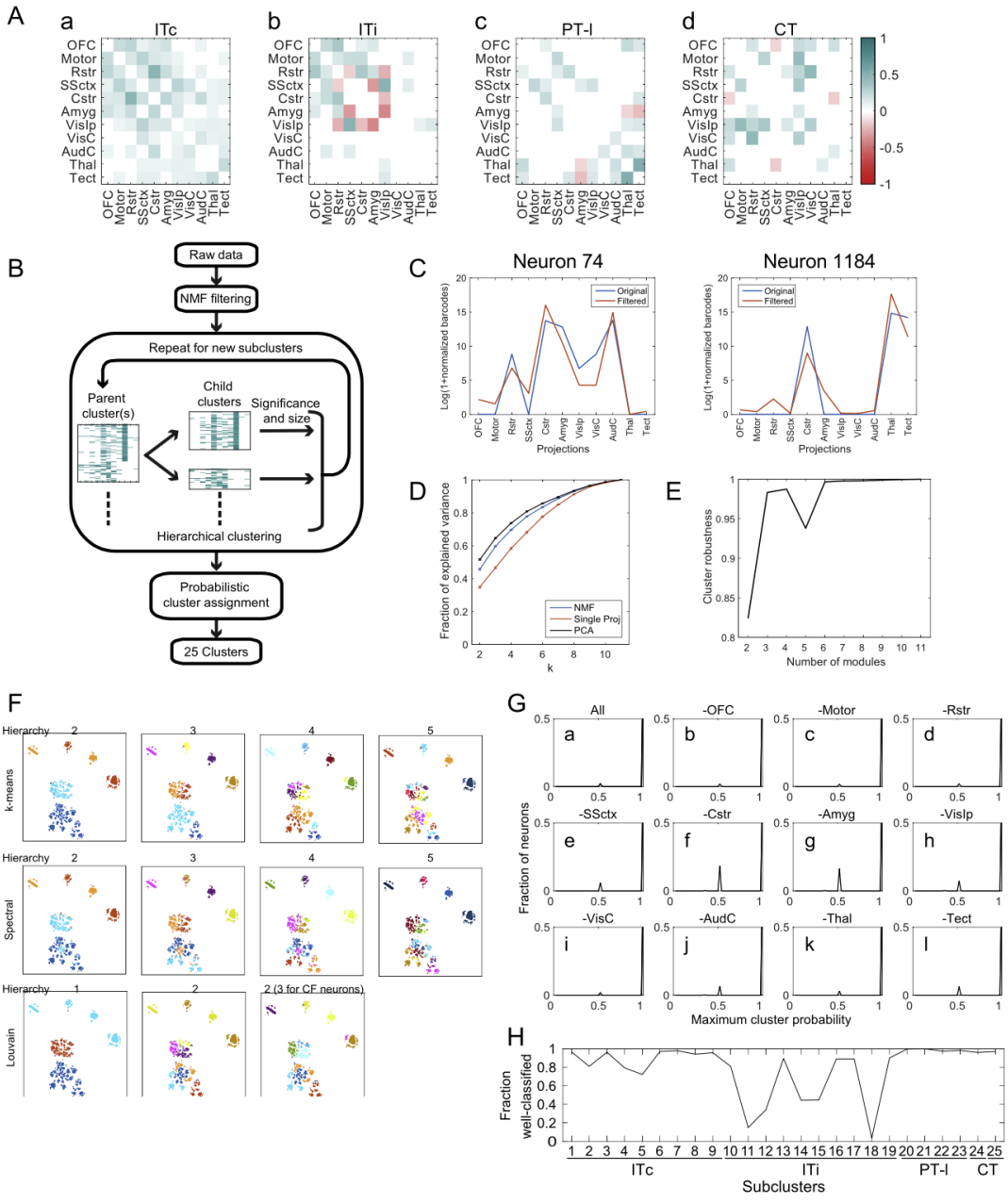


Figure S5. Hierarchical Clustering of Projection Neurons in the Mouse Auditory Cortex, Related to Figure 5 and STAR Methods

(A) Positive (green) or negative (red) Spearman correlation coefficients among projections to the indicated areas in the indicated classes. Only correlations that were statistically significant after Bonferroni correction were shown. (B) The workflow of the hierarchical clustering. (C) Comparison of the original projection strengths (blue) and the filtered projection strengths (red) for two example neurons. (D) The fraction of variance explained (y axis) using non-negative matrix factorization (NMF; blue), individual projections (red), and PCA (black). (E) The fractions 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. (F) 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 are color-coded onto the same t-SNE plot. The colors are randomly assigned to individual clusters. (G) 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. (H) The fraction of well-classified neurons in projection leaf subclusters. The subcluster labels correspond to those in Figure 5C and the major classes they belong to are labeled below.

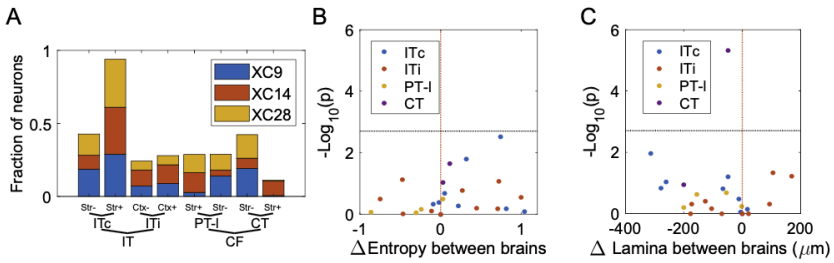


Figure S6. Consistent Projections and Laminar Distribution across Brains, Related to Figure 5

(A) The fractions of neurons from each brain that belong to each indicated high-level projection subcluster. All bars belonging to a brain sum to 1 across the whole plot. (B) 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 are color-coded according to their class-level divisions as indicated. The p values are shown without multiple testing correction. The red vertical dashed line indicates no difference in entropy, and the black horizontal dashed line indicates significance level after Bonferroni correction. (C) 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 are color-coded according to their class-level divisions as indicated. The p values are shown without multiple testing correction. The red vertical dashed line indicates no difference in the mean laminar locations, and the black horizontal dashed line indicates significance level after Bonferroni correction.

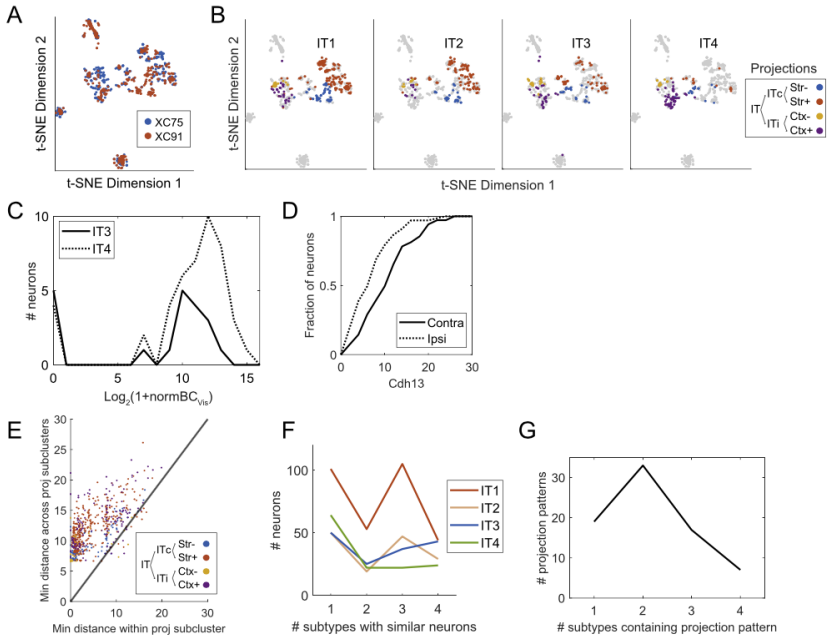


Figure S7. Projections of IT Subtypes Defined by Gene Expression, Related to Figures 6 and 7

(A) t-SNE plots of projections color-coded by brain indices. (B) t-SNE plots of projections of neurons of the indicated IT subtypes defined by gene expression. Neurons of the indicated IT subtypes are color-coded by high-level projection subclusters, and other neurons are grayed out. (C) Histograms of log normalized barcode counts of projections to the ipsilateral visual cortex in IT3 and IT4 neurons. (D) Cumulative probability distribution of *Cdh13* expression in contralateral projection neurons and ipsilateral projection neurons as determined by FISH and CTB retrograde tracing. (E) The minimum distance from a neuron to any neuron within the same projection subcluster (x axis) or in a different projection subcluster (y axis). Neurons are color-coded by high-level projection subclusters. Dashed line indicate same minimum distance within and across subtypes. (F) For each neuron belonging to an indicated IT subtype defined by gene expression, the number of IT subtypes containing neurons with the same binary projection pattern was counted (possible values range from 1 to 4 subtypes). The histogram of this value was then plotted. (G) Histogram of the number of transcriptionally defined subtypes a binary projection pattern was found in.