- 1 High-throughput mapping of long-range neuronal projection using *in situ* sequencing
- 2 Xiaoyin Chen¹, Yu-Chi Sun¹, Huiqing Zhan¹, Justus M Kebschull^{1,2}, Stephan Fischer¹, Katherine
- 3 Matho¹, Z. Josh Huang¹, Jesse Gillis¹, and Anthony M Zador^{1*}
- 4
- ¹Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
- ⁶ ²Watson School of Biological Sciences, Cold Spring Harbor, NY 11724, USA
- 7 *Correspondence: zador@cshl.edu
- 8 Summary

9 Understanding neural circuits requires deciphering interactions among myriad cell types 10 defined by spatial organization, connectivity, gene expression, and other properties. 11 Resolving these cell types requires both single neuron resolution and high throughput, a 12 challenging combination with conventional methods. Here we introduce BARseq, a 13 multiplexed method based on RNA barcoding for mapping projections of thousands of 14 spatially resolved neurons in a single brain, and relating those projections to other properties 15 such as gene or Cre expression. Mapping the projections to 11 areas of 3579 neurons in 16 mouse auditory cortex using BARseq confirmed the laminar organization of the three top 17 classes (IT, PT-like and CT) of projection neurons. In depth analysis uncovered a novel 18 projection type restricted almost exclusively to transcriptionally-defined subtypes of IT 19 neurons. By bridging anatomical and transcriptomic approaches at cellular resolution with 20 high throughput, BARseq can potentially uncover the organizing principles underlying the 21 structure and formation of neural circuits.

22

23 Key words: high throughput, projection mapping, cellular barcoding, in situ sequencing,

24 auditory cortex, cell type

26 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.

32 A particular challenge is to map the long-range axonal projection patterns that form the physical 33 basis for neuronal circuits across brain areas. Traditional neuroanatomical methods based on 34 microscopy can be used to visualize neuronal morphology, including long-range projections, but 35 the throughput of these methods remains low despite recent advances in methodology (up to ~ 50 36 neurons per single cortical area) (Economo et al., 2016; Gong et al., 2016; Guo et al., 2017; Lin et 37 al., 2018; Zheng et al., 2013). Furthermore, these methods usually rely on dedicated imaging 38 platforms specifically designed for neuronal tracing, which are not widely accessible to many 39 laboratories.

To allow high-throughput projection mapping, we previously developed MAPseq (Han et al., 2018; Kebschull et al., 2016a) (Fig. 1A, *left*), a sequencing-based method capable of mapping projections of thousands of single neurons in a single brain. MAPseq achieves multiplexing by uniquely labeling individual neurons with random RNA sequences, or "barcodes." However, like most other sequencing methods, the original MAPseq protocol relies on tissue homogenization, so 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 (<u>Barcoded Anatomy Resolved by Sequencing</u>; Fig. 1A, *right*), a method that combines MAPseq with *in situ* sequencing (Chen et al., 2018; Ke et al., 2013; Lee et al., 2014) of cellular barcodes. BARseq preserves the spatial organization of neurons during projection mapping, and further allows co-registration of neuronal projections with mRNA expression 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-1; we call these neurons pyramidal tract-like because, unlike pyramidal tract neurons in the motor

- 54 cortex, these neurons in the auditory cortex do not project to the spinal cord), and intratelencephalic
- 55 (IT) neurons (Harris and Shepherd, 2015; Shepherd, 2013)—in the mouse auditory cortex and can
- 56 additionally uncover previously unknown organization beyond these major classes. BARseq
- 57 thereby complements traditional neuro-anatomical approaches by providing a high-throughput
- 58 method for linking axonal projections, spatial organization, gene expression, and potentially other
- 59 neuronal properties that define neuronal cell types and circuits.

60 **Results**

In what follows, we first demonstrate robust *in situ* sequencing of cellular barcodes in neuronal culture. We then combine *in situ* barcode sequencing in cortical slices with mapping of projections (MAPseq). We next combine BARseq with *in situ* hybridization 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.

68

69 In situ barcode sequencing in neuronal culture.

70 We set out to develop a method capable of reading out highly diverse ensembles of sequences in 71 situ, with cellular and even subcellular spatial resolution. Although several highly sensitive 72 multiplexed *in situ* hybridization methods have been described [e.g., merFISH (Chen et al., 2015), 73 seqFISH (Eng et al., 2019; Shah et al., 2016), STARmap (Wang et al., 2018), in situ sequencing 74 (Ke et al., 2013) and 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 75 of 10^2 - 10^3 nucleotides. By contrast, cellular barcodes used for projection mapping are much shorter 76 (30-nt), more diverse ($\sim 10^6$ to 10^7), and may differ from each other by just a few nucleotides. 77 78 Cellular barcodes thus cannot be distinguished using currently available multiplexed in situ 79 hybridization approaches. Spatial transcriptomics (Rodriques et al., 2019; Stahl et al., 2016) can 80 read out diverse short barcodes, but at present the spatial resolution is insufficient to resolve single 81 neurons. Laser micro-dissection combined with RNA sequencing can provide cellular resolution, 82 but the throughput is too low for high-throughput projection mapping. We therefore focused on an 83 in situ sequencing approach (Chen et al., 2018; Ke et al., 2013) to achieve both the spatial 84 resolution and the throughput needed for barcode identification in situ.

In situ sequencing is conceptually similar to conventional *in vitro* Illumina DNA sequencing (Fig.
1B), both of which consist of two basic steps: amplification and sequencing. In the amplification
step, the RNA is converted to cDNA by reverse transcription, and the cDNA is amplified using

rolling circle amplification, resulting in the formation of a small (< 1 μ m) nanoball of DNA called a "rolony." In the sequencing step, the rolonies are sequenced in parallel using four fluorescently-

90 labeled nucleotides. The nucleotide sequences are thus transformed into color sequences, and are

91 read out using multi-channel fluorescence microscopy.

92 To capture short diverse barcode sequences, we needed an approach in which the actual target 93 barcode sequence is captured and amplified (Fig. 1Ca-f). Although some in situ hybridization 94 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 95 96 unsuitable for barcode sequencing (Fig. 1Ccd). Circligase-based FISSEQ (Fig. 1Ce) can readout 97 mRNA sequences, but is not sufficiently sensitive for barcode sequencing. We therefore used a 98 targeted approach, BaristaSeq (Chen et al., 2018), in which reverse transcription was used to 99 convert the barcode sequence into cDNA, after which padlock probe hybridization followed by 100 gap-filling and ligation was used to form a circular template for rolling circle amplification (Fig. 101 1Cf). The gap-filling padlock approach allowed for efficient amplification of barcodes in neurons 102 co-expressing barcodes and GFP (Fig. 1D). We also adapted Illumina Sequencing By Synthesis 103 (SBS) chemistry for *in situ* reactions. In combination, these strategies formed the basis for a 104 protocol that allowed efficient and robust sequencing of barcoded cultured neurons (Fig. 1E).



Figure 1

107 Fig. 1. Multiplexed projection mapping using in situ sequencing. (A) Workflow of MAPseq (left) and 108 BARseq (*right*). In both MAPseq and BARseq, a barcoded viral library is delivered to the area of interest. 109 The source area and several target areas are then dissected. In MAPseq, barcodes in all dissected areas are 110 sequenced using Next-Gen Sequencing (NGS). Barcodes at the source site are then matched to those at the 111 targets to find the projection patterns of individual neurons. In BARseq, the injection site is sequenced in 112 situ, thus preserving spatial information. (B) NGS vs. in situ sequencing. In conventional NGS (top), DNAs 113 are anchored to a flow cell (blue), amplified locally, and sequenced. During each sequencing cycle, one 114 fluorescence-labeled base is incorporated using the amplified DNA as template. The sequence is read out 115 from the sequence of fluorescence over cycles. In *in situ* sequencing (*bottom*), RNA is reverse transcribed 116 and amplified *in situ* on a slide (yellow). The amplified cDNAs are then sequenced using the same chemistry 117 as NGS and imaged in situ. Sequences can be read out together with spatial information. (C) Comparison 118 of *in situ* sequencing and hybridization techniques. Methods shown on the left use multiple rounds of 119 hybridization to probe and read out multiple mRNAs. Methods shown in the middle use sequencing to 120 multiplex read out of *in situ* hybridization signals. Methods shown on the right copy target sequences from 121 the mRNA into the rolonies, allowing true sequencing of mRNA in situ. (a) In merFISH, multiple probes 122 with barcode sequences on both ends are hybridized to each target mRNA. The probes are then read out 123 using multiple rounds of hybridization using fluorescently labeled secondary probes. (b) In seqFISH, 124 multiple rounds of HCR (hybridization chain reaction)-amplified FISH are carried out on the same target 125 mRNAs, and the mRNA identities are read out using the color sequence. (c) In STARmap, a linear probe 126 and a padlock probe are hybridized to adjacent sequences on the target. When bound, the linear probe acts 127 as a splint to allow circularization of the padlock probe. The circularized padlock probe is then used as 128 template for rolling circle amplification (RCA). (d) In the non-gapfilling padlock method, the target mRNA 129 is reverse-transcribed to cDNA. A padlock probe then hybridizes to the cDNA and is circularized. The 130 circularized padlock probe is then used as template for RCA. (e) In the Circligase-based FISSEQ, the cDNA 131 is circularized using an ssDNA ligase. This circularized cDNA is then used as template for RCA. (f) The 132 gapfilling padlock method is similar to the non-gapfilling version, except that a gap between the two arms 133 of the padlock is filled by copying from the cDNA, thus allowing the actual cDNA to be sequenced. (D) 134 Representative images of barcode rolonies (yellow) generated in primary hippocampal neuronal culture 135 coexpressing barcodes and GFP (cyan). All GFP positive neurons were filled with barcode amplicons, 136 indicating efficient barcode amplification in neuronal somata. (E) Images of the first four sequencing cycles 137 of the same neurons shown in (D). The bases corresponding to the four colors and the sequences of the 138 three neurons circled in (E) are indicated to the right. In all images, scale bars = $50 \,\mu m$.

140 *In situ* barcode sequencing in brain slices

141 In situ sequencing in brain slices presents additional challenges compared to sequencing in 142 cultured neurons. We developed a chamber system which enabled convenient handling and was 143 compatible with the amplification reactions (Fig. S1A; see STAR Methods). Because barcoded 144 neurons express high levels of GFP (which interferes with imaging during sequencing), we 145 developed a protocol for tissue digestion to reduce the GFP signal from barcoded cells (Fig. S1B; 146 see STAR Methods) and to increase RNA accessibility. Together these optimizations resulted in 147 highly efficient amplification of barcodes (Fig. S1C) compared to other methods (Ke et al., 2013; 148 Lee et al., 2014). We also optimized Illumina sequencing chemistry for brain slices, resulting in a 149 ~10-fold improvement in signal-to-noise ratio (SNR) compared to sequencing by ligation 150 approaches (Ke et al., 2013; Lee et al., 2014) (Fig. S1D, E; see STAR Methods). These 151 modifications resulted in highly efficient and specific barcode amplification (Fig. 2A) and 152 sequencing (Fig. 2B) in barcoded neuronal somata in brain slices.

153 To evaluate barcode sequencing in brain slices, we sequenced 25 bases in a sample infected with a diverse (>10⁶ unique sequences) pool of barcoded Sindbis virus (Kebschull et al., 2016a; 154 155 Kebschull et al., 2016b) injected in the auditory cortex (Fig. 2C). Basecall quality (Fig. 2D; Fig. 156 S1F; see STAR Methods) and signal intensity (Fig. S1G) remained high through all 25 cycles. We 157 observed no bias toward any particular base (Fig. S1H). Read accuracy was high: Barcodes in 158 50/51 (98%) randomly selected cells matched perfectly to known barcodes in the library, and the 159 remaining cell had only a single base mismatch to the closest barcode in the library (Fig. 2E, F). 160 The per-base error rate of $1/(25 \times 51) = 0.0078$ corresponds to a Phred score of 31. Because the barcodes in this library represent a small fraction ($\sim 10^{-9}$) of all possible 25-mer barcodes (>10¹⁵), 161 162 random 25-nt sequences on average had seven mismatches to the closest known barcode, 163 indicating that our barcode reads were unlikely to be false positive matches by chance. These 164 results indicate that in situ sequencing of RNA barcodes in brain slices is both accurate and 165 efficient.





168 Fig. 2. Validation of BARseq. (A) Representative low-magnification images of a barcoded brain slice 169 expressing GFP (*left*) and rolonies (*right*) generated in the same brain slice. GFP intensity does not correlate 170 perfectly with rolony intensity due to differences in protein and RNA expression. Insets: Negative control 171 images of GFP and rolonies of a non-barcoded brain slice taken with the same exposure settings. No GFP 172 or rolonies are visible in these images. Scale bars = 50 µm. (B) Representative high-resolution sequencing 173 image of a barcoded brain slice. The color code is shown in Fig. 1E. Inset shows sequencing image of a 174 non-barcoded brain slice. Scale bar = $50 \ \mu m$. (C) Representative low-resolution images of the indicated 175 cycles of barcode sequencing in a brain slice. The sequences of the three cells base-called are indicated 176 below the images. Only bases corresponding to the images shown are capitalized and color-coded. Scale 177 bars = 100 μ m. (D) The quality of the base calls over 25 cycles of Illumina sequencing *in situ* on the 178 barcoded brain slice. The quality score is defined as the intensity of the called channel divided by the root 179 sum square of all four channels. A quality score of 1 (best) indicates sequencing signal in only one channel, 180 and a score of 0.5 (worst) indicate same intensity across all four channels. (E) Histogram of the number of 181 mismatches between the *in situ* reads and their closest matches from *in vitro* reads (*in situ*) and the number 182 of mismatches between random sequences and their closest matches from *in vitro* reads (Random). (F) An 183 example barcode read in situ and its closest match in vitro, and a random sequence and its closest match in 184 vitro. Red indicates mismatches. (G) A brain was injected with CTB in the contralateral auditory cortex 185 and barcoded in the ipsilateral auditory cortex. BARseq results of the barcoded neurons were then compared 186 to retrograde labeling by CTB. (H) A representative image of a brain slice double labeled with barcodes 187 (cyan) and CTB (magenta) from the contralateral auditory cortex. Dashed lines indicate the top and the 188 bottom of the cortex. Scale bar = $100 \,\mu\text{m}$. A magnified view of the squared area is shown in the inset. The 189 arrow in the inset indicates a GFP+ CTB+ double-labeled neuron. (I) Venn diagram showing the number 190 of GFP expressing neurons labeled with (magenta) or without (white) CTB and/or neurons found to project 191 contralaterally using BARseq (cyan). See also Fig. S1 and Table S1.

Projection mapping using BARseq

194 We next combined *in situ* barcode sequencing with MAPseq. In MAPseq (Fig. 1A, *left*), neurons 195 are barcoded using a Sindbis virus library. Both the source area containing neuronal somata and 196 target projection areas are micro-dissected into "cubelets" and sequenced. Barcodes from the target 197 areas are then matched to those at the source area to reveal projection patterns. The spatial 198 resolution of MAPseq is thus determined by the size of the cubelets. In BARseq, we perform in 199 situ sequencing of barcoded somata at the source; the target projection areas are still micro-200 dissected and sequenced as cubelets (Fig. 1A, right). BARseq thus inherits the throughput and 201 cubelet resolution of projections of MAPseq, but allows the precise somatic origin of each axonal 202 projection to be determined with cellular resolution.

203 The high sensitivity and accuracy of MAPseq is well established (Han et al., 2018; Kebschull et 204 al., 2016a; Klingler et al., 2018). In particular, the sensitivity of MAPseq is indistinguishable from 205 that of conventional GFP-based single neuron reconstruction (Han et al., 2018). To confirm that 206 BARseq maintains the high sensitivity and accuracy of MAPseq, we compared it to conventional 207 retrograde cholera toxin subunit B (CTB) tracing of contralateral projections. We injected a highly diverse (~10⁷ barcodes) viral library into the auditory cortex and CTB in the contralateral auditory 208 209 cortex (Fig. 2G, H). We used BARseq to identify contralaterally projecting neurons and the precise 210 locations of their somata. We then identified barcoded somata that were also labeled with CTB. 211 The majority (12/13, 92%) of CTB-labeled neurons were detected by BARseq, whereas only 28% 212 (12/43) of contralaterally projecting neurons identified by BARseq were labeled with CTB (Fig. 213 2I; Table S1). The strength of the contralateral projections was well above the noise threshold 214 (defined by barcode counts in the olfactory bulb, an area to which the auditory cortex does not 215 project; Fig. S1I), indicating that the higher apparent sensitivity of BARseq was not due to false 216 positives resulting from contaminating barcodes. These results indicate that BARseq retains the 217 previously reported (Kebschull et al., 2016a) high specificity and sensitivity of MAPseq (91 \pm 218 6%), and may exceed the sensitivity of conventional CTB tracing.



220

Figure 3, see also Fig. S2 and Table S2

221 Fig. 3. Correlating gene expression and projections using BARseq. (A) Barcode sequencing in Fezf2-2A-222 CreER::Rosa-CAG-loxp-STOP-loxp-tdTomato (Ai14) animals. Top row: images of Fezf2+ tdTomato 223 expressing cells driven by Fezf2-2A-CreER::Ai14 (left), GFP expressing barcoded cells (middle), and 224 merged image of the two (right). Scale bar = $100 \,\mu m$. Bottom row: sequencing images of the indicated 225 cycle of the area indicated in the merged image. The three arrows indicate tdTomato-expressing barcoded 226 neurons. (B) Projection patterns of all neurons (left) and Fezf2+ neurons (right). Rows indicate neurons and 227 columns indicate projection areas. (C) The cortical depth distribution of barcoded Fezf2-2A-CreER::Ai14 228 labeled somata. Neurons are color-coded by projections as CT (thalamic projections), PT-l (tectal and 229 thalamic projections). IT (callosal projections), and Str (striatal projections only), (D)(E) Representative 230 high-resolution image of FISH (left) against Slc17a7 (magenta) and Gad2 (cyan), and low-resolution 231 sequencing images of cycle 1, 8, and 15 of the same sample are shown in (D). The three cells circled in (D) 232 are base-called in (E) with their expression profile of Slc17a7 and Gad2 shown. Scale bars = 50 μ m. (F) 233 The number of neurons with (Proj+) or without (Proj-) projections in excitatory neurons (S+G-), inhibitory 234 neurons (S-G+), and non-neuronal cells (S-G-). No cell was found expressing both Slc17a7 and Gad2 (S+ 235 G+). (G) Mean log normalized expression of each gene averaged over all barcoded (x-axis) and non-236 barcoded (y-axis) neurons. The gene expression is regressed with a Poisson model to remove the effect of 237 both the percentage of mitochondrial genes and endogenous UMI counts. The diagonal line indicates equal 238 expression in barcoded vs. non-barcoded cells. See also Fig. S2 and Table S2.

239 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 projections
to past or current gene expression by combining with marker-based *Cre* driver lines or *in situ*hybridization.

244 To demonstrate BARseq in Cre-labeled neuronal subpopulations, we performed BARseq in the 245 auditory cortex in a Fezf2-2A-CreER::Rosa-CAG-LSL-tdTomato (Ai14) mouse induced at age 6 246 and 7 postnatal weeks (Fig. 3A). Fezf2 is predominantly expressed in layer 5 PT-l, but is also 247 expressed at lower levels in some layer 5 IT neurons and layer 6 CT neurons (Tasic et al., 2018). 248 Projection patterns obtained from bulk tracing were mostly those expected from PT-l neurons, but 249 also included some callosal projections, suggesting a small contribution from Fezf2-expressing IT 250 neurons (Matho et al., in preparation). We sequenced 1291 projection neurons with good 251 sequencing quality in situ, including 72 neurons co-labeled by tdTomato driven by Fezf2-2A-252 CreER, and assayed their projection patterns to 11 target areas (see STAR Methods). The Fezf2+ 253 population included 53 (74%) PT-l neurons projecting to the tectum and the thalamus, eight (11%) 254 CT neurons projecting to the thalamus only, nine (13%) IT neurons in layer 5 projecting to the 255 contralateral cortex, and two (3%) neurons in layer 5 projecting to the striatum (Fig. 3B, C). The 256 lack of corticotectal projection in the callosal projecting neurons is unlikely an artifact caused by 257 limited detection sensitivity of corticotectal projections, because these neurons also project to other 258 telencephalic areas that PT-l neurons do not. Neurons labeled with Fezf2-2A-CreER thus consisted 259 mainly of PT-l neurons, along with a small fraction of callosally-projecting neurons in layer 5 and 260 CT neurons in layer 6. The projection patterns of Fezf2+ neurons were thus consistent with 261 previous bulk cell-type specific tracing results using the same Cre line (Matho et al., in preparation). 262 These results demonstrate the combination of BARseq with Cre-dependent labeling of neuronal 263 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 fluorescent *in situ* hybridization (FISH) against Slc17a7 [a marker of excitatory neurons;

(Tasic et al., 2016)] and Gad2 (a marker of inhibitory neurons) in the same cells (Fig. 3D, E).
Consistent with the fact that most projection neurons in the cortex are excitatory, we identified 54
neurons with long-range projections, all of which expressed Slc17a7, but not Gad2 (Fig. 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.

274 We next assessed whether barcoding disrupted the transcriptomic landscape of single neurons. We 275 performed single-cell RNAseq in 398 barcoded and 1088 non-barcoded cells dissociated from the 276 mouse auditory cortex. Although fewer endogenous mRNA molecules were recovered from 277 barcoded compared to non-barcoded neurons, the relative levels of endogenous gene expression 278 remained largely unchanged in barcoded cells (Fig. 3G; Fig. S2A; see STAR Methods for details). 279 The fact that relative levels were preserved is compatible with a model in which infection with the 280 barcoded viral vector affects the expression of most endogenous genes about equally. Importantly, 281 top principal components of genes identified in the non-barcoded neurons were equally effective 282 in describing gene expression in barcoded neurons (Fig. S2B). These principal components 283 contained many known neuronal subtype markers (Fig. S2C). These results indicate that the 284 expression pattern of genes, especially those that contribute to cell typing, remain largely intact in 285 barcoded cells.

286

287 **Projection diversity in the mouse auditory cortex**

288 Having established the sensitivity and specificity of BARseq, we applied it to study the 289 organization of long-range projections from the mouse auditory cortex. We performed BARseq on 290 two brains (XC9 and XC28) and MAPseq on an additional brain (XC14). We allowed one 291 mismatch when matching *in situ* barcodes to the projection barcodes (see STAR Methods; Fig. 292 4A; Fig. S3A), and found 1806 (55%) of 3237 cells sequenced in situ projected to at least one 293 target area (Table S3). We further excluded cells obtained from tissue deformed during processing, 294 and from labeled cells outside of the cortex. This resulted in 1309 neurons sequenced in situ and 295 5082 neurons sequenced *in vitro* using conventional MAPseq (Table S3; Fig. 4B). 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 6391 neurons had non-zero barcode counts in the olfactory bulb (four neurons with a count of one molecule, and one neuron with a count of four molecules), indicating that the technical false-positive rate for a single target area is extremely low at 5 / 6391 = 0.08%. Using the false-positive rate and the number of barcodes with zero counts in each area, we estimated that a total of 40 out of 18851 projections detected in our dataset might be false positives, corresponding to a false-discovery rate of 0.2% (i.e., for each 1000 detected projections, two were likely false positives). BARseq thus has a very low false discovery rate.

313 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 314 315 contribution of potential artifacts, including those due to non-uniform barcode transport and 316 variable barcode expression strength, have been carefully quantified in previous work (Han et al., 317 2018; Kebschull et al., 2016a). Although one potential challenge of MAPseq arises from fibers of 318 passage, in practice these have not represented a major source of artifact (Han et al., 2018; 319 Kebschull et al., 2016a). One reason is that the total amount of axonal material due to fibers of 320 passage is typically small compared with the rich innervation of a fiber at its target. A second 321 reason is that the barcode carrier protein is designed to be enriched at synapses (Kebschull et al., 322 2016a), which further minimizes the contribution of passing fibers to barcode counts. To determine 323 the extent of contamination from fibers of passage, we examined the striatal projections of putative 324 CT neurons (i.e. neurons that project to the thalamus, but not to any cortical area or the tectum)

325 and putative PT-l neurons (i.e. neurons that project to the tectum, the thalamus, but not to any 326 cortical area) from the two BARseq brains (the MAPseq brain was excluded because it may include 327 PT-l neurons from neighboring areas that did not project to the tectum, and may be misidentified 328 as CT neurons). Both types of neurons send descending fibers through the internal capsule adjacent 329 to the caudal striatum, and thus can be used to evaluate fibers of passage contamination in the 330 striatum. Consistent with the fact that PT-l but not CT neurons project to the striatum, only 8 / 250 331 (3.2%) putative CT neurons sequenced in situ projected to the caudal striatum, whereas 59 / 245 332 (24%) putative PT-l neurons sequenced *in situ* projected to the caudal striatum. The fact that ~97% 333 of putative CT neurons did not show projection to the caudal striatum reinforce previous 334 observations that fibers of passage have minimal impact on projections mapped in these 335 experiments.

336 At the mesoscopic level, the projection patterns of the auditory cortex revealed by BARseq were 337 consistent with anterograde tracing in the Allen Mouse Connectivity Atlas (Oh et al., 2014). In 338 MAPseq and BARseq, the strength of the projection of a neuron to a target is given by the number 339 of individual barcode molecules from that neuron recovered at the target. This is analogous to a 340 conventional fluorophore mapping experiment, in which projection strength is assumed to be 341 proportional to GFP intensity. Consistent with conventional bulk GFP tracing (Oh et al., 2014) 342 (Fig. 4C; Pearson correlation coefficient r = 0.94, p < 0.0001), projections from the auditory cortex 343 to the thalamus, the tectum, the contralateral auditory cortex, and the caudal striatum were 344 particularly strong (Fig. 4D). BARseq thus provides quantitative measurements of long-range 345 projections in the auditory cortex consistent with those obtained by conventional bulk labeling 346 techniques.

347 The projection patterns of individual neurons were remarkably diverse. To quantify this diversity, 348 we binarized projection patterns, using a conservative threshold for projection detection (four 349 molecules per barcode per area, given by the OB negative control). We observed 264 distinct 350 patterns, or 13% of the $2^{11} = 2048$ possible patterns. This high diversity was unlikely an artifact of 351 projections missed by BARseq due to false negatives: 237 to 247 unique projection patterns remain 352 even when a false negative rate of 10-15% was assumed (see STAR Methods). This high diversity 353 was also unlikely to be caused by limited sensitivity of MAPseq for neurons with fewer barcodes 354 overall: A significant fraction of these projection patterns (112/264 = 42%) can be accounted for

solely by strong projections (see STAR Methods), and eliminating neurons with weaker primary projections resulted in only a moderate reduction in the diversity of projection patterns (Fig. 4E, blue line). Furthermore, this reduction was largely due to the reduction in sample size (Fig. 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.

361 The majority (85%) of cortical projection neurons projected to two or more target areas (Fig. 4F) 362 after binarization. The ratio of the projection strengths of the two strongest projections follow a 363 log normal distribution (Fig. 4G), and about half of all neurons have a secondary projection that is 364 at least 20% as strong as their primary projections. These estimates were based on neurons whose 365 primary projections were strong enough to observe a secondary projection, and therefore were not 366 limited by the sensitivity of MAPseq/BARseq or by the variability in the expression level of 367 barcodes. This high projection diversity and multiple projections per neuron are consistent with 368 findings in the visual cortex (Han et al., 2018), and may be a general feature of cortical projections.



370

371 Fig. 4. Mapping projections of the auditory cortex using BARseq. (A) Histogram of the minimal pairwise 372 hamming distance of the first 15 bases of barcodes recovered from brain XC9. (B) Single-cell projection 373 patterns sorted by clusters. Each row represents a barcode and each column represents projection strengths 374 to the indicated brain area. OFC: orbitofrontal cortex; Motor: motor cortex; Rstr: rostral striatum; SSctx: 375 somatosensory cortex; Cstr: caudal striatum; Amyg: amygdala; VisIp: ipsilateral visual cortex; VisC: 376 contralateral visual cortex: AudC: contralateral auditory cortex: Thal: thalamus: Tect: tectum. (C) 377 Conventional bulk GFP tracing intensities (x-axis) were plotted against the bulk projection strength 378 obtained from BARseq (y-axis). Error bars indicate SEM. N = 5 for GFP tracing and N = 3 for BARseq. 379 Pearson correlation coefficient r = 0.94, p < 0.0001. (D) The distribution of projection intensity in each 380 projection area. The y-axis indicates the logarithms of raw barcode counts in each area, and the x-axis 381 indicates the number of cells. (E) The numbers of binarized projection patterns (y-axis) after filtering for 382 primary projection strength (x-axis). Blue line indicates filtering at the indicated thresholds, and red line 383 indicates random subsampling to the same sample size. Black lines and error bars indicate 95% confidence 384 interval for subsampling. (F) Histogram of the number of projections per neuron. (G) The fractions of multi-385 projection neurons (y-axis) are plotted against the ratio between the secondary and primary projections (x-386 axis). Blue line indicates actual distribution and red line indicates fitting with a log normal distribution. See 387 also Fig. S3 and Table S3.



388

Figure 5, see also Fig. S4, S5, S6

389 Fig. 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 390 391 projection strength normalized to the strongest projection for that class. Blue arrows indicate projections to 392 contralateral brain areas and black arrows indicate projections to ipsilateral brain areas. (B) The sequenced 393 projection neurons from a brain (XC9) are color-coded by class identities and plotted at their locations in 394 the cortex. The top and bottom of the cortex are indicated by the red and blue dashed lines, respectively. 395 The laminae and their boundaries are marked. Scale bar = $100 \mu m$. Inset: histograms of each class of 396 projection neurons in the pooled BARseq dataset. The y-axis indicates cortical depth and the x-axis 397 indicates the number of neurons at that depth in a particular class. The laminae are indicated on the right 398 and the boundaries between two layers are marked by dashed lines. (C) Hierarchical clustering of single-

399 cell projection data. Top: dendrogram of the hierarchical structure of the clusters. An index is assigned to 400 each leaf as indicated. Nodes representing major classes are labeled by class names, and nodes representing 401 finer divisions are labeled by major differences in projections. Middle: the mean projection patterns of the 402 corresponding leaf cluster. Each column represents a leaf cluster. Each row represents projection to the 403 indicated area as labeled on the left. *Bottom*: The laminar distribution of the corresponding leaf clusters. 404 Individual neurons are superimposed on top of the distribution plots (light grey). Neurons whose cluster 405 identity were less confident were marked in gray. The number of cells that belong to each leaf cluster is 406 indicated below. Neurons of subcluster 25 were likely misidentified PT-l neurons (see STAR Methods). 407 (D) t-SNE plot of the projection neurons. The neurons are color-coded by their first level subcluster 408 identities post-hoc. (E) The normalized entropy of nodes/leaves (y-axis) in the indicated clustering 409 hierarchy (x-axis). The branch of the subcluster nodes/leaves were color coded as indicated. Grev bars 410 indicate mean \pm stdev of all nodes/leaves of a specific hierarchy. Hierarchies 1-3 correspond to class 411 divisions and hierarchies 4-7 correspond to subcluster divisions. See also Fig. S4-S6.

413 **BARseq recovers structured projections across major classes of projection neurons**

414 Simple hierarchical clustering on the non-binarized projections (see STAR Methods) partitioned 415 neurons into established (Harris and Shepherd, 2015) cortical classes: CT (corticothalamic), PT-1 416 (pyramidal tract-like), and IT (intratelencephalic, Fig. 5A). CT neurons project to the thalamus; 417 PT-l neurons project to both the thalamus and the tectum; and IT neurons project to the cortex but 418 not subcortically to the thalamus or tectum. IT neurons were further subdivided into those that 419 project to both the ipsilateral and contralateral cortex (the ITc class), and those that project 420 ipsilaterally only (the ITi class). These patterns correspond to the classically-defined top-level 421 subdivisions of cortical projection patterns.

422 The classification of cortical excitatory projection patterns into PT-l, IT and CT classes was 423 supported by the corresponding segregation of these neurons into distinct laminae. As expected, 424 the major classes segregated strongly with laminar position: Superficial layers contained 425 predominantly IT neurons, whereas deep layers contained predominantly PT-l and CT neurons 426 (Harris and Shepherd, 2015) (Fig. 5B; see STAR Methods and Fig. S3B-E for the estimation of 427 layers using layer-specific marker genes). Thus major classes of cortical projection neurons 428 defined by BARseq had distinct laminar distributions, consistent with previous observations using 429 conventional methods (Harris and Shepherd, 2015).

430 Our data revealed two interesting features of the PT-l neurons in the auditory cortex. First, although 431 IT neurons projected to both the caudal striatum—an area implicated in auditory decision making 432 (Znamenskiy and Zador, 2013)—and the rostral striatum, PT-l neurons projected only to the caudal 433 striatum. We confirmed this finding by triple retrograde tracing (Fig. S4A): A significant fraction 434 of neurons (66 / 296) projecting to the rostral striatum also projected to the caudal striatum, but 435 none (0/296) projected to the tectum (Fig. S4B, C). Second, we identified a minor population of 436 PT-l neurons in layer 6 of the auditory cortex in addition to the more common layer 5 PT-l neurons 437 (Fig. 5B). This population of layer 6 PT-l neurons has previously been reported to differ from the 438 layer 5 PT-l neurons based on physiological and other properties (Slater et al., 2013). Similar to 439 the differences observed between the layer 5 and layer 6 PT-l neurons, the corticotectal projections 440 were slightly stronger in layer 5 than those in layer 6, but with considerable overlap [Fig. S4D] 441 (PT-Tect); 11.7 \pm 2.2 (mean \pm stdev), N = 80 for layer 6 neurons and 13.8 \pm 1.6, N = 185 for layer

5 neurons, p < 0.0001 using bootstrap K-S test]. The strengths of other major projections were indistinguishable between the two groups (Fig. 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.

446

447 **Projection subclusters do not segregate by laminae**

Even within the major classes, projections from the auditory cortex to individual targets were often correlated (Fig. S5A). For example, an ITi neuron with a strong projection to the somatosensory cortex was also likely to have a strong projection to the visual cortex, but a weak striatal projection (Fig 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.

455 To explore the statistical structure underlying these projection patterns, we performed hierarchical 456 clustering on the projection patterns. For this analysis, we did not binarize the projection patterns; 457 we considered not only whether a neuron projected to an area, but also the strength of that 458 projection. This clustering uncovered structure well beyond the previously established top-level 459 classes (Fig. 5C, D; see Fig. S5B-E and STAR Methods for details of clustering methods). 460 Depending on the precise clustering algorithm and statistical criteria used (Fig. S5F), as many as 25 subclusters were revealed. Most of these subclusters were robust: 5682/6391 (89%) of neurons 461 462 were unambiguously assigned into one of the 25 subclusters using a probabilistic approach (see 463 STAR Methods; Fig. S5G, H). All top-level subclusters were observed in all three brains (Fig. 464 S6A: see STAR Methods for likely misclassification of some striatal-projecting PT-l neurons as 465 CT neurons due to choice of projection targets sampled). BARseq thus uncovered organization in 466 projections well beyond the previously described major divisions.

467 As noted above, the three major cortical classes—PT-l, IT and CT—are spatially segregated into 468 laminae. By contrast, most subclusters defined by projection pattern, especially those within the 469 IT class, spanned many layers (Fig. 5C, E; p = 0.95 for cluster hierarchy 2-7, Kurskal-Wallis test;

see STAR Methods). For example, a subcluster consisting of IT neurons that project only to contralateral cortical areas, but not the ipsilateral cortex or the striatum (Leaf 1, Fig. 5C), can be found across all laminae. This mixed laminar distribution of projection patterns was consistent across brains (Fig. S6B, C; see STAR Methods). Therefore, laminar position could not fully explain projection patterns or diversity.

475

476 **BARseq associates projections with transcriptionally defined IT subtypes**

Although projection subclusters did not segregate cleanly into laminae, we hypothesized that the 477 478 conjunction of laminar position and gene expression might predict projection patterns. Although 479 gene expression patterns of single neurons have been studied in other cortical areas, including 480 visual, somatosensory, and motor cortices (Tasic et al., 2018; Zeisel et al., 2015), gene expression 481 patterns of auditory cortical neurons have not previously been explored. We therefore performed 482 single-cell RNAseq for 735 neurons in the auditory cortex using 10x Genomics Chromium. As 483 expected from work in other cortical regions, the top-level partition among neurons was between 484 excitatory and inhibitory classes (Fig. 6A). Within the excitatory neuron class, gene expression 485 further segregated into CT and IT subtypes. As in other studies using the 10x platform (Zeisel et 486 al., 2018), few PT-l neurons were recovered, possibly due to biases in cell survival (see STAR 487 Methods). The transcriptional taxonomy of neurons in the auditory cortex were thus consistent 488 with other sensory cortices.

489 To further explore whether transcriptionally-defined subtypes corresponded to subclusters defined 490 by projection pattern, we focused on subdivisions within the IT class. IT neurons were the most 491 diverse class based on projection mapping, with as many as 19 distinct subclusters (Fig. 5C). 492 Clustering of gene expression revealed four top-level IT subtypes, which we denote IT1, IT2, IT3, 493 and IT4 (Fig. 6A, B). These four subtypes corresponded well to the four major subtypes of IT 494 neurons previously identified in visual cortex (Tasic et al., 2018)(Fig. 6C; see STAR Methods). 495 We thus sought to determine how the 19 projection subclusters were partitioned among the 4 496 transcriptionally defined subtypes.

497 To assess the relationship between projection pattern and gene expression among IT neurons at 498 the single neuron level, we combined BARseq with fluorescence *in situ* hybridization (FISH) 499 against Cdh13, Slc30a3, and Slc17a7 (Fig. 6D). The combination of laminar position and 500 expression of these markers allowed us to assign neurons to subtypes: IT1 neurons consisted of all 501 IT neurons in layer 2/3; IT2 neurons consisted of Cdh13-negative IT neurons in layer 5; IT3 502 neurons consisted of both Cdh13-expressing neurons in layer 5 and Slc30a3-expressing neurons 503 in L6; and IT4 neurons consist of Slc30a3-negative IT neurons in layer 6 cortex (Tasic et al., 504 2018)(Fig. 6D, E; see STAR Methods). We collected the same projection areas as in the previous 505 BARseq experiments, but achieved higher spatial resolution of the projection targets by collecting 506 upper and lower cortical layers separately. We obtained good sequencing and FISH in 979 507 projection neurons from two brains (Fig. 6D; Fig. S7A), of which 735 were IT neurons defined by 508 projection (note that this set of 735 IT neurons was independent of the 735 neurons studied using 509 single-cell RNAseq; the same number of neurons is coincidental). These IT neurons were assigned 510 to one of the four genetically-defined subtypes based on laminae and gene expression, and to one 511 of the projection subclusters based on projection patterns (Fig. S7B). Interestingly, subtypes that 512 were further apart based on mean gene expression were also further apart based on clustering of 513 projections (Fig. 6A), suggesting a relationship between gene expression and projection pattern 514 within IT subtypes.



516

517 Fig. 6. Subtypes of IT neurons defined by gene expression in the auditory cortex. (A) Histograms of the log 518 normalized expression of the indicated marker genes in the indicated clusters obtained from single-cell 519 RNAseq in the auditory cortex. The dendrogram on the left shows distances of mean gene expression among 520 transcriptomic clusters, and the dendrogram on the right show distances of mean projection pattern among 521 the same subtypes of neurons obtained through BARseq and FISH. (B) t-SNE plot of the gene expression 522 of neurons color-coded by cluster identity as in (A). (C) MetaNeighbor comparison of neuronal clusters 523 obtained in the auditory cortex to those in the visual cortex from Tasic et al. (2018). (D) Projections (left) 524 and the expression of genes (right) of neurons obtained using combination of BARseq and FISH are shown 525 on a log scale. The transcriptionally defined subtype labels are shown on the left. The color bar on the left 526 indicates Log_2 values of normalized barcode counts and the color bar on the right indicates Log_{10} values of 527 normalized mRNA counts. Projection areas are the same as in Fig. 4B, except that each cortical area is 528 divided into upper (u) and lower (l) layers. (E) Distributions of laminar positions of neurons. Individual 529 neurons (red) are superimposed on the smoothed distribution (black). See also Fig. S7.

530



533 Fig. 7. Projections across IT subtypes defined by gene expression. (A) The fraction of ITi-Ctx neurons in 534 each indicated IT subtype defined by gene expression. The number of ITi-Ctx neurons in each subtype and 535 the total numbers of each subtype are indicated on top of each bar. Inset: the mean projection pattern of ITi-536 Ctx neurons. (B) Histograms of the layer bias of contralateral projections of neurons in each IT subtype. 537 The histograms are normalized so that the maximum value for a bin is 1 for each subtype. See STAR 538 Methods for the definition of projection layer bias. (C) The Log normalized barcode count of projections 539 to the contralateral auditory cortex (x-axis) is plotted against that of projections to the ipsilateral visual and 540 somatosensory cortex (y-axis) for each neuron. Neurons are color-coded by IT subtypes defined by gene 541 expression. (D) The fractions of neurons of IT subtypes defined by gene expression that belong to each 542 indicated IT projection leaf subcluster. All bars belonging to a transcriptionally defined subtype sum to 1 543 across the whole plot. The projection cluster labels correspond to those in Fig. 5C. Vertical dashed line 544 separates ITc and ITi subclusters. (E) The number of neurons with (AudC+) or without (AudC-) projections 545 to the contralateral auditory cortex in each IT subtype defined by gene expression. Neurons in each IT 546 subtype are further divided into those expressing Cdh13 and those that do not. *p < 0.005 using Fisher's 547 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 from the
 indicated subtypes. The number of neurons belonging to a projection subcluster and a transcriptionally

below defined subtype and the total number of neurons in the transcriptionally defined subtype are indicated below

each example. (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). Neurons are color-coded by

553 IT subtypes defined by gene expression. Dashed line indicates same minimum distance within and across 554 subtypes. See also Fig. S7.

556 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 novel 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.

563 Perhaps the most striking observation was the identification of a specific projection subcluster, 564 which we denote ITi-Ctx, found almost exclusively in the two deep-layer subtypes IT3 and IT4, 565 but absent in IT1 or IT2 (Fig. 7A). ITi-Ctx neurons are a subset (leaves 14 and 15 in Fig. 5C) of 566 IT ineurons which project exclusively to the ipsilateral cortex, but not to the contralateral cortex 567 or the striatum. Whereas ITi-Ctx neurons accounted for 13% (20 / 155) of IT3 neurons and 34% 568 (45 / 123) of IT4 neurons, it accounted for only 2% (6 / 303) of IT1 neurons and fewer than 1% (1 569 / 145) of IT2 neurons (Fig. 7A; p < 1e-4 comparing IT3 to either IT1 or IT2, and p < 5e-15570 comparing IT4 to either IT1 or IT2, all using Fisher's exact test with Bonferroi correction). These 571 ITi-Ctx neurons would have been difficult to detect using conventional retrograde tracing because 572 they are defined by a combinatorial projection pattern, i.e. projection to ipsilateral but not 573 contralateral cortex.

Second, we also found that when IT3 and IT4 neurons had contralateral projections, the projections specifically targeted the deep layers of the contralateral cortex (Fig. 7B; p < 5e-8 comparing IT3 to IT1 or IT2, and p < 5e-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.

581 Third, we found that if an IT3 or IT4 neuron had an ipsilateral projection, then it was unlikely to 582 also have a contralateral projection. By contrast, neurons in IT1 and IT2 often projected both ipsi-583 and contralaterally (Fig. 7C; p < 5e-4 comparing IT3 to either IT1 or IT2, and p < 5e-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 (Fig. 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.

590 Finally, we discovered finer quantitative differences across individual subtypes defined by gene 591 expression. Although we identified two projection patterns (contralateral projections to deep layers 592 and ITi-Ctx projections) shared by both IT3 and IT4, these two projection patterns were 593 differentially enriched between the two subtypes. ITi-Ctx projections were both stronger (Fig. S7C, 594 p = 0.012 comparing IT3 and IT4 using rank sum test) and more frequent in IT4 subtype compared 595 to those in IT3 (Fig. 7D; 13% in IT3 compared to 34% in IT4 neurons, p = 3e-5 using Fisher's 596 exact test). In contrast, contralateral projections were twice as frequent in IT3 neurons compared 597 to IT4 (Fig. 7D; 89 / 155 IT3 neurons compared to 34 / 132 IT4 neurons, p = 6e-8 using Fisher's 598 exact test). Within a subtype defined by gene expression, projections further correlate with gene 599 expression: The probability of projecting contralaterally was higher among IT3 neurons with high 600 Cdh13 expression (p = 0.001, Fisher's exact test, Fig. 7E), an observation further confirmed using 601 CTB retrograde tracing (Fig. S7D, p < 0.0001 using rank sum test).

602 Although the analyses above support the idea that gene expression and laminar position are 603 correlated with projection pattern, the relationship between IT subtypes defined by gene expression 604 and subclusters defined by projection patterns was in general complex (Fig. 7D). Thus, although 605 some projection patterns were enriched in some subtypes [e.g., ITi-Ctx neurons (leaf 14 and 15) 606 and ITc neurons of leaf 4 and 5; Fig. 7D], no projection subcluster was specific to a single subtype. 607 Indeed, in many cases, the same projection subcluster could be found in all four IT subtypes; two 608 examples of projection subclusters found in all four transcriptomic subtypes are shown in Fig. 7F. 609 Such shared projection patterns across subtypes were common: 54% of IT neurons (400 / 735) 610 were more similar in projections to a neuron of a different IT subtype than to any neuron of its 611 own subtype (Fig. 7G; Fig. S7E-G). This mixed projection pattern is in contrast to the clear 612 separation of projection patterns across the three higher-level classes of neurons (IT/PT-l/CT). 613 Thus whereas at the top levels of the hierarchies, the classical partitioning into IT/PT-l/CT neurons

- 614 captures the correlational structure of multiple cellular properties, including gene expression,
- 615 laminar position and projection pattern, our results suggest that further subclustering based on one
- 616 experimental property (such as gene expression) leads to categories that do not map neatly onto
- 617 clustering based on another property (such as projections).

618 **Discussion**

619 Here we have described BARseq, a barcoding-based neuroanatomical method that can relate high-620 throughput projection mapping with other neuronal properties at cellular resolution through *in situ* 621 sequencing. As a proof of principle, we applied BARseq to the projections in the auditory cortex. 622 BARseq of 3579 neurons recapitulated the organization of cortical projections into the classic 623 IT/PT-I/CT subtypes. Within these classic subtypes, BARseq also revealed the impressive 624 diversity of projection patterns: We observed 264 distinct projection patterns, falling into as many 625 as 25 distinct clusters. We then combined BARseq with single cell analysis of gene expression of 626 an additional 735 IT neurons. We identified projection patterns exclusive to, or enriched in, 627 specific transcriptionally defined subtypes. BARseq thus revealed rich and complex relationships 628 between gene expression and projection patterns that would have been difficult to uncover using 629 conventional transcriptomic and/or anatomical techniques.

630

631 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.

636 BARseq fundamentally differs from conventional optical approaches in that projections are 637 determined by matching barcodes, not tracing. Barcode matching, unlike optical tracing, does not 638 accumulate error over distance. In conventional optical approaches to mapping projections, axons 639 are reconstructed by observing continuity between successive optical or physical sections. Any 640 lost or distorted section may result in an error, and in general the probability of error increases 641 exponentially with the length of the axonal projection and the number of neurites multiplexed. For 642 example, even with a low error rate of 1% per 50 μ m, more than half of axons traced would be in 643 error within 5 mm. By contrast, because BARseq relies on matching sequences to reconstruct 644 projections, errors do not accumulate for long axons: A distant subcortical projection is just as 645 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.

653 In BARseq, the spatial resolution at which projections are resolved is determined by the size of 654 the cubelets dissected. In this study, we chose to dissect brains relatively coarsely-only sufficient 655 to distinguish among brain areas and between superficial and deep cortical layers, but this 656 resolution was sufficient to resolve the organization of projections across neuronal subtypes. For 657 questions requiring higher spatial resolution, BARseq can be further improved with laser capture 658 microdissection (Huang et al., 2018) or direct in situ sequencing of projection barcodes, thus 659 potentially resolving axonal projections and dendritic morphology at subcellular resolution. Such 660 an approach would yield an "infinite color Brainbow," allowing reconstruction of densely labeled 661 neurons.

662

663 The organization of projections across neuronal subtypes

664 The partitioning of neurons into types and subtypes is useful to the extent that these classes can 665 infer multiple neuronal properties. The utility of these classes and subclasses arises from the fact 666 that these properties co-vary. For example, knowing whether a cortical neuron is excitatory or 667 inhibitory, or the excitatory subtype (IT/PT-1/CT) to which it belongs (Harris and Shepherd, 2015), 668 allows us to infer a great deal about the neuron's pattern of gene expression, morphology, 669 connectivity and projection pattern. The partitioning of cortical neurons into excitatory vs. 670 inhibitory, and of excitatory neurons into IT/PT-I/CT subtypes, is well supported by many lines of 671 data.

Within these major subtypes of excitatory neuron, the predictive value of further subdivisions hasbeen less clear. It has been reported that some PT-l subtypes defined by gene expression

674 correspond to specific projection patterns (Economo et al., 2018), and some CT subtypes show 675 subtype-dependent biases in their projections (Chevee et al., 2018). However, a systematic 676 approach to determining the co-variance of two or more properties would require high-throughput 677 measurements of multiple properties simultaneously, a feat that would be challenging using 678 previous approaches. Indeed, it has only recently been possible to obtain such data on even a single 679 property such as gene expression or projection pattern (Kebschull et al., 2016a).

680 The combination of BARseq with *in situ* transcriptional mapping has the potential to 681 systematically determine the co-variation of multiple properties. As a proof-of-principle, we 682 combined BARseq and FISH to explore the projections of transcriptionally defined IT subtypes. 683 Using this approach, we identified a novel projection pattern (ITi-Ctx) restricted almost entirely 684 to two transcriptionally defined IT subtypes (IT3 and IT4). We also identified several other ways 685 in which patterns of gene expression covaried with projection patterns (Fig. 7). However, in 686 marked contrast to the one-to-one relationship between top-level subtypes (IT/CT/PT-1) defined 687 by gene expression and projection pattern, we did not observe a simple mapping of these properties 688 within IT neurons.

689 The mixed projection patterns observed across IT subtypes defined by gene expression are unlikely 690 to have resulted from errors in classifying neurons to subtypes, because we were able to identify 691 both specific and promiscuous projection patterns in the same IT subtypes. This question can be 692 further tested by correlating projections with more marker genes to better define IT subtypes by 693 gene expression. To achieve this, BARseq can potentially be combined with highly-multiplexed 694 FISH (Chen et al., 2015; Codeluppi et al., 2018; Eng et al., 2019; Shah et al., 2016) or in situ 695 sequencing methods (Ke et al., 2013; Lee et al., 2014; Wang et al., 2018) to correlate the expression 696 of dozens to hundreds of genes to projections. Such an approach could further dissect the 697 relationship between projections and gene expression beyond IT subtypes, and has the potential to 698 generate a comprehensive understanding of long-range projections and gene expression at cellular 699 resolution.

700

701 Rosetta Brain

702 Neuronal types differ by a combination of multiple properties (Cadwell et al., 2016; Economo et 703 al., 2018; Paul et al., 2017), including anatomical characteristics such as morphology and 704 connectivity (Economo et al., 2016; Gerfen et al., 2016), molecular characteristics such as gene 705 (Luo et al., 2017; Tasic et al., 2016; Tasic et al., 2018; Zeisel et al., 2015) or protein expression, 706 and functional characteristics such as behaviorally-evoked activity (Cadwell et al., 2016). Defining 707 the biological functions of neuronal classes thus requires a multi-faceted approach that combines 708 measurements of several neuronal properties together at cellular resolution (Huang and Paul, 2019; 709 Zeng and Sanes, 2017). The ability to relate diverse characteristics of many single neurons 710 simultaneously, in a co-registered fashion, within single brains, is an important challenge in 711 neuroscience (Marblestone, 2014). Even simultaneous co-registration of two characteristics can 712 be challenging, but the few successes in this arena have led to insights about the functional 713 organization of neural circuits (Bock et al., 2011; Raj et al., 2018; Sorensen et al., 2015).

714 Co-registering multiple cellular properties at single-cell resolution is crucial for understanding 715 biological systems with heterogeneous cell types, such as the nervous system. By sequencing 716 barcodes in situ, BARseq offers a solution to incorporate high-throughput measurements based on 717 cellular barcoding, such as lineage tracing (Raj et al., 2018), high-throughput screening (Feldman 718 et al., 2018), and projection mapping (Kebschull et al., 2016a) into such integrated approaches. In 719 this study, we focused on combining projections with adult gene expression and the laminar 720 location of neuronal somata. Adapting BARseq for other barcoding techniques and adapting 721 additional *in situ* methods, such as *in vivo* two-photon imaging, to combine with BARseq would 722 yield a "Rosetta Brain"—an integrative dataset that could constrain theoretical efforts to bridge 723 across levels of structure and function in the nervous system (Marblestone, 2014).

725 Acknowledgements

726 The authors would like to acknowledge Kay Tye, Gordon Shephard, and Jessica Tollkuhn for 727 useful discussions, and Barry Burbach, Nour El-Amine, and Stephen Hearn for technical support. 728 This work was supported by the following funding sources: National Institutes of Health 729 [5RO1NS073129 to A.M.Z., 5RO1DA036913 to A.M.Z.]; Brain Research Foundation [BRF-SIA-730 2014-03 to A.M.Z.]; IARPA MICrONS [D16PC0008 to A.M.Z.]; Simons Foundation 731 [382793/SIMONS to A.M.Z.]; Paul Allen Distinguished Investigator Award [to A.M.Z.]; 732 postdoctoral fellowship from the Simons Foundation to X.C. This work was performed with 733 assistance from CSHL Shared Resources, which are funded, in part, by the Cancer Center Support 734 Grant 5P30CA045508.

735

736 Author contributions

X.C. and A.M.Z conceived the study. X.C., H.Z. and Y.S. optimized and performed BARseq. X.C.

and Y.S. performed BARseq in combination with FISH and in Cre labeled mice. J.M.K, and H.Z.

performed MAPseq. X.C. performed single-cell RNAseq. K.M. and Z.J.H generated Cre-labeled

740 mice. X.C. and A.M.Z. analyzed the BARseq and MAPseq data. X.C., S.F., and J. G. analyzed the

single-cell RNAseq data. X.C. and A.M.Z. wrote the paper.

742

743 Competing interests

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

745

746 Materials & Correspondence

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

748

749 Supplementary Figure and Table legends:

750 Fig. S1, related to Fig. 2 and STAR Methods. Optimization of *in situ* barcode sequencing for 751 brain slices. (A) Amplification of barcodes in brain slices in the indicated reaction chambers. 752 Scale bars = $100 \,\mu\text{m}$. (B) Merged images of rolonies (yellow) generated in barcoded brain slices 753 and the residual GFP signals (cyan) with the indicated time of pepsin treatment. Scale bars = 100 754 μm. (C) Comparison of barcode amplicons generated using BaristaSeq (a), the original padlock 755 method (b), and FISSEQ (c). Scale bars = $50 \,\mu$ m. (D) Sequencing images of cycles 2, 4 and 6 of 756 barcoded brain slices sequenced using SOLiD sequencing chemistry (top) and using Illumina 757 sequencing chemistry (bottom). Imaging conditions were kept constant throughout each 758 sequencing run. Scale bars = $100 \,\mu$ m. (E) Average signal-to-noise ratio of Illumina (red) and 759 SOLiD (blue) sequencing *in situ* over cycles. Error bars indicate the standard errors for the SNR 760 for pixels. (F-H) Sequencing quality and signal intensity of individual base calls (F), mean signal 761 intensity over cycles (G), and the fraction of the bases over cycles (H) are plotted. (I) The 762 distribution of barcode counts in contralateral auditory cortex compared to those in the olfactory 763 bulb (negative control). Dashed line indicates threshold above which a projection is considered 764 real.

765 Fig. S2, related to Fig. 3. Comparison of gene expression in barcoded neurons to that in non-766 barcoded neurons. (A) Mean log normalized expression of each gene averaged over all barcoded 767 (x-axis) and non-barcoded (y-axis) neurons. The gene expression is regressed linearly using the 768 percentage of mitochondrial genes. The diagonal line indicates equal expression in barcoded vs. 769 non-barcoded cells. (B) Fraction of variance explained by the same top principal components (x-770 axis) in non-barcoded neurons (blue), barcoded neurons (red), and randomized non-barcoded 771 neurons (yellow). Vertical bars along the non-barcoded line indicate 95% confidence interval. 772 (C) Heat maps of the expression of top genes in the first six principal components from non-773 barcoded neuros.

Fig. S3, related to Fig. 4 and STAR Methods. Locations of neurons in the auditory cortex
mapped by BARseq. (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

778 the workflow of these experiments does not preserve whole brain slices. Scale bar = 500 μ m. 779 (B)(C) Representative images of FISH against Cux2 (B) and Fezf2 (C) in two adjacent slices. Scale 780 bars = 50 μ m. (D) Normalized layer boundaries determined in three pairs of slices across the 781 auditory cortex. In each slice, the thickness of the cortex was normalized to that in the BARseq 782 brains, and the boundary positions were scaled accordingly. (E) Violin plots of the laminar 783 distribution of all BARseq neurons in XC9 and XC28 (All) and those with (Proj) or without (Non-784 proj) detected projections. Individual neurons (red) are plotted on top of smoothed distribution 785 plots.

786 Fig. S4, related to Fig. 5. Projections of the auditory cortex revealed by BARseq. (A) Triple 787 retrograde tracing of neurons projecting to the rostral striatum (CTB-647), the caudal striatum 788 (CTB-488), and the tectum (RetroBeads). (B) A representative image of triple retrograde 789 labeling in the auditory cortex showing neurons projecting to the rostral striatum (magenta), the 790 caudal striatum (cyan), and the tectum (yellow). Scale bar = $100 \,\mu$ m. (C) Venn diagram showing 791 the number of neurons projecting to each of the three areas. (D) The strengths of the indicated 792 projection (y-axes) of individual neurons of the indicated classes are plotted against their laminar 793 position (x-axes). The mean projection strengths are marked by black lines. The layer boundaries 794 are marked by dotted vertical lines.

795

796 Fig. S5, related to Fig. 5 and STAR Methods. Hierarchical clustering of projection neurons in 797 the mouse auditory cortex. (A) Positive (green) or negative (red) Spearman correlation coefficients 798 among projections to the indicated areas in the indicated classes. Only correlations that were 799 statistically significant after Bonferroni correction were shown. (B) The workflow of the 800 hierarchical clustering. (C) Comparison of the original projection strengths (blue) and the filtered 801 projection strengths (red) for two example neurons. (D) The fraction of variance explained (y-axis) 802 using non-negative matrix factorization (NMF; blue), individual projections (red), and PCA 803 (black). (E) The fractions of neurons that remain in the same class-level clusters (y-axis) when 804 filtering the projection data with the indicated number of projection modules (x-axis) compared to 805 the clusters without filtering. (F) Comparison of clusters obtained using k-means (upper row), 806 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 Fig. 5C and the major classes they belong to are labeled below.

813 Fig. S6, related to Fig. 5. Consistent projections and laminar distribution across brains. (A) The 814 fractions of neurons from each brain that belong to each indicated high-level projection subcluster. 815 All bars belonging to a brain sum to 1 across the whole plot. (B) Differences in normalized entropy 816 (x-axis) of individual subclusters between the two brains are plotted against the negative logarithm 817 of the p values (y-axis). The subclusters are color-coded according to their class-level divisions as 818 indicated. The p values are shown without multiple testing correction. The red vertical dashed line 819 indicates no difference in entropy, and the black horizontal dashed line indicates significance level 820 after Bonferroni correction. (C) Differences in mean laminar locations (x-axis) of individual 821 subclusters between the two brains are plotted against the negative logarithm of the p values (y-822 axis). The subclusters are color-coded according to their class-level divisions as indicated. The p 823 values are shown without multiple testing correction. The red vertical dashed line indicates no 824 difference in the mean laminar locations, and the black horizontal dashed line indicates 825 significance level after Bonferroni correction.

826 Fig. S7, related to Fig. 6 and Fig. 7. Projections of IT subtypes defined by gene expression. (A) 827 t-SNE plots of projections color-coded by brain indices. (B) t-SNE plots of projections of neurons 828 of the indicated IT subtypes defined by gene expression. Neurons of the indicated IT subtypes are 829 color-coded by high-level projection subclusters, and other neurons are grayed out. (C) Histograms 830 of log normalized barcode counts of projections to the ipsilateral visual cortex in IT3 and IT4 831 neurons. (D) Cumulative probability distribution of Cdh13 expression in contralateral projection 832 neurons and ipsilateral projection neurons as determined by FISH and CTB retrograde tracing. (E) 833 The minimum distance from a neuron to any neuron within the same projection subcluster (x-axis) 834 or in a different projection subcluster (y-axis). Neurons are color-coded by high-level projection 835 subclusters. Dashed line indicate same minimum distance within and across subtypes. (F) For each 836 neuron belonging to an indicated IT subtype defined by gene expression, the number of IT subtypes

- 837 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
- 839 of transcriptionally defined subtypes a binary projection pattern was found in.

840

842 Table S1, related to Fig. 2. Comparison between BARseq and CTB retrograde tracing. Each row 843 represents a single neuron recovered from BARseq with visible GFP signal from the barcodes and 844 good sequencing quality (quality score > 0.75, column 4). The first four columns indicate the raw 845 barcode counts in the olfactory bulb (OB), contralateral auditory cortex (c1), the cortical area 846 surrounding the contralateral auditory cortex with CTB signals visible to the naked eyes (c2), and an even larger cortical area surrounding the tracer area with CTB signals visible under the 847 848 microscope (c3). The rest of the columns indicate minimal sequencing scores and whether the cell 849 projects contralaterally based on CTB and/or BARseq.

Table S2, related to Fig. 3. Correlation of gene expression and projections using BARseq. Each row represents a single barcoded neuron with a minimum quality score over 0.8. Columns 2-7 indicate barcode molecule counts in each collected area, and column 8 indicates total barcode molecule counts. Columns 9 and 10 indicate whether the cells expressed Slc17a7 and/or Gad2.

854 Table S3, related to Fig. 4. Summary of BARseq datasets. The number of target areas collected, 855 the number of slices sequenced in situ, The number of barcodes sequenced per brain from the 856 projection sites, the number of cells sequenced per brain from the auditory cortex, and the number 857 of *in situ* sequenced barcodes matching barcodes at the projection sites with or without quality 858 filtering are indicated for each brain. *In XC14, the auditory cortex was sequenced in vitro, and 859 no further filtering was applied in addition to those during standard MAPseq data processing. **In 860 experiments combining BARseq and FISH, low-quality neurons were filtered out before mapping 861 to projection pattern. The number of cells in ACx were thus those after filtering. *** In the 862 experiments combining BARseq and Cre labeling, the quality filtering was done similar to the 863 experiments combining BARseq and FISH. The number of cells in ACx listed thus only included 864 high-quality neurons. The number of filtered cells here indicate projection neurons double labeled 865 by the Cre.

867 STAR Methods

868 Viruses, constructs, and oligos

- 869 The plasmid encoding the Sindbis barcode library (JK100L2, <u>https://benchling.com/s/EKtQttOe</u>)
- 870 is available from Addgene (#79785). The RT primer (XC1215), the padlock probe / sequencing
- 871 primer (XC1164), and the fluorescent probe for visualization (XC92) were described previously
- 872 (Chen et al., 2018).
- 873 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
- 876 experiments, we used a separate diverse barcode library with $\sim 10^7$ diversity (Han et al., 2018).
- 877 This library was not fully sequenced *in vitro*.
- 878

879 Animals and tissue processing

880 All animal procedures were carried out in accordance with Institutional Animal Care and Use 881 Committee protocol 16-13-10-07-03-00-4 at Cold Spring Harbor Laboratory. Eight to nine week 882 old male C57BL/6 mice were injected in the left auditory cortex at -4.3 mm ML, -2.6 mm AP from 883 bregma, with 140nL 1:3 diluted Sindbis virus at each of the following depths (200 µm, 400 µm, 884 $600 \,\mu\text{m}$, and $800 \,\mu\text{m}$) at 30° angle. For samples prepared for BaristaSeq only, we transcardially 885 perfused the animal with 10% formalin, then postfixed the tissue for 24 hrs. We then cryo-886 protected the brain in PBS with 10% sucrose for 12 hrs, 20% sucrose PBS for 12 hrs, and 30% 887 sucrose PBS for 12 hrs. We then embedded the brain in OCT (Electron Microscopy Sciences) and 888 cryo-sectioned to 14 µm slices onto SuperFrost Plus slides (VWR).

For BARseq samples, we trans-cardially perfused the animal with PBS 43-45 hrs post-injection. We cut out the left auditory cortex from the brain and post-fixed it in 10% formalin at 4°C for 8 hrs, and snap-froze the rest of the brain on a razor blade on dry ice. The snap-frozen brain (without the injection site) was then processed for conventional MAPseq as described (Kebschull et al.,

893 2016a). The post-fixed auditory cortex was cryo-protected, embedded, and cryo-sectioned as894 described above.

895 For combined BARseq/FISH experiments in Fig. 3D-F, we injected 8-week old C57BL/6 male 896 mouse at -4 mm ML, -2.6 mm AP from bregma, with 140nL 1:3 diluted Sindbis virus at depths 897 200 µm, 400 µm, 600 µm, and 800 µm straight down. After 24 hours, the animal was processed 898 as described above for BARseq experiments. For the experiment in Fig. 6D-E and Fig. 7, 8-week 899 old C57BL/6 male mice were injected at -4.3 mm ML, -2.6 mm AP from bregma with 140nL 1:3 900 diluted Sindbis virus at depths 300 µm, 600 µm, and 900 µm at 30° angle. After 24 hours, the 901 animal was processed as described above for BARseq experiments, except that the injection site 902 was flash frozen in OCT using isopentane and liquid nitrogen immediately after dissection from 903 the brain. These brains were cryo-sectioned to $14 \,\mu m$ sections using a home-made tape-transfer 904 system (Pinskiy et al., 2015) and glued to slides using NOA 81 (Norland) to reduce tissue 905 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 μ m, 400 μ m, 600 μ m, and 800 μ m) at 30° angle. After 48 hrs, we injected 140 nL 1:3 diluted JK100L2 virus into the left auditory cortex at -4.3 mm ML, -2.6 mm AP, with 140nL Sindbis virus at each depth (200 μ m, 400 μ m, 600 μ m, and 800 μ m) at 30° angle. After another 44 hrs, 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 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
for single-cell dissociation as described below.

916

917 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

920 in 70%, 85%, and 100% ethanol for an hour. After rehydration in PBST, we incubated the samples 921 in 0.1M HCl for 5 mins, followed by three PBST washes. We then reverse transcribed the samples 922 [1 U/µl RiboLock RNase inhibitor (Thermo Fisher Scientific), 0.2 µg/µl BSA, 500 µM dNTPs 923 (Thermo Fisher Scientific), 1 µM RT primer, and 20 U/µl RevertAid H Minus M-MuLV reverse 924 transcriptase (Thermo Fisher Scientific) in 1× RT buffer] at 37 °C overnight. After reverse 925 transcription, we crosslinked the cDNAs in 50 mM BS(PEG)₉ for 1 hr and neutralized with 1M 926 Tris-HCl for 30 mins. We then gap-filled and ligated padlock probes [100 nM padlock probe, 50 927 µM dNTPs, 1 U/µl RiboLock RNase inhibitor, 20% formamide (Thermo Fisher Scientific), 0.5 928 U/µl Ampligase (Epicentre), 0.4 U/µl RNase H (Enzymatics), and 0.2 U/µl Phusion DNA 929 polymerase (Thermo Fisher Scientific) in $1 \times$ ampligase buffer supplemented with additional 50 930 mM KCl] for 30 mins at 37 °C and 45 mins at 45 °C. Following PBST washes, we performed 931 rolling circle amplification (RCA) [20 µM aadUTP, 0.2 µg/µl BSA, 250 µM dNTPs, and 1 U/µl 932 ϕ 29 DNA polymerase in 1× ϕ 29 DNA polymerase buffer supplemented with 5 % additional 933 glycerol] overnight at room temperature. After crosslinking the rolonies using BS(PEG)₉ and 934 neutralizing with Tris-HCl, we hybridized 2.5 µM sequencing primers or 0.5 µM fluorescent 935 probes in 2× SSC with 10 % formamide, washed three times in the same buffer, and proceeded to 936 sequencing or imaging.

937 For BaristaSeq on brain tissues, we tested three commercially available reaction chambers that 938 were physically compatible with our samples (Fig. S1Aa-c), and found that the HybriWell-FL 939 sealing system was the only system that did not inhibit rolony formation (Fig. S1Ab). The 940 ImmEdge hydrophobic barrier pen also produced good amplification (Fig. S1Ad), but the 941 HybriWell-FL system offered better control of liquid evaporation during heating steps and easier 942 handling. All slides with brain slices were thus first sealed in HybriWell-FL chambers (22 mm x 943 22 mm x 0.25 mm; Grace Bio-labs) for reactions. The brain slices were washed three times in PBS 944 supplemented with 0.5% Tween-20 (PBST), followed by a pepsin digestion step. This step was 945 necessary to increase accessibility of fixed RNAs (Fig. S1B) and to reduce the GFP signal from the cells (cyan in Fig. S1B), which may interfere with sequencing signals. We found that 3 mins 946 947 of 0.2% pepsin digestion in 0.1 M HCl at room temperature greatly increased rolony formation 948 (Fig. S1Bb) compared to no pepsin treatment (Fig. S1Ba), whereas 5 mins of pepsin digestion 949 caused excessive tissue loss (Fig. S1Bc). We therefore used 3 mins of pepsin digestion for

BaristaSeq in most brain slices, but the optimal timing could vary with different fixation
conditions. We then proceeded with ethanol dehydration, followed by reverse transcription,
padlock gap-filling and ligation, and RCA as described above for cultured neurons.

953 To sequence the barcodes using Illumina sequencing chemistry, we based our sequencing protocol 954 on the HiSeq recipe files (Chen et al., 2018) and reduced the CRM incubation time to two minutes 955 each due to efficient heat transfer of the reaction chambers. We also increased the PBST washes 956 to four to eight times after the IRM reactions to counteract the increased background staining in 957 tissue slices. For slices sectioned using the tape-transfer system (including XC54, XC75, XC91, 958 and XC92), we further incubated the samples in 0.4% MMTS (Pierce) in PBST at 60 °C for 3 mins 959 after CRM and SB3 wash. This optimized Illumina sequencing protocol resulted in an average 960 signal-to-noise ratio 39 ± 4 in the first six sequencing cycles (Fig. S1D, E), ~10-fold higher than 961 that of sequencing by ligation (SOLiD; 4 ± 1) used by other sequencing methods (Ke et al., 2013; 962 Lee et al., 2014).

963

964 Imaging

965 All imaging except for the experiments in Fig. 2C-F and Fig. 3A-C was performed on an UltraView VoX spinning disk confocal microscope (Perkin Elmer) with Volocity 6.3 software as previously 966 967 described (Chen et al., 2018). The sequencing channels were: Channel G, 514 nm laser excitation, 968 405/440/514/640 quad dichroic, 550/49 emission filter, exposure time 500 ms; Channel Y, 561 nm 969 laser excitation, 405/488/561/640 dichroic, 445/60 and 615/70 dual band emission filter, exposure 970 time 120 ms; Channel A, 640 nm laser excitation, 405/488/561/640 dichroic, 676/29 emission 971 filter, exposure time 250 ms; Channel C, 640 nm laser excitation, 405/488/561/640 dichroic, 972 775/140 emission filter, exposure time 500 ms. The experiments in Fig. 2C-F were produced on a 973 Zeiss LSM 710 laser scanning confocal microscope with Zeiss Zen software as previously 974 described (Chen et al., 2018). The experiment in Fig. 3A-C was imaged on a Nikon TE2000 with 975 Crest X-light spinning disk, Photometrics Prime 95B camera, and an 89north 7-line LDI laser. The 976 scope was controlled by micro-manager (Edelstein et al., 2014). The four sequencing channels 977 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).

985 For XC9 and XC28, a 3×1 stitched image at $10\times$ was acquired for each slice, with each tile position 986 including 11 z-positions with 10 μ m step size. For XC75 and XC91, a 3×3 stitched image at 10× 987 was acquired for each slice. Imaging time was about one minute per slice for XC9 and XC28, and 988 three minutes per slice for XC75 and XC91. During each sequencing run, 24-36 slices on three 989 slides were imaged together, resulting in a per-cycle imaging time of 0.5-1 hour. For XC75 and 990 XC91, an additional 5×5 tiles of 20× image was acquired for the first sequencing cycle, with each 991 tile position including 17 z-positions with 3 µm step size. This high-resolution imaging typically 992 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 Fig. S1D, which were shown with only the camera blackpoint
subtracted from the images.

996

997 Base-calling

998 Max projection images were first processed through a median filter and a rolling ball background 999 subtraction. The processed images were then corrected for channel bleed through. The barcodes 1000 were then base-called by picking the channel with the strongest signal. The sequencing quality 1001 score is defined as the intensity of the called channel divided by the root sum square of all four 1002 channels.

1003

1004 Comparison to other *in situ* sequencing techniques

1005 The original padlock probe-based barcode amplification was performed similar to BaristaSeq 1006 amplification as described above, except that the Stoffel fragment (DNA Gdansk) was used in 1007 place of Phusion DNA polymerase and the cDNA crosslink was done using 4% paraformaldehyde 1008 in PBST for 10 mins (Ke et al., 2013).

- 1009 To perform targeted FISSEQ in brain slices, we processed the sample in the same way as in 1010 BaristaSeq in brain slices to the cDNA crosslink step. After cDNA crosslinking, we digested the 1011 RNAs [10 μ l RiboShredder (Epicentre) and 5 μ l RNase H in 1× RNase H buffer] for 1 hour at 37 1012 °C. After washing the samples twice in water, we circularized the cDNAs [0.5 mM DTT, 1M 1013 Betaine, 2.5 mM MnCl₂, and 1 U/µl Circligase II in 1× Circligase buffer] for 1 hour at 60 °C. After 1014 washing the samples with PBST, we hybridized 1.5 μ M RCA primers in 2× SSC with 10% 1015 formamide for 1 hour. We then washed the samples three times with the same buffer, twice more 1016 with PBST, and proceeded to RCA as in BaristaSeq.
- 1017 To compare Illumina sequencing chemistry to SOLiD sequencing chemistry for *in situ* sequencing 1018 in tissues, we performed SOLiD sequencing as described previously (Chen et al., 2018; Lee et al., 1019 2014). To calculate the signal to noise ratio, we first converted the four-channel images into one 1020 single channel, taking the maximum value of the four channels for each pixel. We selected areas 1021 containing barcoded cells or areas containing tissues but no barcoded cells by thresholding as the 1022 signal and background areas. We then subtracted the black point of the camera from both the 1023 signals and the backgrounds and calculated the SNR. The SNR was calculated using the same 1024 selected areas in all six cycles.
- 1025

1026 BARseq

Animals were injected and processed as described above. Cryo-sectioned brain slices were first imaged to generate DIC and GFP images, before they were processed for BaristaSeq. The GFP images from neighboring slices were aligned to each other manually to reduce deformation during sectioning. The sequencing images were then registered back to the GFP images to locate the

positions of the neurons within the slice. Each basecall ROI was thus registered back to the alignedGFP images.

1033 We dissected 12 projection sites from the frozen brains for sequencing. These 12 sites included 1034 the olfactory bulb, the orbitofrontal cortex, the motor cortex, the rostral and caudal striatum, the 1035 somatosensory cortex, the ipsilateral and contralateral visual cortex, the contralateral auditory 1036 cortex, the thalamus, and the tectum. The auditory cortex dissections also included the neighboring 1037 temporal association area, and the visual cortex collected included the posterior parietal cortex. 1038 Slice images, dissected areas, and their correspondence to the Allen Reference Atlas are available 1039 at Mendeley (see Data Availability). The projection sites were sequenced as described for MAPseq 1040 (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.

1044

1045 BARseq data processing

We first filtered the MAPseq generated barcodes so that all barcodes had at least 10 molecules but no more than 10000 molecules at the strongest projection site. We recovered 26840 barcodes using these criteria from the three brains (Table S3). We then matched these barcode sequences at the projection sites to those at the injection site, allowing three mismatches for conventional MAPseq or one mismatch for BaristaSeq. In the conventional MAPseq brain (XC14), 5082 out of 8418 barcodes were confirmed to be from the auditory cortex with more than 20 molecules per barcode at the injection site and were used for the subsequent analyses.

1053 In BARseq experiments, the injection sites were sequenced *in situ* to 15 bases. The 15 bases read 1054 length *in situ* was sufficient to distinguish unambiguously all infected barcodes allowing one 1055 mismatch. For the XC9 brain, barcodes recovered through MAPseq had a mean hamming distance 1056 of 4.5 ± 0.7 (mean \pm stdev; Fig. 5A). Only one pair (0.04%) out of 4841 barcodes had a hamming 1057 distance of 1 and 10 pairs (0.4%) out of 4841 had a hamming distance of 2. Because the sequencing

1058 experiment in Fig. 2 showed only a single error for 51 barcodes, each sequenced 25 bases, our 1059 sequencing error rate was approximately $1/(51 \times 25) = 0.08\%$. Therefore, assuming that 1060 sequencing errors have no bias toward a particular base, the probability of matching an in situ barcode to the wrong MAPseq barcode, while allowing one mismatch, is $\frac{2}{4841} \times 0.08\% \div 3 =$ 1061 1e-7. The probability of an *in situ* barcode matching to two MAPseq barcodes is $\frac{20}{4841} \times$ 1062 $0.08\% \times 2 \div 3 + \frac{2}{4841} = 4e - 4$. Although we cannot detect false positive matches, an 1063 ambiguous match could be detected. In the XC9 data, however, no ambiguous match between the 1064 1065 in situ barcodes and the MAPseq barcodes has occurred.

1066 In addition, XC9 had three pairs of barcodes whose first 15 bases were the same. These appeared 1067 to have arisen from amplification errors in homopolymer stretches of the same barcode rather than 1068 different barcodes, because each pair had a single in-del and had almost identical projection 1069 patterns. These three pairs were not recovered *in situ* and thus did not affect the analyses.

1070 Similarly, out of 13581 total sequences, XC28 had 5 pairs of barcodes within one mismatch and 1071 106 pairs of barcodes within two mismatches for the first 15 bases. No XC28 barcodes had 1072 identical sequences in the first 15 bases. The probability of a wrong match in XC28 is $\frac{10}{13581} \times$ 1073 $0.08\% \div 3 = 2e - 7$, and the probability of an ambiguous match in XC28 is $\frac{212}{13581} \times 0.08\% \times$ 1074 $2 \div 3 + \frac{10}{13581} = 7e - 4$. No actual ambiguous match was seen in XC28. Therefore, allowing one 1075 mismatch for a 15-base sequence was sufficient to match barcodes in the somata to those at the 1076 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 theanalyses.

1086

1087 Comparison of BARseq to retrograde tracer

1088 The animals were injected and processed as described above. We collected four target sites, 1089 including the olfactory bulb (negative control), the contralateral auditory cortex (i.e. the center of 1090 the CTB injection), the remaining areas where CTB was visible to the naked eye, and the 1091 surrounding areas where CTB was visible under a fluorescent microscope. The last three samples 1092 thus formed concentric rings around the CTB injection site. All three samples gave consistent 1093 results regarding contralateral projections (Table S1)

1094 Before library preparation for BaristaSeq, we also imaged the Alexa 647 channel to locate 1095 retrograde-labeled neurons. These images were aligned to the sequencing images. We then only 1096 counted neurons clearly expressing GFP, which was essential to properly judge colocalization with 1097 CTB. To find the fraction of CTB labeled neurons that were also labeled by BARseq, we identified 1098 neurons labeled by both CTB and barcodes with a minimum sequencing quality of 0.75, and 1099 counted the number of neurons with barcodes in the contralateral auditory cortex above the noise 1100 floor. The noise floor was set to be the maximum count of individual barcodes recovered in the 1101 olfactory bulb.

1102

1103 BARseq in Fezf2-2A-CreER::Rosa-CAG-loxP-STOP-loxP-tdTomato (Ai14) 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

1110 flash frozen for dissection of projection areas. The injection site punch was then post-fixed in 4% 1111 PFA for 24 hours, cryo-protected in 10%, 20%, and 30% sucrose in PBS for 12 hours each, 1112 mounted in OCT, and frozen. The punch was then cryo-sectioned to 14 um slices using the tape-1113 transfer system and imaged for GFP, RFP, and DIC channels on a Nikon TE2000 microscope with 1114 a Crest X-light spinning disk confocal. The slices were then processed for BARseq as described 1115 above. Sequencing was done as described above for 14 cycles, except that the imaging was done 1116 on a Nikon TE2000 microscope with a Crest X-light spinning disk confocal. In addition, a set of 1117 sequencing images was taken after each cleavage step. This background image records the residual 1118 tdTomato signals in the sequencing images. After registering and subtracting the background 1119 image from sequencing images, we then median filtered the images, performed background 1120 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 1121 1122 taken before library preparation. We then picked cell bodies from the sequencing images using the 1123 "Find Maxima" function in ImageJ, base-called for barcode sequences, and read out the intensity 1124 from the tdTomato images. We filtered out cells with sequencing quality lower than 0.7. We 1125 considered cells with tdTomato signal over 20000 as Fezf2+ neurons. These cells were visually 1126 inspected to make sure that they had good morphology and were labeled by tdTomato. These cells 1127 were then used for analyses.

1128

1129 Validation of BARseq identified projections using retrograde tracing

1130 To validate the striatal projections of PT-l and IT neurons, we injected red RetroBeads (LumaFluor) 1131 diluted 1:1 in PBS in the superior colliculus at -4.8 mm AP, -0.7 mm ML from bregma at depths 1132 $500 \,\mu\text{m}$, $700 \,\mu\text{m}$, $900 \,\mu\text{m}$, $1100 \,\mu\text{m}$, and $1300 \,\mu\text{m}$ ($70 \,\mu\text{l}$ per depth) from the surface of the brain, 1133 Alexa 488 labeled CTB in the caudal striatum at -1.6 mm AP, -3.2 mm ML at depths 2.5 mm and 1134 3 mm (50 µl per depth) from the surface of the brain, and Alexa 647 labeled CTB in the rostral 1135 striatum at 0.6 mm AP, -2 mm ML at depths 2.5 mm and 3 mm (50 µl per depth) from the surface 1136 of the brain. After 96 hrs, we perfused the animal and sliced the auditory cortex coronally into 70 1137 µm slices. We then imaged the slices on an UltraView VoX spinning disk confocal microscope 1138 (Perkin Elmer).

1139 Validation of combination of BARseq and FISH

1140 To correlate projections to the expression of Slc17a7 and Gad2, we collected the olfactory bulb, 1141 ipsilateral cortical areas (mainly the visual cortex), contralateral cortical areas (the auditory and 1142 visual cortex), the striatum, the thalamus, and the tectum. These areas were processed for MAPseq 1143 as described above and sequenced on an Illumina MiSeq. The injection site was cut out from the 1144 brain, post-fixed and cryo-protected for cryo-sectioning.

1145 We cryo-sectioned the injection site to $14 \,\mu m$ slices using the tape transfer system. The brain slices 1146 were then processed for BARseq to the first cross-link step after reverse transcription. After 1147 neutralization of additional crosslinkers with Tris, we proceeded with ViewRNA ISH (Thermo 1148 Fisher Scientific) using a class 1 probeset for Slc17a7 and a class 6 probeset for Gad2 according 1149 to the manufacturer's protocol. We then imaged each area around the injection site (identified by 1150 the presence of GFP) using a 20×0.75 NA objective on a spinning disk confocal for both FISH 1151 channels (RFP and Cy5), the GFP channel, and the DIC channel. We then stripped away the FISH 1152 probes using 80% formamide for 5 mins twice, followed by three washes in 10% formamide in $2\times$ 1153 SSC, and two washes with PBST. We then proceeded to the padlock ligation step in BARseq and 1154 produced rolonies as described above.

1155 The sequencing of rolonies was done as described above except that the first sequencing cycle was 1156 imaged first using a 20×0.75 NA objective without binning for all four sequencing channels and 1157 the DIC channel. This was followed by imaging using a 10×0.45 NA objective with $2 \times$ binning. 1158 All subsequent sequencing cycles were imaged at the lower resolution to increase throughput.

1159 We then registered the high-resolution FISH images to the high-resolution first cycle sequencing 1160 images using the DIC channel. To match the high-resolution first cycle sequencing images to the 1161 low-resolution sequencing images, we down-sampled the first-cycle sequencing images by four 1162 folds and applied a Gaussian filter to mimic the lower optical resolution. The two images were 1163 then roughly aligned manually. We then extracted cell-body locations using the "Find Maxima" 1164 function in ImageJ, and further aligned the cell-body locations using Iterative Closest Points (Besl 1992) 1165 and McKay, in MATLAB

1166 (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.

1173 Of the 99 barcoded cells with high sequencing quality and good morphology (Fig. 3F; Table S2), 1174 80 were excitatory neurons (Slc17a7+ and Gad2-), 3 were inhibitory neurons (Slc17a7- and 1175 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., 1176 1177 2011), p = 0.13 using Fisher's exact test]. Out of these 99 cells, 54% (54/99) projected to at least 1178 one of the sampled areas with strengths above the noise floor defined by the olfactory bulb (Table 1179 S2). Consistent with the fact that most projection neurons in the cortex are excitatory, all 54 1180 projection neurons identified by BARseq expressed Slc17a7. The non-projecting excitatory 1181 neurons likely projected locally or to nearby cortical areas we did not sample.

1182

1183 Estimation of laminar boundaries using fluorescent *in situ* hybridization (FISH)

1184 To estimate the boundaries of cortical layers, we performed FISH against two known layer-specific 1185 marker genes, Cux2 (Custo Greig et al., 2013) (Fig. S3B), and Fezf2 (Fig. S3C). Cux2 was strongly 1186 expressed in L2/3 and only sporadically in other layers; Fezf2 was strongly expressed in L5 and 1187 weakly in L6. Because L4 is poorly defined in the auditory cortex (Linden and Schreiner, 2003), 1188 we omitted L4 and defined only the remaining two borders. We defined the L2/3 and L5 border as 1189 below the Cux2 band and above the strong Fezf2 band, and defined the L5 and L6 border as 1190 between the strong and weak bands of Fezf2. To account for variation in cortex thickness and 1191 sample preparation, we examined three slices spanning 800 µm in the auditory cortex, normalized 1192 all cortical thickness to 1200 µm (i.e. the same cortical thickness as the BARseq brains), and 1193 calculated the mean positions of layer boundaries (Fig. S3D), The L2/3 and L5 border defined by 1194 Fezf2 agreed with that defined by Cux2. Based on these measurements, we defined the $L^{2/3}$ and

L5 border to be at 590 μm and the L5 and L6 border to be at 830 μm. These borders were used for
the BARseq analyses when layer identities were involved.

We saw few projection neurons in superficial L2/3 in our dataset. This is partially due to smaller number of neurons labeled near the cortical surface, and partially due to an enrichment of neurons without detectable projections in superficial L2/3 (Fig. 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.

1203

1204 Comparison of BARseq bulk projection pattern to bulk GFP tracing

1205 For bulk projection comparison to GFP tracing data, we used the bulk GFP tracing data from five 1206 brains in the Allen connectivity database (Oh et al., 2014) (experiment 116903230, 100149109, 1207 120491896, 112881858, and 146858006; © 2011 Allen Institute for Brain Science. Allen Mouse 1208 Brain Connectivity Atlas. Available from: http://connectivity.brain-map.org/). All five brains had 1209 cells labeled in the primary auditory cortex and no labeling in non-auditory area. Several areas we 1210 collected only corresponded to part of the areas of the same name in the Allen database, including 1211 the somatosensory cortex (restricted to the upper-limb, lower-limb, and the trunk areas), the two 1212 visual cortices (restricted to mostly area pm and am in the Allen Mouse Brain Connectivity Atlas; 1213 these area labels were different from labels in the Allen Reference Atlas), and the striatum (the 1214 rostral striatum and the caudal striatum samples were separated by two brain slices, or $600 \,\mu$ m). 1215 The auditory cortex area we collected also included the temporal association area. For both 1216 BARseq/MAPseq bulk projections and GFP bulk tracing data, we normalized the strengths of 1217 projection to individual areas to total projection strengths in all examined areas for that brain first, 1218 and then averaged across brains (five brains for GFP tracing and three brains for 1219 BARseq/MAPseq). We then calculated the Pearson correlation coefficient and the associated p 1220 value between the GFP tracing and BARseq/MAPseq bulk projection strengths of the 1221 corresponding brain areas.

1222

1223 Population analysis of projections

1224 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.

1236 We then estimated whether each projection pattern could have resulted from missing projections 1237 due to false negatives. For each "target" projection pattern, we identified all "parent" projection 1238 patterns with one more projection. Neurons with the parent projection patterns might have been misidentified as the target projection pattern of interest if a projection was missed. If the ratio of 1239 1240 neurons with the target projection pattern to the ratio of neurons with parent projection patterns 1241 was smaller than the false negative rate, then we eliminated the target projection pattern. We 1242 repeated this process for all projection patterns, and counted the remaining projection patterns. 1243 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; Fig. 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.

1250 To estimate the distribution of the strengths of secondary projections in multi-projection neurons 1251 (i.e. neurons with two or more projections), we filtered out neurons whose primary projections 1252 were too weak to allow more than one barcode molecules in the secondary projections. For 1253 example, when calculating the fraction of neurons with secondary to primary projection ratio of 1254 0.1, we only included neurons whose primary projection was at least 20 barcodes. This would 1255 allow at least 2 barcodes to be observed in the secondary projection. We then separated the neurons 1256 into 100 bins according to their ratio of strengths between secondary and primary projections, and 1257 calculated the probability density function from the cumulative density function. We then fitted 1258 with a log normal distribution using gradient descent.

1259

1260 Hierarchical clustering

1261 All clustering analyses (Fig. S5B) were done using the logarithm of the spike-in-normalized 1262 projection strength. We first filtered the projection data using non-negative matrix factorization 1263 (NMF)(Lee and Seung, 1999) in Matlab. We approximated the projection pattern X of m neurons 1264 to *n* areas as the product of *Y*, an *m* by *k* matrix containing the loadings of the *k* projection modules 1265 for each neuron, and A, a k by n matrix containing the projection pattern of each projection module. 1266 Each of the k projection modules represents a set of projections that correlate with each other. The 1267 projection patterns of individual neurons can thus be approximated by a weighted sum of the 1268 projection modules (Fig. S5C). We chose a k value that captured most of the variance in the data 1269 (Fig. S5D) and resulted in similar classification of neurons in the first two hierarchies (Fig. S5E). This resulted in k = 6. We then reconstructed the filtered projection data X' = Y * A. The filtered 1270 1271 projection data X' was used for clustering.

1272 During each step of the hierarchical clustering (Fig. S5B), we split each node into two groups using 1273 k-means clustering on the squared Euclidean distance of the projection patterns. We then evaluated 1274 whether the split was significant using a Matlab implementation of SigClust (available from 1275 <u>http://www.unc.edu/~marron/marron_software.html</u>). We kept the new clusters if the split was 1276 significant after Bonferroni correction and the sizes of the resulting clusters were larger than 1% 1277 of all data points. This procedure was repeated for each new node until no new clusters were found.

1278 This clustering did not separate all different binary projection patterns into their own clusters, 1279 suggesting that it was conservative. However, it captured high-level differences in projections that 1280 were likely reproducible, and were sufficient for our downstream analyses. Additional levels of 1281 clustering would not likely affect downstream analyses since most of our findings did not rely on 1282 the leaf clusters.

1283 We also compared our clustering to graph-based clustering using Louvain community detection 1284 (Blondel et al., 2008) and spectral clustering (Ng et al., 2002) (Fig. S5F). Louvain community 1285 detection identified 2-4 clusters at each hierarchy, and therefore did not fully correspond to the 1286 clusters obtained by bifurcation only at any hierarchical level. However, the resulting clusters from 1287 both methods, especially high-level nodes, were similar to those obtained using k-means. We 1288 chose to base all further analyses on clustering using k-means, because the major classes were 1289 better separated than using spectral clustering and the imposed bifurcation was easier to interpret 1290 than the clusters produced by Louvain community detection.

1291 We then used a probabilistic approach (Tasic et al., 2016) to assign neurons to these clusters. For 1292 each pair of clusters, we trained a random forest classifier on 80% of the data. We then used the 1293 classifier to classify the remaining 20% of the data. We repeated this process five times, each time 1294 using a mutually exclusive group of 20% of data, so that all data were classified once. We repeated 1295 this whole process 10 times for each pair of clusters, so that all data were classified between each 1296 pair of clusters 10 times. For each barcode, we then removed all cluster memberships that were 1297 scored 0 out of 10 in any one of the pairs involving that cluster. The remaining clusters (i.e. ones 1298 that have scored at least 1 out of 10 comparisons in any pairwise comparison) were assigned to the 1299 barcode, with the main identity as the cluster with the highest sum of scores across all pairwise 1300 comparisons involving that cluster. The majority of neurons (5968/6391, or 93% of all neurons) 1301 were uniquely assigned to a single cluster with high probability (>98% probability), but a small 1302 set of neurons (421/6391, or 7% of all neurons) were assigned to two clusters at around 50% 1303 probability each (Fig. 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

1307 projections, thus simulating the effect of a projection being uninformative (Fig. S5H). This 1308 analysis revealed similar distribution of neurons with the majority being well-classified and a 1309 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 1310 1311 assigned to the wrong cluster using any of the 10 projections if the resulting cluster was consistent 1312 with a dropped projection rather than a false-positive projection. This resulted in 5682/6391 (89%) 1313 well-classified neurons and 709/6391 (11%) ambiguous neurons. These estimates represent an 1314 upper bound on the number of ambiguous neurons, because it did not take into consideration the 1315 actual low false-negative rate of BARseq. Therefore the majority of neurons were unambiguously 1316 assigned to a single cluster even when potentially missed projections were taken into 1317 consideration.

1318 For Louvain community detection, we used a similarity matrix S, in which each element S_{ij} was 1319 the difference between the Euclidean distance of projections between two neurons and the 1320 maximum Euclidean distance of projections between any two neurons in the dataset. Louvain 1321 community detection was performed using a MATLAB implementation of the algorithm 1322 (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 *S* as described above. Spectral clustering was performed using a MATLAB implementation of the algorithm (<u>https://www.mathworks.com/matlabcentral/fileexchange/34412-fast-and-efficient-</u> <u>spectral-clustering</u>).

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 (Fig. S6A), with the exception for the striatal projecting neurons in the CT branch (Leaf 25 in Fig. 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

1335 rather than the tectum. This is especially likely since we did not sample many potential target areas 1336 for PT-l neurons, such as the brain stem. Our analyses of projection neurons *in situ* support the 1337 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 1338 1339 group was obtained in the *in situ* sequenced brains, compared to 229 of the former group (Fig. 1340 S6A). Therefore, most of the striatum and thalamus projecting neurons were from the conventional 1341 MAPseq brain (XC14). Because we only sequenced neurons at the center of each injection site in 1342 situ (XC9 and XC28), but collected a much larger injection site that may have included 1343 neighboring cortical areas in the conventional MAPseq brain (XC14), these neurons were rare in 1344 the auditory cortex, and were more likely in neighboring cortical areas, where PT-l neurons could 1345 project to other targets. Furthermore, the 10 neurons that projected to the striatum had a laminar 1346 profile similar to that of PT-l neurons, but different from those of the layer 6 CT neurons (Fig. 5C). These results suggest that most of the striatum projecting "corticothalamic" neurons were likely 1347 1348 PT-l neurons in layer 5 and deep layer 6 in neighboring cortical areas. The lack of such neurons in 1349 XC9 and XC28 thus likely reflect differences in the locations of neurons sampled due to the 1350 differences between MAPseq and BARseq.

1351

1352 Normalized entropy for the laminar distribution of a group of neurons

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

1359 To calculate the entropy of the laminar distribution of a group of neurons, we discretized the 1360 laminar location of the neurons into 13 bins, each covering 100 µm. We then calculated the entropy 1361 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 1362 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 $E' = -E/\log_{10}(1/13)$. The normalized entropy thus equals 0 when all neurons fall into one bin, and 1 when the neurons randomly distribute across all 13 bins.

1365 We did not see significant difference in the entropy of laminar distribution of subclusters across 1366 brains (Fig. S6B; p > 0.05 for all subclusters after Bonferroni correction). Furthermore, the 1367 distribution of neurons was also consistent across brains: the average laminar locations of neurons 1368 from each brain were similar for all but one subcluster (Fig. S6C; p > 0.05 after Bonferroni 1369 correction, Mann-Whitney U test). The one exception was CT neurons, which were on average 50 1370 μ m deeper in XC28 than in XC9 (Fig. S6C; p < 0.0005 after Bonferroni correction), probably 1371 because more deep L6 neurons were labeled in XC28 than in XC9 (XC9 and XC28 had 165 and 1372 132 L6 labeled neurons with laminar depths $< 1000 \,\mu$ m, respectively, compared to 56 and 129 L6 1373 neurons with laminar depths > 1000 μ m; p < 1e-7 using Fisher's exact test). Therefore, this 1374 observed lack of laminar specificity of subclusters was consistent across samples.

1375

1376 Single-cell dissociation for RNAseq

To dissociate neurons for single-cell RNAseq, the animal was anesthetized with isofluorane and 1377 1378 decapitated. We then dissected out the brain and used a 2 mm biopsy punch to remove the auditory 1379 cortex. The auditory cortex was immediately dissected in ice cold HABG medium [40 mL 1380 Hibernate A (Brainbits), 0.8 mL B27 (Thermo Fisher Scientific) and 0.1 mL Glutamax (Thermo 1381 Fisher Scientific)] into small pieces and placed in 3mL papain solution [3mL Hibernate A-Ca 1382 (Brainbits), 6 mg papain (Brainbits), and 7.5 µL Glutamax] pre-warmed to 30 °C in a 15 mL tube. 1383 The tube was closed and gently rocked at 30 °C for 40 mins. The digested tissues were then 1384 transferred to a tube containing 2 mL HABG pre-warmed to 30 °C and triturated 10 times using a 1385 salinized pipette with 500 µm opening. The undissociated tissues were then transferred to a new 1386 tube with 2 mL HABG and triturated 10 times using a salinized pipette with 500 µm opening. The 1387 remaining undissociated tissues were again transferred to a new tube with 2mL HABG and 1388 triturated 5 times. The three tubes of 2 mL HABG each were then combined, and carefully laid on 1389 top of a density gradient of 17.3%, 12.4%, 9.9%, and 7.4% (v/v) Optiprep in HABG. This tube 1390 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 5mL HABG and centrifuged at 300 g for 5 mins. The pellet was washed in 5 mL HABG, pelleted again, and resuspended in 100 μ L HABG. The cell suspension was then kept on ice and processed for library preparation using 10x Genomics Chromium according to the manufacturer's protocol. The prepared libraries were sequenced using Illumina NextSeq.

1397

1398 Comparison of barcoded and non-barcoded neurons using single-cell RNAseq

1399 We injected the brain in the auditory cortex as described above. After 24 hours, we dissociated 1400 neurons from the auditory cortex as described above, and prepared library using Chromium single 1401 cell 3' reagent kit v2 (10x Genomics). The sequencing library was then sequenced using one lane of Illumina NextSeq. The raw sequencing data were processed using Cell Ranger v2 (10x 1402 1403 Genomics) to generate the expression matrix. We then further processed the data using Seurat v2. 1404 We first filtered out cells with less than 200 unique genes per cell and more than 10% mitochondria 1405 RNA per cell) and log normalized the data. We then regressed out either the effect of the 1406 percentage of mitochondria genes linearly (Fig. S2A) or both the effect of the percentage of 1407 mitochondria genes and endogenous UMI counts (i.e. excluding barcode reads) using a poisson 1408 model (Fig. 3G). We then identified barcoded neurons and non-barcoded neurons by thresholding 1409 the level of Snap25 expression and barcode counts. We observed a general reduction in gene 1410 expression across most genes, but the relative level of expression among these genes were 1411 preserved in barcoded cells (Fig. S2A). Although a minor fraction of genes were over-expressed 1412 in barcoded cells, we could further correct for both the under-expression among most genes and 1413 over-expression in this subset of genes in the barcoded neurons by simply regressing out the effect 1414 of endogenous read counts per cell (Fig. 3G), thus restoring the relative gene expression level to 1415 the level of the non-barcoded cells. These results indicate that the relative gene expression level is 1416 preserved in barcoded neurons.

1417 To further confirm that the relationship among expression of genes were preserved in individual 1418 cells, we examined the amount of variance explained by principal components in neurons (Fig.

1419 S2B). We reasoned that if Sindbis infection perturbs major structures in gene expression, then 1420 principal components of gene expression in non-barcoded neurons should fail to capture the 1421 variance of gene expression in barcoded neurons. We identified genes with variable expression 1422 and exported the expression of these genes for both barcoded and non-barcoded cells to MATLAB. 1423 We then scaled and centered the data in MATLAB and calculated the explained variance of top 1424 principal components. Our results (Fig. S2B) showed that the same top principal components 1425 explained similar amount of variance in both barcoded and non-barcoded neurons, thus further 1426 indicating that the structures in gene expression is preserved in barcoded neurons. Heatmaps of 1427 top genes of principal components (Fig. S2C) were generated in Seurat.

1428

1429 Single-cell RNAseq of neurons in the auditory cortex

1430 For non-barcoded cells in the auditory cortex, the library was prepared using Chromium single cell 1431 3' reagent kit v3 (10x Genomics) and sequenced using one lane of Illumina NextSeq. The raw data 1432 was processed using Cell Ranger v3 (10x Genomics) followed by Seurat v3 (Stuart et al., 2019). 1433 We analyzed 4471 cells with more than 3000 UMI, more than 800 unique RNAs, and less than 1434 15% mitochondria RNA counts. We then linearly regressed out the effect of the percentage of 1435 mitochondria RNA counts, and clustered using the top 40 principal components using a resolution 1436 of 1 in Seurat. This produced seven clusters (735 cells) that appeared to be neurons with high 1437 Snap25 expression. We then pooled these neuronal clusters together and performed a second round 1438 of clustering using the top 40 principal components of this subset of data using a resolution of 0.3. 1439 This produced the seven neuronal clusters shown in Fig. 6A, B.

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 et al. (2018) has been merged into a single label "L4/L5 IT". This analysis revealed that the "CT" cluster in the auditory cortex 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 segregatethese neuronal types by clustering, nor did we attempt to enrich for PT-l neurons.

1449

1450 Correlating projections to transcriptionally defined IT subtypes using BARseq

1451 The four IT subtypes defined by gene expression are largely segregated by laminae (Tasic et al.,

1452 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

1454 layer 6 contains both IT3 and IT4 neurons, laminar position alone was not sufficient to

1455 distinguish between them. We thus utilized the differential expression of two marker genes,

1456 Cdh13 and Slc30a3, to further distinguish these IT subtypes. Cdh13 is highly expressed in IT3,

1457 but not IT2 neurons, whereas Slc30a3 is expressed in IT3, but not IT4 neurons.

1458 Based on the gene expression (Fig. 6A) and the laminar distribution of the four IT subtypes, we

1459 could distinguish the four IT subtypes using the laminar position of neurons and the expression

1460 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

1462 with Slc30a3; IT4 neurons are layer 6 IT neurons without Slc30a3. We further included Slc17a7,

1463 a highly expressed excitatory neuron marker in the cortex, as a control for both cell morphology

and FISH quality.

1465 We collected the same 12 projection targets as in the BARseq-only experiments, but each

1466 cortical site was further dissected into superficial and deep layers. This resulted in a total of 18

1467 projection sites. The projection sites were processed as in the BARseq-only experiments and

1468 sequenced on a Illumina MiSeq.

1469 The injection site (auditory cortex) was cryo-sectioned to 14 um slices using the home-made tape

1470 transfer system. The slides were then fixed and permeated according to RNAscope (ACD Bio)

1471 pretreatment procedures for fresh-frozen tissues (cold 4% PFA fixation for 15 mins, followed by

1472 dehydration in ethanol and 30 mins of Protease IV treatment at room temperature). We then

1473 washed the slide three times in PBST, deactivated residual protease by incubating twice in 40%

1474 formamide, 2× SSC, 0.01% Triton-X100 for 10 mins at 60 °C, then washed an additional three

1475 times in PBST. We then performed reverse transcription, crosslinking, and neutralization of

1476 crosslinkers as in BARseq, except that the reverse transcription was performed for only two

- 1477 hours at 37 °C. We then followed procedures for RNAscope according to the manufacturer's
- 1478 protocol.

1479 The FISH signals produced by RNAscope were then imaged on a spinning disk confocal using a

1480 20×0.75 NA Plan-Apo objective with $2 \times$ binning. After imaging, the FISH probes were stripped

1481 twice in 40% formamide, 2× SSC, 0.01% Triton-X100 for 10 mins at 60 °C. We then continued

1482 with BARseq from padlock ligation as described above. Sequencing of barcodes was performed

1483 as in BARseq-only experiments, except that the first sequencing cycle was imaged with a $10\times$

1405 as in DARseq-only experiments, except that the first sequencing eyere was imaged with a 107

1484 0.45 NA Plan-Apo objective and a 20×0.75 NA Plan-Apo objective, both with $2 \times$ binning.

1485 The FISH images were then registered to the first cycle high-resolution sequencing images using 1486 the DIC channel using intensity-based registration in MATLAB. The high-resolution first-cycle 1487 sequencing images were registered to the low-resolution first-cycle sequencing images using 1488 intensity-based registration. We then extracted cell-body locations using the "Find Maxima" 1489 function in ImageJ, and base-called the barcoded somata from the low-res images. We then filtered 1490 out cells with low sequencing quality (below 0.8). We then manually counted FISH reads for 1491 excitatory neurons with high FISH quality. Excitatory neurons with high FISH quality is defined 1492 by: (1) The neuronal somata were clearly delineated by barcodes; (2) the Slc17a7 reads clearly 1493 defined the same soma shape; (3) the neuron did not overlap significantly with neighboring 1494 neurons. Cells with fewer than 10 copies of Slc17a7 per cell were further excluded (these might 1495 have been cells with low FISH quality, partial cells that were cut during cryo-sectioning, and non-1496 excitatory or non-neuronal cells). The remaining cells, including 781 excitatory neurons from 1497 XC75 and 737 excitatory neurons from XC91, were mapped to barcodes at projection sites for 1498 further analyses.

1499

1500 Analysis of projections of IT subtypes defined by gene expression

1501 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 data set 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 projections specifically to the superficial layers, and a bias of 0 indicate projections to the deep layers only.

1513 To quantify whether ipsilateral and contralateral projections were more likely to co-occur or were 1514 mutually exclusive, we compared the fraction of co-projecting neurons to the fraction of either 1515 ipsilateral-only or contralateral-only projection neurons between pairs of IT subtypes defined by 1516 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 1517 1518 likely to co-project contralaterally than both IT1 (p = 8e-23) and IT2 neurons (p = 4e-17). Among 1519 contralaterally projecting neurons, IT3 was more likely to exclude ipsilateral projections than IT1 1520 neurons (p = 0.02). No other significant bias was found among the contralaterally projecting 1521 neurons (p = 0.3 comparing IT1 to IT4; p = 1 comparing IT2 to either IT3 or IT4). All statistical 1522 significance was obtained using Fisher's exact test after Bonferroni correction. Therefore, IT3 and 1523 IT4 neurons that project ipsilaterally were more likely to exclude contralateral projections than 1524 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 (Fig. 7G). As a control the same analysis was done on IT neurons within or across high-level projection subclusters (Fig. S7E). As expected, most neurons were more similar to neurons within the same projection subcluster compared to neurons in different projection subclusters.

1530

1531 Retrograde tracing validation of correlation between Cdh13 and contralateral projections

1532 12-week old C57BL/6 animals were injected in the somatosensory cortex with 150 nL CTB-Alexa 1533 488 at 0.6 mm AP, -2.2 mm ML from begma, at 300 and 600 um depths, and in the contralateral 1534 auditory cortex with 150 nL CTB-Alexa 647 at 2.3 mm AP, 4.2 mm ML, at 500 and 700 um depths. 1535 After 96 hours, the animal was perfused with fresh 4% PFA, post-fixed for 24 hours, and cut to 70 1536 µm coronal sections using a vibratome. Four slices containing the ipsilateral auditory cortex were 1537 mounted and imaged for both the green and far red channel on an Ultraview spinning-disk confocal 1538 (Perkin Elmer). We then probed for Cdh13 in these slices using RNAscope Fluorescent Multiplex 1539 Assay following the manufacturer's instructions. The Cdh13 probe was visualized with the red 1540 Amp4 in the RNAscope kit. We then imaged the green channel, far red channel, and the Cdh13 1541 channel (red), and registered the images to the green/far red channel images taken before 1542 RNAscope using those two channels. We then manually counted the Cdh13 expression in 1543 ipsilateral projecting or contralateral projecting neurons in layer 6 (defined by a normalized 1544 cortical depth of over 830 µm). Rank sum test was used to calculate statistical significance between 1545 the distribution of Cdh13 in ipsilaterally projecting neurons and contralaterally projecting neurons.

1546

1547 **Quantification and statistical analysis**

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.

1551

1552 Data availability

1553 All *in vitro* high throughput sequencing datasets (including two BARseq brains, XC9 and XC28,

1554 one MAPseq brain, XC14, one brain for the comparison between BARseq and retrograde tracing,

1555 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 deposited to SRA (PRJNA557267 and PRJNA448728). *In situ* sequencing data, images for the MAPseq dissections, and a list of Allen
- 1558Reference Atlas slices corresponding to the dissection slices will be available from Mendeley Data
- 1559 (doi:10.17632/mk82s9x82t.1, 10.17632/g7kdxznt6w.1, 10.17632/86wf7xfz5x.1,
- 1560 10.17632/2w649fccnt.1). The same data are provided at
- 1561 <u>https://www.dropbox.com/sh/xz7kwj5aj9yxmee/AAD_4d8DppGA6heaO0EVytua?dl=0</u> for
- review purposes.
- 1563

1564 **Code availability**

Analysis code and data needed to reproduce all analyses are available for review purposes at:
<u>https://www.dropbox.com/sh/xz7kwj5aj9yxmee/AAD_4d8DppGA6heaO0EVytua?dl=0</u>. The
same code and data will be available from Mendeley Data (doi: 10.17632/mk82s9x82t.1,
10.17632/g7kdxznt6w.1, 10.17632/86wf7xfz5x.1, 10.17632/2w649fccnt.1) upon publication.

1570 **References**

- Besl, P., and McKay, N. (1992). A Method for Registration of 3-D Shapes. IEEE Trans PatternAnal Mach Intell *14*, 239-256.
- 1573 Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of
- 1574 communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008,1575 P10008.
- 1576 Bock, D.D., Lee, W.C., Kerlin, A.M., Andermann, M.L., Hood, G., Wetzel, A.W., Yurgenson,
- 1577 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.
- 1579 Cadwell, C.R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S.,
- 1580 Bethge, M., Tolias, K.F., *et al.* (2016). Electrophysiological, transcriptomic and morphologic
- profiling of single neurons using Patch-seq. Nat Biotechnol *34*, 199-203.
- Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). RNA imaging.
 Spatially resolved, highly multiplexed RNA profiling in single cells. Science *348*, aaa6090.
- 1584 Chen, X., Sun, Y.C., Church, G.M., Lee, J.H., and Zador, A.M. (2018). Efficient in situ barcode 1585 sequencing using padlock probe-based BaristaSeq. Nucleic Acids Res *46*, e22.
- 1586 Chevee, 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.
- 1589 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.
- 1592 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. NatCommun 9, 884.
- 1595 Custo Greig, L.F., Woodworth, M.B., Galazo, M.J., Padmanabhan, H., and Macklis, J.D. (2013).
- Molecular logic of neocortical projection neuron specification, development and diversity. Nat
 Rev Neurosci 14, 755-769.
- 1598 Economo, M.N., Clack, N.G., Lavis, L.D., Gerfen, C.R., Svoboda, K., Myers, E.W., and
- 1599 Chandrashekar, J. (2016). A platform for brain-wide imaging and reconstruction of individual 1600 neurons. Elife *5*, e10566.
- 1601 Economo, M.N., Viswanathan, S., Tasic, B., Bas, E., Winnubst, J., Menon, V., Graybuck, L.T.,
- 1602 Nguyen, T.N., Smith, K.A., Yao, Z., *et al.* (2018). Distinct descending motor cortex pathways 1603 and their roles in movement. Nature *563*, 79-84.
- Edelstein, A.D., Tsuchida, M.A., Amodaj, N., Pinkard, H., Vale, R.D., and Stuurman, N. (2014).
 Advanced methods of microscope control using muManager software. J Biol Methods *1*.
- 1606 Eng, C.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C.,
- 1607 Yuan, G.C., *et al.* (2019). Transcriptome-scale super-resolved imaging in tissues by RNA 1608 seqFISH. Nature *568*, 235-239.

- 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.
- 1611 Gerfen, C.R., Economo, M.N., and Chandrashekar, J. (2016). Long distance projections of
- 1612 cortical pyramidal neurons. J Neurosci Res.
- 1613 Gong, H., Xu, D., Yuan, J., Li, X., Guo, C., Peng, J., Li, Y., Schwarz, L.A., Li, A., Hu, B., et al.
- 1614 (2016). High-throughput dual-colour precision imaging for brain-wide connectome with
- 1615 cytoarchitectonic landmarks at the cellular level. Nat Commun 7, 12142.
- 1616 Guo, C., Peng, J., Zhang, Y., Li, A., Li, Y., Yuan, J., Xu, X., Ren, M., Gong, H., and Chen, S.
- 1617 (2017). Single-axon level morphological analysis of corticofugal projection neurons in mouse
- 1618 barrel field. Sci Rep 7, 2846.
- 1619 Han, Y., Kebschull, J.M., Campbell, R.A.A., Cowan, D., Imhof, F., Zador, A.M., and Mrsic-
- 1620 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. NatNeurosci 18, 170-181.
- 1623 Huang, L., Kebschull, J.M., Furth, D., Musall, S., Kaufman, M.T., Churchland, A.K., and Zador,
- 1624 A.M. (2018). High-throughput mapping of mesoscale connectomes in individual mice. Biorxiv.
- Huang, Z.J., and Paul, A. (2019). The diversity of GABAergic neurons and neuralcommunication elements. Nat Rev Neurosci.
- 1627 Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wahlby, C., and Nilsson, M.
- (2013). In situ sequencing for RNA analysis in preserved tissue and cells. Nat Methods *10*, 857-860.
- 1630 Kebschull, J.M., Garcia da Silva, P., Reid, A.P., Peikon, I.D., Albeanu, D.F., and Zador, A.M.
- 1631 (2016a). High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded
 1632 RNA. Neuron *91*, 975-987.
- 1633 Kebschull, J.M., Garcia da Silva, P., and Zador, A.M. (2016b). A New Defective Helper RNA to
- 1634 Produce Recombinant Sindbis Virus that Infects Neurons but does not Propagate. Front 1635 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.
- Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrixfactorization. Nature *401*, 788-791.
- 1640 Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R.,
- 1641 Jeanty, S.S., Li, C., Amamoto, R., *et al.* (2014). Highly multiplexed subcellular RNA sequencing 1642 in situ. Science *343*, 1360-1363.
- 1643 Lin, H.M., Kuang, J.X., Sun, P., Li, N., Lv, X., and Zhang, Y.H. (2018). Reconstruction of
- 1644 Intratelencephalic Neurons in the Mouse Secondary Motor Cortex Reveals the Diverse 1645 Projection Patterns of Single Neurons. Front Neuroanat *12*, 86.
- 1646 Linden, J.F., and Schreiner, C.E. (2003). Columnar transformations in auditory cortex? A
- 1647 comparison to visual and somatosensory cortices. Cereb Cortex 13, 83-89.

- 1648 Luo, C., Keown, C.L., Kurihara, L., Zhou, J., He, Y., Li, J., Castanon, R., Lucero, J., Nery, J.R.,
- 1649 Sandoval, J.P., *et al.* (2017). Single-cell methylomes identify neuronal subtypes and regulatory 1650 elements in mammalian cortex. Science *357*, 600-604.
- 1651 Marblestone, A.H., Daugharthy, E.R., Kalhor, R., Peikon, I.D., Kebschull, J.M., Shipman, S.L.,
- 1652 Mishchenko, Y., Lee, J.H., Kording, K.P., Boyden, E.S., Zador, A.M., Church, G.M. (2014).
- 1653 Rosetta Brains: A Strategy for Molecularly-Annotated Connectomics. *arXiv*.
- 1654 Meyer, H.S., Schwarz, D., Wimmer, V.C., Schmitt, A.C., Kerr, J.N., Sakmann, B., and
- 1655 Helmstaedter, M. (2011). Inhibitory interneurons in a cortical column form hot zones of
- 1656 inhibition in layers 2 and 5A. Proc Natl Acad Sci U S A *108*, 16807-16812.
- Ng, A.Y., Jordan, M.I., and Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm.
 Adv Neur In *14*, 849-856.
- 1659 Oh, S.W., Harris, J.A., Ng, L., Winslow, B., Cain, N., Mihalas, S., Wang, Q., Lau, C., Kuan, L.,
- 1660 Henry, A.M., *et al.* (2014). A mesoscale connectome of the mouse brain. Nature 508, 207-214.
- 1661 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-
- 1663 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*, 522539 e520.
- Pinskiy, V., Jones, J., Tolpygo, A.S., Franciotti, N., Weber, K., and Mitra, P.P. (2015). HighThroughput Method of Whole-Brain Sectioning, Using the Tape-Transfer Technique. PLoS One
- 1669 *10*, e0102363.
- 1670 Raj, B., Wagner, D.E., McKenna, A., Pandey, S., Klein, A.M., Shendure, J., Gagnon, J.A., and
- 1671 Schier, A.F. (2018). Simultaneous single-cell profiling of lineages and cell types in the vertebrate 1672 brain. Nat Biotechnol *36*, 442-450.
- 1673 Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R., Welch,
- 1674 J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: A scalable technology for
- 1675 measuring genome-wide expression at high spatial resolution. Science 363, 1463-1467.
- 1676 Shah, S., Lubeck, E., Zhou, W., and Cai, L. (2016). In Situ Transcription Profiling of Single
- 1677 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.
- 1682 Sorensen, S.A., Bernard, A., Menon, V., Royall, J.J., Glattfelder, K.J., Desta, T., Hirokawa, K.,
- 1683 Mortrud, M., Miller, J.A., Zeng, H., et al. (2015). Correlated gene expression and target
- 1684 specificity demonstrate excitatory projection neuron diversity. Cereb Cortex 25, 433-449.
- 1685 Stahl, P.L., Salmen, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello,
- 1686 S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis of gene
- 1687 expression in tissue sections by spatial transcriptomics. Science *353*, 78-82.

- 1688 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 e1821.
- 1691 Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T.,
- 1692 Sorensen, S.A., Dolbeare, T., *et al.* (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci *19*, 335-346.
- 1694 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.
- 1699 Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K.,
- 1700 Liu, C., Ramakrishnan, C., Liu, J., *et al.* (2018). Three-dimensional intact-tissue sequencing of 1701 single-cell transcriptional states. Science.
- 1702 Zeisel, A., Hochgerner, H., Lonnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Haring,
- 1703 M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse
- 1704 Nervous System. Cell 174, 999-1014 e1022.
- 1705 Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A.,
- 1706 Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the
- 1707 mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138-1142.
- Zeng, H., and Sanes, J.R. (2017). Neuronal cell-type classification: challenges, opportunities and
 the path forward. Nat Rev Neurosci *18*, 530-546.
- 1710 Zheng, T., Yang, Z., Li, A., Lv, X., Zhou, Z., Wang, X., Qi, X., Li, S., Luo, Q., Gong, H., et al.
- 1711 (2013). Visualization of brain circuits using two-photon fluorescence micro-optical sectioning
- 1712 tomography. Opt Express 21, 9839-9850.
- 1713 Znamenskiy, P., and Zador, A.M. (2013). Corticostriatal neurons in auditory cortex drive
- 1714 decisions during auditory discrimination. Nature *497*, 482-485.
- 1715