bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Prospective, brain-wide labeling of neuronal subclasses with enhancer-driven AAVs

One sentence Summary:

Combining scATAC-seq and scRNA-seq identifies subclass-specific enhancers to label cells in mouse cortex at high resolution.

Authors:

Lucas T Graybuck, Adriana E Sedeño-Cortés, Thuc Nghi Nguyen, Miranda Walker, Eric Szelenyi, Garreck Lenz, La'Akea Sieverts, Tae Kyung Kim, Emma Garren, Brian Kalmbach, Shenqin Yao, Marty Mortrud, John Mich, Jeff Goldy, Kimberly Smith, Nick Dee, Zizhen Yao, Ali Cetin, Boaz Levi, Ed Lein, Jonathan Ting, Hongkui Zeng, Tanya Daigle, and Bosiljka Tasic

Author Contributions:
B.T., L.T.G., and T.D. designed the study; J.T., B.L,. and J.M. provided viral genome constructs and cloning protocols; G.L. performed cloning experiments. A.S.-C., T.N., and L.T.G. performed scATAC-seq experiments. B.K. performed electrophysiology experiments. A.C., S.Y., and M.M. performed viral packaging and purification. T.N. and M.M. performed retrograde injections. T.N. and E.S. performed stereotaxic injections. M.W. performed sectioning and IHC experiments. T.N. and E.G. performed additional validation experiments. N.D. managed tissue processing for RNA-seq experiments. K.S. managed RNA-seq experiments. Z.Y. and L.T.G. performed RNA-seq analysis. L.T.G. and A.S.-C. performed scATAC-seq analysis. H.Z. and E.L. lead the Cell Types Program at the Allen Institute. L.T.G. and B.T. wrote the manuscript, with input from all coauthors.

Abstract

Labeling and perturbation of specific cell types in multicellular systems has transformed our ability to understand them. The rapid pace of cell type identification by new single-cell analysis methods has not been met with efficient access to these newly discovered types. To enable access to specific neural populations in the mouse cortex, we have collected single cell chromatin accessibility data from select cell types. We clustered the single cell data and mapped them to single cell transcriptomics to identify highly specific enhancers for cell subclasses. These enhancers, when cloned into AAVs and delivered to the brain by retro-orbital injections, enable transgene expression in specific cell subclasses throughout the mouse brain. This approach will enable functional investigation of cell types in the mouse cortex and beyond.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

The functional interplay of neural cell types gives rise to the complex functions of neural tissues. To fully understand the biology of the brain, we need to be able to distinguish and describe these cell types, and identify markers that can be used to selectively label and perturb them for further study[1,2]. In mouse, recombinase driver lines have been used to great effect to label cell populations that share marker gene expression[3–5]. However, the creation, maintenance, and use of lines that label cell types with high specificity can be costly, frequently requiring triple transgenic crosses, which yield low frequency of experimental animals.

Recent advances in single-cell profiling, such as single-cell RNA-seq[6,7] and surveys of neuronal electrophysiology and morphology[8], have revealed that many recombinant driver lines label many different cell types, often from multiple subclasses. For example, the *Rbp4-Cre* mouse driver line, which labels layer 5 (L5) glutamatergic neurons, labels transcriptomically and connectionally distinct subclasses: L5 intratelencephalic (L5 IT) and pyramidal tract (L5 PT) neurons[9].

Here, we demonstrate that a combination of single-cell RNA-seq (scRNA-seq) and single-cell ATAC-seq (scATAC-seq) can be used to identify functional and specific enhancer elements. These elements can be combined with recombinant adeno-associated viruses (AAVs) to generate subclass-specific or cell type-specific viral labeling tools. These tools can be delivered using minimally invasive retro-orbital injections to consistently label or perturb neural cell subclasses in mice with specificity that surpasses existing driver lines (Figure 1).

Results

We isolated individual neuronal and non-neuronal cells from transgenically-labeled mouse cortex by FACS (Supp Fig 1) and examined them using the Assay for Transposase-Accessible Chromatin with next generation sequencing (scATAC-seq; **Fig 1**) [10,11]. This strategy allowed us to interrogate both abundant (e.g. L4 IT neurons, ~17% of VISp neurons) and very rare cell types (e.g. Sst Chodl neurons, ~0.1% of VISp neurons) with the same method. To sample cells both broadly and specifically in the mouse brain, we utilized 25 different Cre or Flp-driver lines or their combinations crossed to appropriate reporter lines (**Supp Fig. 2**, **Supp Table 1**). We characterized many of the same lines previously by single-cell RNA-seq[6]. In addition, we employed retrograde labeling by recombinase-expressing viruses to selectively sample cells with specific projections (Retro-ATAC-seq; **Supp Fig 3**, **Supp Table 2,3**). Our method yielded scATAC-seq libraries of comparable quality to previously published scATAC-seq studies (**Supp Fig 4**).

To generate scATAC-seq directly comparable with our scRNA-seq dataset (Tasic 2018), we focused our dissections on visual cortex for glutamatergic cell types, but allowed broader cortical sampling for GABAergic cell types. This strategy is rooted in our observation that GABAergic cell types are shared across two distant poles of mouse cortex, wherease the glutamatergic cell types are distinct among different cortical regions[6] Retro-ATAC-sSeq cells

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

were collected only from the visual cortex. In total, we collected 3,381 single cells from 25 driver-reporter combinations in 60 mice, 126 retrogradely labeled cells from injections into 3 targets across 7 donors, and 96 samples labeled in by 1 retro-orbital injection from of a viral tool generated in this study (**Supp Table 1**). After FACS, individual cells were processed using ATAC-seq, and were sequenced in 60-96 sample batches using a MiSeq (Methods). We performed quality control filtering to select 2,416 samples with >10,000 uniquely mapped paired-end fragments, >10% of which had a fragment size longer than 250 bp, and with >25% of fragments overlapping high-depth cortical DNAse-seq peaks generated by ENCODE [12] (**Fig 2a**, **Supp Fig 2**, **Supp Table 4**) .

Previous studies have shown that most recombinase driver lines label more than one transcriptomic cell type[6,7]. To increase the cell-type resolution of chromatin accessibility profiles beyond that provided by driver lines, we clustered the scATAC-seq data using a novel, feature-free method for computation of pairwise Jaccard distances (Figure 2b). These distances were used for principal component analysis (PCA) and t-stochastic neighbor embedding (t-SNE), followed by Phenograph clustering[13] (**Fig 2c**). This clustering method clearly grouped cells from class-specific driver lines together, and segregated them into multiple clusters as expected based on transcriptomic analyses (**Fig 2d**). Cluster identity was then assigned by comparison of accessibility near transcription start sites (TSS ± 20 kb) to our scRNA-seq dataset for VISp[6] using median correlation (**Fig 2c**, **Methods**). We found that subclass-level assignments for each driver line matched closely with those observed for the same driver lines by scRNA-seq (**Supp Fig 5**). Once assigned, clusters from the same subclass (e.g. Vip or L5 IT) or distinct cell type (e.g. Pvalb Vipr2) were aggregated for peak calling and examination of accessibility patterns (**Fig 2e**). Comparisons of these scATAC-seq aggregate profiles to previously published ATAC-seq from cortical populations showed strong correspondence between aggregate profiles and populations (**Supp Fig 6**), and comparisons to previously published cortical scATAC-seq data demonstrate an increase in cell type resolution using our current dataset (**Supp Fig 7**).

Layer 5 (L5) of mouse cortex contains three major subclasses of excitatory neurons: intertelencephalic (IT) neurons that project to other cortical regions, near-projecting (L5 NP) neurons that have mostly local projections, and cortico-fugal (a subset of which is called pyramidal tract, L5 PT) neurons that project to subcortical brain regions such as thalamus[6,14]. The driver line *Rbp4-Cre* labels both L5 IT and L5 PT neurons in cortex, but not L5 NP[7]. Our scATAC-seq clustering identified L5 PT and L5 IT neurons in our dataset based on correlation with scRNA-seq cell types (**Fig 2c**). Labeling of these cells by *Rbp4-Cre* and retrograde labeling from a known L5 PT target region, the lateral posterior nucleus of the thalamus (LP, **Supp Fig 3**), validated that these cells are likely L5 IT (*Rbp4-Cre+* only) and L5 PT neurons (Rbp4 and LP Retro-ATAC-seq). We searched near transcriptomic marker genes for ~500 bp putative enhancer regions that were specific to L5 PT or L5 IT cells, and which had strong sequence conservation (**Fig 2e)**. We refer to these regions as mouse single-cell regulatory elements (mscREs, **Supp Table 5**).

To functionally test mscREs, we cloned their genomic sequences upstream of a minimal beta-globin promoter driving fluorescent proteins SYFP2 or EGFP in a recombinant adeno-

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

associated virus (AAV) genome (**Fig 3a**). These constructs were packaged using a PHP.eB serotype, which can cross the blood-brain barrier[15], to enable delivery by retro-orbital injection. We screened four mscREs for L5 PT cells and two for L5 IT (**Supp Table 5**). Two weeks after retro-orbital injection, we collected the brains of infected mice and screened expression by visual inspection of native fluorescence and immunohistochemistry to enhance SYFP2 and EGFP signal. Two of these enhancers provided specific labeling of cells in L5 (**Fig 3b**), and were selected for further validation.

To assess the utility of enhancer-driven fluorophores as viral tools, we performed retro-orbital injection of the mscRE4-SYFP2 virus in additional animals. From two of these, we dissected L5 of VISp, sorted labeled cells by FACS (**Supp Fig 2**) ,and performed scRNA-seq as described previously[6]. We compared scRNA-seq expression profiles to a VISp reference dataset[6] using centroid classification of cell types (**Methods**). We found that the mscRE4-SYFP2 virus yielded >91% specificity for L5 PT cells within L5 (**Fig 3c**). We confirmed labeling of L5 PT cells by electrophysiological characterization of labeled vs unlabeled cells in the cortex (**Fig 3c,d**, **Supp Fig 8**). Cells labeled by mscRE4 had characteristics of L5 PT neurons, whereas cells that were label-negative more closely matched L5 IT neurons[16]. This experiment demonstrates the utility of these viral tools for electrophysiology experiments targeted to specific neuronal subclasses for which driver lines are not available. Finally, we tested stereotaxic injection of the mscRE4 fluorophore viruses directly into VISp. We found that we could achieve extremely bright and specific labeling using stereotaxic injection, although the specificity depended on the volume of injection, likely reflecting a loss of specificity at high numbers of genome copies per cell (**Supp Fig 9**).

L5 PT cells are often difficult to isolate from single-cell suspensions when in a heterogeneous mixture with other cell types due to differential cell survival[6,7], and there is currently no reliable driver line to selectively label L5 PT cells. We used retro-orbital injection of the mscRE4-SYFP2 virus to enhance our scATAC-seq dataset by sorting cells labeled by mscRE4 for FACS. As expected based on scRNA-seq analysis, 55 of 61 high-quality mscRE4 scATAC-seq profiles clustered together with other L5 PT samples (90.2%).

Although fluorophore expression provided enough signal to sort cells by FACS or perform patch-clamp experiments, expression of a recombinase from a specific enhancer virus would enable the utility of these tools as drivers for reporter lines that express fluorophores, activity reporters, opsins, or other genes that are too large to package in AAVs[3,17]. To test the specificity of enhancer-driven recombinase expression, we cloned mscRE4 into constructs containing a minimal beta-globin enhancer driving destabilized Cre (dgCre), iCre, FlpO, or tTA2, and packaged them in PHP.eB viruses (**Fig 4a**). These viruses were delivered by retro-orbital injection into mice with genetically encoded reporters for each recombinase (Ai14 for dgCre and iCre[18]; Ai65F for FlpO[17]; and Ai63 for tTA2[3]). Labeling was assessed by sectioning and microscopy of native fluorescence (**Fig 4a**). FlpO, iCre, and tTA2 viral constructs yielded highly specific labeling of cells in L5 of the mouse cortex, while dgCre had non-specific labeling of cortical layers. We applied the same strategy to screen both mscRE4 and mscRE16 drivers of FlpO, iCre, and/or tTA2 by retro-orbital injection at two different titers (1×10^{10} and 1×10^{11} total

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

genome copies, GC). We found that the specificity and completeness of labeling depended heavily on both the injected titer and the recombinase-reporter combination used in these experiments (**Supp Fig 10**). Based on these experiments, we chose a single titer for each virus for in-depth characterization, and injected additional animals for scRNA-seq and whole-brain two-photon tomography by TissueCyte (**Figure 4b-e**, **Fig 5a-b**, **Supp Fig 11**). We found that each of these viruses had a high degree of layer and subclass specificity in the cortex, with 87.5% of cells labeled by mscRE4-FlpO corresponding to L5 PT cells (**Fig 4c**) and 42% of cells labeled by mscRE16-FlpO corresponding to L5 IT cells (**Fig 4e**), with little overlap. TissueCyte imaging revealed that two viruses labeled additional subcortical populations (mscRE4 in APr, CEa, and HIP, **Fig 5a**; and mscRE16 in pons, BLA, and HIP, **Supp Fig 11**).

Viruses can also be coadministered to label multiple populations of cells, either exclusively or intersectionally (**Fig 5c**). This strategy reduces the need for triple- or quadruple crosses to obtain co-labeled populations of cells. We tested brain-wide co-labeling of both L5 IT and L5 PT populations by retro-orbital injection of mscRE4-iCre (to label L5 PT cells, green) and mscRE16-FlpO (to label L5 IT cells, red) in the same Ai65F;Ai140 animal (**Fig 5d**). We found distinct labeling of these two cell populations in L5 by microscopy (**Fig 5e**), demonstrating that multiple enhancer-driven viruses can be used to simultaneously label or perturb populations of prospectively defined subclasses in the same animal.

Discussion

Single-cell genomics methods continue to uncover the diversity of cell types in complex tissues. We have demonstrated that a combination of scRNA-seq and scATAC-seq identifies functional subclass-specific enhancers that can label or perturb specific cell populations using viral tools at unprecedented cell type resolution. Previous methods for enhancer identification, such as STARR-seq[19], CRE-seq[20], and EDGE[21] rely on large screens of genomic elements identified in bulk tissues, with retrospective identification of labeled cell types. In contrast, our methodology extends concepts previously used in the retina, where defined populations of cell types were used to identify putative regulatory elements for screening[22,23]. Utilizing single-cell measurements of chromatin accessibility allowed us to prospectively generate highly targeted tools with minimal screening of putative enhancers.

The rich scATAC-seq dataset in this study provides a foundational resource for locating enhancer elements for targeting and perturbation of diverse cell types in the mouse cortex. We were able to resolve cell subclasses in Layer 5 of the mouse cortex, yielding viral tools with cell type resolution beyond available driver lines. These viral tools are immediately useful in conjunction with existing reporter and/or driver lines to augment and expand existing genetic toolkits[3]. Further optimization of these natural regulatory elements may result in increased specificity and higher on-target expression, as demonstrated for AAV tools targeting cell types in muscle[24]. In some cases, these highly targeted viral tools may supercede germline transgenesis for labeling and perturbation of specific cell types. Targeted, enhancer-driven viral reagents open new frontiers for the exploration of cell types in the brain.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1.

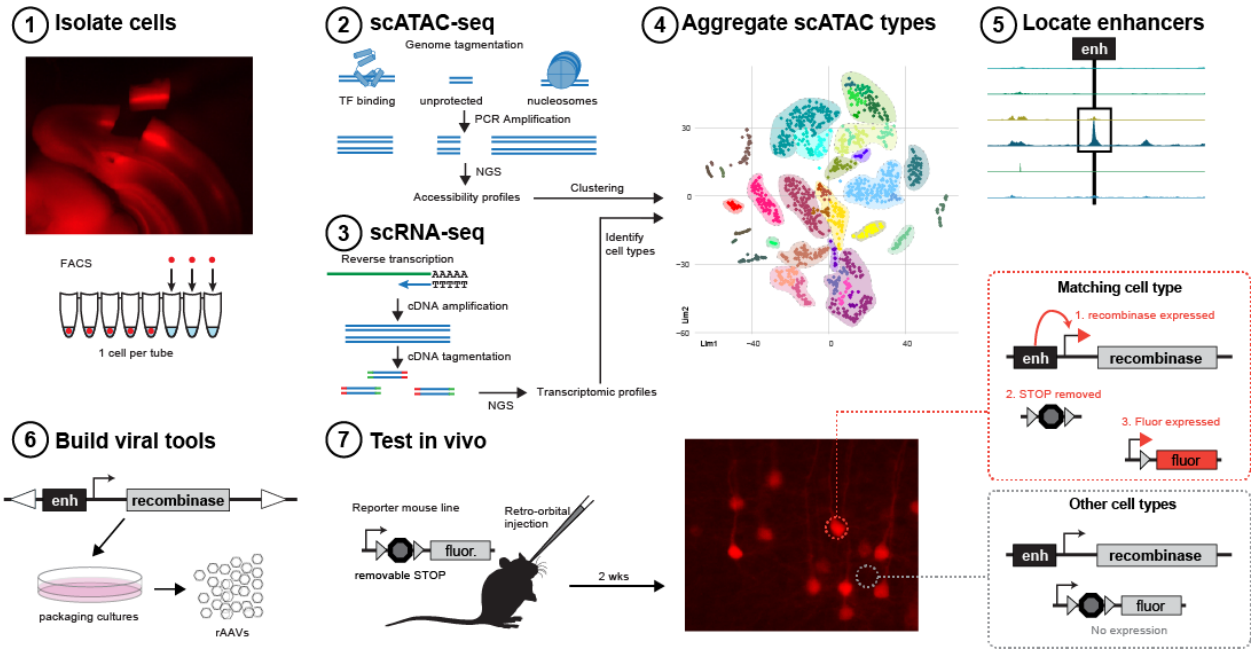
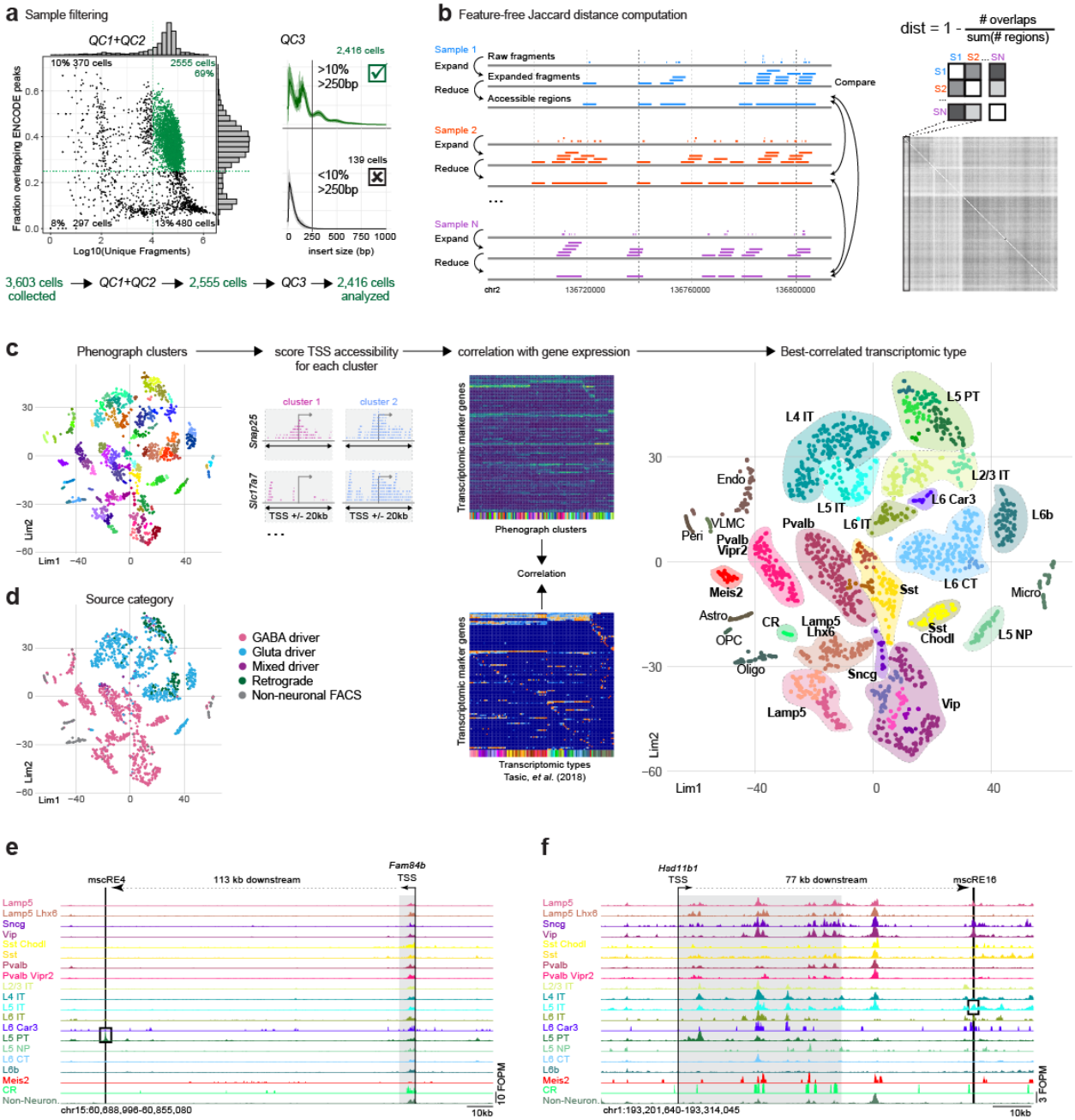


Figure 1. Overview of enhancer discovery for viral tools. To build cell type-specific labeling tools, we isolated cells from adult mouse cortex, performed scATAC-seq, clustered samples, and compared them to scRNA-seq datasets to identify the clusters. We then pooled single cells matching the same transcriptomic types and searched the genome for type-specific putative enhancers. These regions were cloned upstream of a minimal promoter in an AAV genomic backbone, which was used to generate scAAVs or rAAVs. These viral tools were delivered by retro-orbital or stereotaxic injection to label specific cortical populations. In principle, in cells with a matching cell type, enhancers recruit their cognate transcription factors to drive cell type-specific expression. In other cells, viral genomes are present, but transcripts are not expressed. However, it is necessary to test the enhancer constructs for specificity, as not all enhancers behave as expected.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2.



bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2. Identification of cell types in scATAC-seq data. (a) Quality control criteria for scATAC-seq. For analysis, we retained samples with >10,000 uniquely mapped fragments (QC1) that overlapped ENCODE whole-cortex DNase-seq peaks with > 25% of fragments (QC2), and which had nucleosomal fragment size structure identified by > 10% of all aligned fragments with an insert size > 250bp (QC3). **(b)** Samples that passed QC were prepared for clustering by downsampling to 10,000 unique fragments, extending fragment lengths to 1 kb, and reducing overlapping fragments to identify accessible regions. Samples were then compared by computing the number of overlaps between accessible regions, and a Jaccard distance was computed. These distances were used as input for t-SNE projection. **(c)** Samples were clustered in t-SNE space using the Phenograph implementation of Louvain clustering. To identify the cell types within these clusters, cells from each cluster were pooled, and we counted the number of fragments within 20kb of each TSS. We then selected marker genes for transcriptomic clusters from Tasic, *et al*. (2018), and performed correlations between TSS accessibility scores and log-transformed gene expression. The scRNA-seq cluster with the highest correlation score was assigned as the identity for each Phenograph cluster, and clusters with the same transcriptomic mapping were combined for downstream analyses. **(d)** t-SNE as in (c) with samples labeled according to the predominant cell class of their source driver line or experimental source. Cells from the same class are grouped as expected. **(e)** scATAC-seq data were grouped by subclass or distinct type (Methods) based on transcriptomic mapping, and aggregated fragment overlaps were plotted near the Fam84b gene after normalization for fragment count. A peak of accessibility specific to L5 PT samples was identified 113 kb downstream from the Fam84b TSS. FOPM, fragment overlaps per million reads. (**f**) Tracks of fragment overlaps per million fragments (FOPM) for each scATAC-seq group near the Hsd11b1 gene, a marker for L5 IT neurons. A putative regulatory element with higher accessibility in L5 IT cells than in other excitatory types was identified 77 kb downstream of the Hsd11b1 TSS.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3.

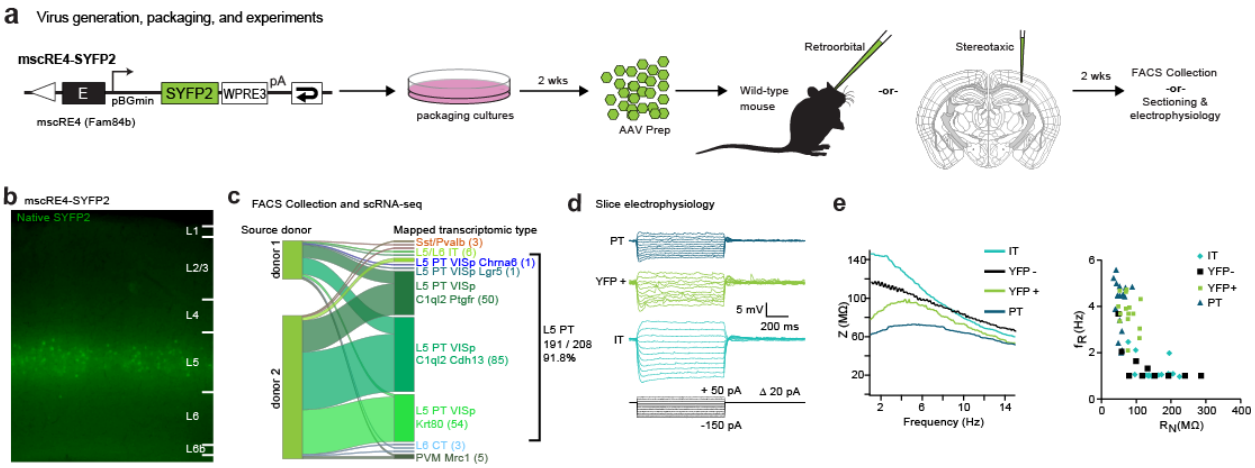


Figure 3. Direct fluorophore labeling of L5 PT neurons by an enhancer-driven virus. (a) Direct enhancer-driven expression of a fluorophore was tested by cloning the putative Fam84b enhancer, mscRE4, in a self-complementary AAV backbone (scAAV) construct with a minimal promoter driving an SYFP2-WPRE3. After packaging, purification, and titering scAAVs were retro-orbitally injected into a wild-type mouse. **(b)** 2 weeks post-retro-orbital injection, distinct labeling was seen in L5 cells, which was microdissected, dissociated, and sorted by FACS for SYFP2+ cells for scRNA-seq. **(c)** After sorting and sequencing, we were able to map scRNA-seq data to the Tasic, *et al*. (2018) cell type reference, and found 91.8% of all sorted cells were from L5 PT cell types. **(d)** Cortical slices from an animal labeled with the mscRE4-SYPF2 virus were used for electrophysiological characterization. Example impedance amplitude profiles obtained from a YFP+ and a YFP- neuron in VISp. For comparison, impedance amplitude profiles from an unlabeled PT-like and an IT-like neuron from somatosensory cortex are also shown. Resonance frequency is plotted as a function of input resistance. **(e)** Example voltage responses to a series of hyperpolarizing and depolarizing current injections for a YFP+ and a YFP– neuron. Example voltage responses obtained from unlabeled PT-like and IT-like neurons are also shown for reference.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4.

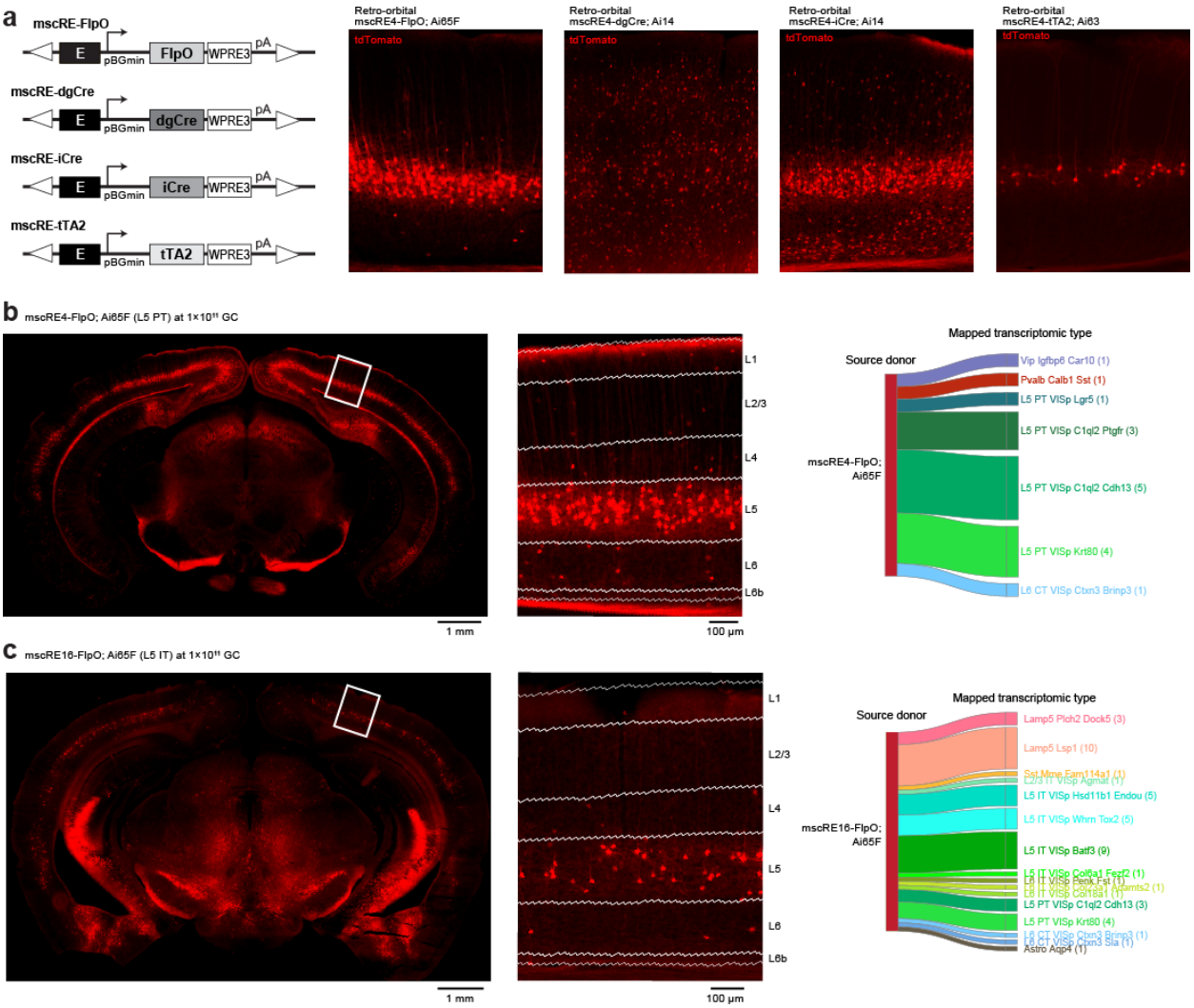
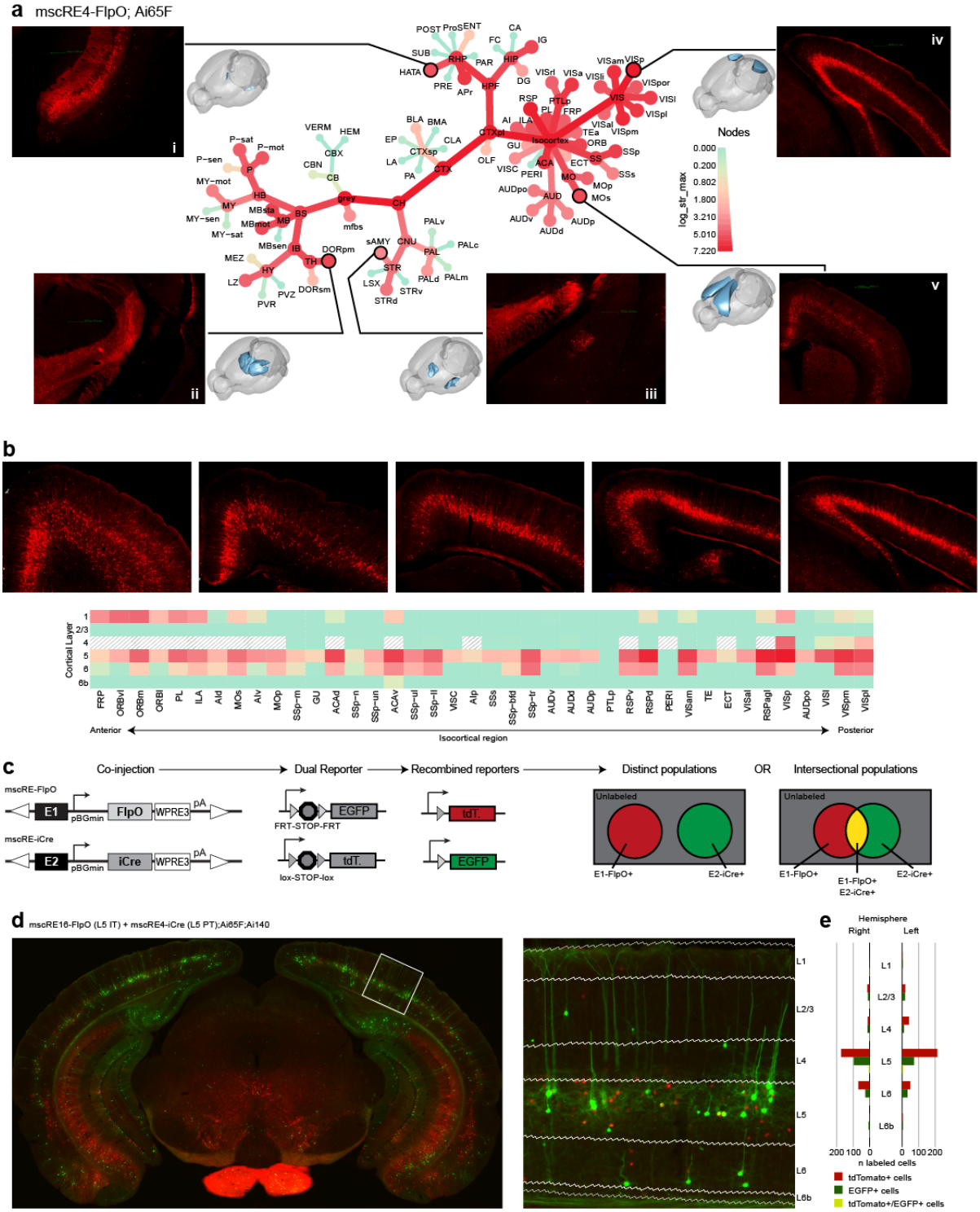


Figure 4. Enhancer-driven recombinase viruses as drivers for cell labeling.
(a) To test the use of regulatory elements to drive recombinases or transactivators in reporter mouse lines, we cloned mscRE4 upstream of a minimal promoter driving FlpO, dgCre, iCre, or tTA2. These viruses were then packaged in PHP.eB-serotype rAAVs, and were injected retro-orbitally into reporter mice. Images show native fluorescence in visual cortex 2 weeks post-injection. **(b)** Full-section imaging of a mscRE4-FlpO injection shows labeling throughout L5 of the posterior cortex. Inset region on the right corresponds to the white box on the left. Layer overlays from the Allen Brain Reference Atlas shows labeling restricted primarily to L5. **(c)** tdTomato+ cells were dissected from the full cortical depth and were collected by FACS for scRNA-seq. Transcriptomic profiles were mapped to reference cell types from Tasic, et al. (2018). 87.5% of cells (28 of 32) mapped to L5 PT cell types. **(d)** Full-section imaging of mscRE16-FlpO injection shows labeling in L5 of the cortex. Inset region on the right corresponds to the white box on the left. **(e)** scRNA-seq of tdTomato+ cells shows that L5 IT cell types are the most frequently labeled subclass of neurons at 42% (n = 20 of 48), but other subclasses are also labeled at this titer (Lamp5, 27%; L6 IT, 6%; L5 PT, 15%).

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5.



bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5. Brain-wide and intersectional labeling of cell types

(a) Results from full-brain imaging using TissueCyte. Sections throughout the whole brain of an Ai65F mouse after retro-orbital injection of mscRE-FlpO were aligned to the Allen Institute Common Coordinate Framework (CCF), and mapped to the Allen Brain Atlas structural ontology. A high-level overview of cell labeling throughout the structural ontology is represented by the taxonomic plot. “Grey” is the root of the plot, representing all grey-matter regions, and each branching of nodes shows child structures within each region. The size and color of nodes represents the maximum signal found among all children of the nodes, which allows one to follow the tree to the source of high signal within each structure. Insets display selected regions of high or specific signal. Region acronyms correspond to the Allen Brain Adult Mouse Atlas. **(b)** Further division of the isocortical regions in the TissueCyte dataset to the level of cortical layers allows brain-wide quantification of layer-specific signal. Representative cortical sections from the TissueCyte dataset are shown along the top, from most anterior to most posterior (left to right). The heatmap shows quantification of the signal in each region and layer. Agranular regions, which lack layer 4, have hashing in the L4 row. **(c)** Diagrams showing the use of co-injected recombinase viruses in a dual-reporter system for co-labeling or intersectional labeling of cell types. In this experiment, one virus driving FlpO and a second driving iCre are co-injected into a mouse with genetically-encoded Flp-dependent and Cre-dependent reporters. In target cell types, enhancers will drive the recombinases, which will permanently label their target cell types. If the enhancers selected are mutually exclusive, distinct populations will be labeled. If they overlap, intersectional labeling is possible. **(d)** Native fluorescence imaging of an Ai65F;Ai140 dual-reporter mouse line retro-orbitally injected with mscRE16-FlpO (red) and mscRE4-iCre (green). These enhancers are expected to label mutually-exclusive cell types in L5 of the cortex. The region in the white box corresponds the inset image, showing strong labeling of cells in L5. **(e)** Cell counts within each layer for all cortical regions labeled with EGFP (mscRE4; L5 PT), tdTomato (mscre16; L5 IT), or both in the image in (d).

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and Methods

Mouse breeding and husbandry

Mice were housed under Institutional Care and Use Committee protocols 1508 and 1802 at the Allen Institute for Brain Science, with no more than five animals per cage, maintained on a 12 hr day/night cycle, with food and water provided *ad libitum*. Animals with anophthalmia or microphthalmia were excluded from experiments. Animals were maintained on a C57BL/6J genetic background.

Retrograde labeling

We performed stereotaxic injection of CAV-Cre (gift of Miguel Chillon Rodrigues, Universitat Autònoma de Barcelona)[25] into brains of heterozygous or homozygous Ai14 mice using stereotaxic coordinates obtained from Paxinos adult mouse brain atlas[26]. Specific coordinates used for each injection are provided in **Supp. Table 3**. TdT+ single cells were isolated from VISp by FACS. Example FACS gating is provided in **Supp. Fig 2**.

Single cell ATAC

Single-cell suspensions of cortical neurons were generated as described previously[27], with the exception of use of papain in place of pronase for some samples, and the addition of trehalose to the dissociation and sorting medium for some samples as shown in **Supp. Table 2**. We then sorted individual cells using FACS with gating of negative-DAPI and positive-fluorophore labeling (tdTomato, EGFP, or SYFP2) to select for live neuronal cells or negative-DAPI and negative-fluorophore labeling for live non-neuronal cells. Example FACS gating is provided in **Supp. Fig 2**.

For Gm12878 scATAC, cells were grown on 10 cm plates in RPMI 1640 Medium (Gibco, Thermo Fisher Cat#11875093) supplemented with 10% fetal bovine serum (FBS). At ~80% confluence, plates were washed with PBS and treated with Tripsyn-EDTA (Thermo Fisher Scientific, Cat#25300054) to release cells. Cells were placed in a 15 mL conical tube, centrifuged, and washed with PBS containing 1% FBS. Cells were then resuspended in PBS with 1% FBS and 2 ng/mL DAPI (DAPI*2HCl, Life Technologies Cat#D1306) for FACS sorting.

Single cells were sorted into 200 µL 8-well strip tubes containing 1.5 µL tagmentation reaction mix (0.75 µL Nextera Reaction Buffer, 0.2 µL Nextera Tn5 Enzyme, 0.55 µL water). After collection, cells were briefly spun down in a bench-top centrifuge, then immediately tagmented at 37˚C for 30 minutes in a thermocycler. After tagmentation, we added 0.6 µL Proteinase K stop solution to each tube (5 mg/mL Proteinase K solution (Qiagen), 50 mM EDTA, 5 mM NaCl, 1.25% SDS) followed by incubation at 40˚C for 30 minutes in a thermocycler. We then purified the tagmented DNA using AMPure XP beads (Beckman Coulter) at a ratio of 1.8:1 resuspended beads to reaction volume (3.8 µL added to 2.1 µL), with a final elution volume of 11 µL. Libraries were indexed and amplified by the addition of 15 uL 2X Kapa HiFi HotStart ReadyMix and 2 uL Nextera i5 and i7 indexes to each tube, followed by incubation at 72˚C for 3 minutes and PCR

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(95˚C for 1 minute, 22 cycles of 98˚C for 20 seconds, 65˚C for 15 seconds, and 72˚C for 15 seconds, then final extension at 72˚C for 1 minute). After amplification, sample concentrations were measured using a Quant-iT PicoGreen assay (Thermo Fisher) in duplicate. For each sample, the mean concentration was calculated by comparison to a standard curve, and the mean and standard deviation of concentrations was calculated for each batch of samples. Samples with a concentration greater than 2 standard deviations above the mean were not used for downstream steps, as these were found in early experiments to dominate sequencing runs. All other samples were pooled by combining 5 µL of each sample in a 1.5 mL tube. We then purified the combined library by adding Ampure XP beads in a 1.8:1 ratio, with final elution in 50 µL. The mixed library was then quantified using a BioAnalyzer High Sensitivity DNA kit (Agilent).

scATAC sequencing, alignment, and filtering

Mixed libraries, containing 60 to 96 samples, were sequenced on an Illumina MiSeq at a final concentration of 20-30 pM. After sequencing, raw FASTQ files were aligned to the GRCm38 (mm10) mouse genome using Bowtie v1.1.0 as described previously[27]. After alignment, duplicate reads were removed using samtools rmdup, which yielded only single copies of uniquely mapped paired reads in BAM format. For analysis, we filtered to remove samples with fewer than 10,000 paired-end fragments (20,000 reads), and with at least 10% of sequenced fragments longer than 250 bp. An additional filter was created using ENCODE whole cortex DNase-seq HotSpot peaks (sample ID ENCFF651EAU from experiment ID ENCSR00COF). Samples with less than 25% of paired-end fragments that overlapped DNase-seq peaks were removed from downstream analysis. Cells passing these criteria both had sufficient unique reads for downstream analysis, and had high-quality chromatin accessibility profiles as assessed by fragment size analysis (**Supp Fig 1**). As an additional QC check, we compared aggregate scATAC-seq data to bulk ATAC-seq data from matching Cre-driver lines, where available. We found that aggregate single-cell datasets matched well to previously published bulk datasets (**Supp Fig 6**).

Jaccard distance calculation, PCA and tSNE embedding, and density-based clustering

To compare scATAC-seq samples, we downsampled all cells to an equal number of uniquely aligned fragments (10,000 per sample), extended these fragments to a length of 1kb, then collapsed any overlapping fragments within each sample into regions based on the outer boundaries of overlapping fragments. We then counted the number of overlapping regions between every pair of samples, and divided by the total number of regions in both samples to obtain a Jaccard similarity score. These scores were converted to Jaccard distances (1 – Jaccard similarity), and the resulting matrix was used as input for t-stochastic neighbor embedding (t-SNE). After t-SNE, samples were clustered in t-SNE space using the RPhenograph package with k = 6 to obtain small groups of similar neighbors[13]. Cluster assignments used for correlation with transcriptomic data are showin **Supp. Table 6**.

Correlation with single-cell transcriptomics

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Phenograph-defined neighborhoods were assigned to cell subclasses and clusters by comparison of accessibility near transcription start site (TSS) to median expression values of scRNA-seq clusters at the cell type (e.g. L5 PT Chrna6) and at the subclass level (e.g. Sst) from mouse primary visual cortex[6]. To score each TSS, we retrieved TSS locations from the RefSeq Gene annotations provided by the UCSC Genome Browser database, and generated windows from TSS +/- 20kb. We then counted the number of fragments for all samples within each cluster that overlapped these windows. For comparison, we selected differentially expressed marker genes from the Tasic, *et al*. (2018) scRNA-seq dataset using the scrattch.hicat package for R. We then correlated the Phenograph cluster scores with the log-transformed median exon read count values for this set of marker genes for each scRNA-seq cluster from primary visual cortex, and assigned the transcriptomic cell type with the highest-scoring correlation. We found that this strategy of neighbor assignment and correlation allowed us to resolve cell types within the scATAC-seq data close to the resolution of the scRNA-seq data, as types that were split too far would resolve to the same transcriptomic subclass or type by correlation.

scATAC-seq grouping and peak calling

For downstream analysis, we grouped cell type assignments to the subclass level, with the exception of highly distinct cell types (Lamp5 Lhx6, Sst Chodl, Pvalb Vipr2, L6 IT Car3, CR, and Meis2). Unique fragments for all cells within each of these subclass/distinct type groups were aggregated to BAM files for analysis. Aligned reads from single cell subclasses/clusters were used to create Tag Directories and peaks of chromatin accessibility were called using HOMER[28] with settings “findPeaks -region -o auto”. The resulting peaks were converted to BED format.

Population ATAC of Sst neurons

We performed population ATAC-seq of neurons from Sst-IRES2-Cre;Ai14 mice as described previously[27]. Briefly, cells from the visual cortex of an adult mouse were microdissected and FACS sorted into 8-well strips as described above, but with 500 cells per well instead of single cells as for scATAC-seq. Cell membranes were lysed, and nuclei were pelleted before resuspension in the same tagmentation buffer described above at a higher volume (25 µL). Tagmentation was carried out at 37 C for 1 hour, followed by addition of 5 µL of Cleanup Buffer (900 mM NaCl, 300 mM EDTA), 2 µL 5% SDS, and 2 µL Proteinase K and incubation at 40°C for 30 minutes, and cleanup with AMPure XP beads (Beckman Coulter) at a ratio of 1.8:1 beads to reaction volume. Samples were amplified using KAPA HotStart Ready Mix (Kapa Biosystems, Cat# KK2602) and 2uL each of Nextera i5 and i7 primers (Illumina), quantified using a Bioanalyzer, and sequenced on an Illumina MiSeq.

Comparisons to bulk ATAC-seq data

For comparison to previously published studies, we used data from GEO accession GSE63137 from Mo, *et al*. (2015)[29] for Camk2a, Pvalb, and Vip neuron populations, and GEO accession GSE87548 from Gray, *et al*. (2017)[27] for Cux2, Scnn1a-Tg3, Rbp4, Ntsr1, Gad2, mES, and genomic controls. For these comparisons, we also included population ATAC-seq of Sst

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

neurons, described above. For each population, we merged reads form all replicates and downsampled each region to 6.4 million reads. We then called peaks using HOMER as described for aggregated scATAC-seq data, above. We used the BED-formatted peaks for scATAC-seq aggregates with or without bulk ATAC-seq datasets as input for comparisons using the DiffBind package for R as described previously[27].

Identification of mouse single-cell regulatory elements

We performed a targeted search for mouse single cell regulatory elements (mscREs) by performing pairwise differential expression analysis of scRNA-seq clusters to identify uniquely expressed genes in L5 PT and L5 IT subclasses across all glutamatergic subclasses. We then searched for unique peaks within 1 Mbp of each marker gene, and manually inspected these peaks for low or no accessibility in off-target cell types and for conservation. If a region of high conservation overlapped the peak region, but the peak was not centered on the highly conserved region, we adjusted the peak selection to include neighboring highly conserved sequence. For cloning, we centered our primer search on 500 bp regions centered at the middle of the selected peak regions, and included up to 100 bp on either side. Final region selections and PCR primers are shown in **Supplementary Table 5**.

Viral genome cloning

Enhancers were cloned from C57Bl/6J genomic DNA using enhancer-specific primers and Phusion high-fidelity polymerase (M0530S; NEB). Individual enhancers were then inserted into an rAAV or scAAV backbone that contained a minimal beta-globin promoter, gene, and bovine growth hormone polyA using standard molecular cloning approaches. Plasmid integrity was verified via Sanger sequencing and restriction digests to confirm intact inverted terminal repeat (ITR) sites.

Viral packaging and titering

Before transfection, 105 µg of AAV viral genome plasmid, 190 µg pHelper, and 105 µg AAV-PHP.eB were mixed with 5 mL of Opti-MEM I media (Reduced Serum, GlutaMAX; ThermoFisher Scientific) and 1.1 mL of a solution of 1 mg/mL 25 kDa linear Polyethylenimine (Polysciences) in PBS at pH 4-5. This cotransfection mixture was incubated at room temperature for 10 minutes. Recombinant AAV of the PHP.eB serotype was generated by adding 0.61 mL of this cotransfection mixture to each of ten 15-cm dishes of HEK293T cells (ATCC) at 70-80% confluence. 24 hours post-transfection, cell medium was replaced with DMEM (with high glucose, L-glutamine and sodium pyruvate; ThermoFisher Scientific) with 4% FBS (Hyclone) and 1% Antibiotic-Antimycotic solution. Cells were collected 72 hours post transfection by scraping in 5mL of medium, and were pelleted at 1500 rpm at 4C for 15 minutes. Pellets were suspended in a buffer containing 150 mM NaCl, 10 mM Tris, and 10 mM MgCl2, pH 7.6, and were frozen in dry ice. Cell pellets were thawed quickly in a 37°C water bath, then passed through a syringe with a 21-23G needle 5 times, followed by 3 more rounds of freeze/thaw, and 30 minutes of incubation with 50 U/ml Benzonase (Sigma-Aldrich) at 37°C. The suspension was then centrifuged at 3,000 × g and purified using a layered iodixanol step

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

gradient (15%, 25%, 40%, and 60%) by centrifugation at 58,000 rpm in a Beckman 70Ti rotor for 90 minutes at 18°C by extraction of a volume below the 40-60% gradient layer interface. Viruses were concentrated using Amicon Ultra-15 centrifugal filter unit by centrifugation at 3,000 rpm at 4°C, and reconstituted in PBS with 5% glycerol and 35 mM NaCl before storage at -80°C.

Virus titers were measured using qPCR with a primer pair that recognizes a region of 117bp in the AAV2 ITRs (Forward: GGAACCCCTAGTGATGGAGTT; Reverse: CGGCCTCAGTGAGCGA). QPCR reactions were performed using QuantiTect SYBR Green PCR Master Mix (Qiagen) and 500 nM primers. To determine virus titers, a positive control AAV with known titer and newly produced viruses with unknown titers were treated with DNAse I. Serial dilutions (1/10, 1/100, 1/500, 1/2500, 1/12500, and 1/62500) of both positive control and newly generated viruses were loaded on the same qPCR plate. A standard curve of virus particle concentrations vs C_q values was generated based on the positive control virus, and the titers of the new viruses were calculated based on the standard curve.

Retro-orbital injections

To introduce AAV viruses into the blood stream, 21 day old or older C57Bl/6J, Ai14, Ai65F, or Ai63 mice[17] were briefly anesthetized by isoflurane and $1x10^{10}$-$1x10^{11}$ viral genome copies (gc) was delivered into the retro-orbital sinus in a maximum volume of 50 µL or less. This approach has been utilized previously to deliver AAV viruses across the blood brain barrier and into the murine brain with high efficiency[15]. For delivery of multiple AAVs, the viruses were mixed beforehand and then delivered simultaneously into the retro-orbital sinus. Animals were allowed to recover and then sacrificed 1-3 weeks post-infection in order to analyze virally-introduced transgenes within the brain.

Stereotaxic injections

Viral DNA was packaged in a PHP.eB serotype[15] to produce recombinant adeno-associated virus (rAAV) for mscRE4-minBGprom-EGFP-WPRE3, mscRE4-minBGprom-tTa2-WPRE3, and mscRE4-minBGprom-FlpO-WPRE3 viruses (titers: 1.64 x 1014, 5.11 x 1013, 6.00 x 1013, respectively), or self-complementary AAV (scAAV) for mscRE4-minBGprom-SYFP2-WPRE3-BGHpA virus (titer 1.34 x 1013). Each virus was delivered bilaterally at 250 and 50 nL or 50 and 25 nL into the primary visual cortex (VISp; coordinates: A/P: -3.8, ML: -2.5, DV: 0.6) of male and female C57Bl6/J and wild-type transgenic mice (Htr2a-Cre (-), SST-IRES-Cre; Ai67(-), Cck-IRES-Cre (-)) for rAAV-mscRE4-minBGprom-EGFP-WPRE3 and scAAV-mscRE4-minBGprom-SYFP2-WPRE3 viruses, or heterozygous Ai65F and Ai63 mice for rAAV-mscRE4-minBGprom-FlpO-WPRE3 and rAAV-mscRE4-minBGprom-tTa2-WPRE3 viruses, respectively, using a pressure injection system (Nanoject II, Drummond Scientific Company, Catalog# 3-000-204). To mark the injection site, rAAV-EF1a-tdTomato or rAAV-EF1a-EGFP was co-injected at a dilution of 1:10 with experimental virus. The expression for all viruses was analyzed at 14 days post-injection. For tissue processing, mice were transcardially perfused with 4% paraformaldehyde (PFA) and post-fixed in 30% sucrose for 1-2 days. 50 µm sections were prepared using a

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

freezing microtome and fluorescent images of the injections were captured from mounted sections using a Nikon Eclipse TI epi-fluorescent microscope.

Immunohistochemistry

Mice were transcardially perfused with 0.1M phosphate buffered saline (PBS) followed by 4% paraformaldehyde (PFA). Brains were removed, post-fixed in PFA overnight, followed by an additional incubation overnight in 30% sucrose. Coronal sections (50 µm) were cut using a freezing microtome and native fluorescence or antibody-antibody enhanced was analyzed in mounted sections. To enhance the EGFP fluorescence, a rabbit anti-GFP antibody was used to stain free floating brain sections. Briefly, sections were rinsed three times in PBS, blocked for 1 hour in PBS containing 5% donor donkey serum, 2% bovine serum albumin (BSA) and 0.2% Triton X-100, and incubated overnight at 4°C in the anti-GFP primary antibody (1:2000; Abcam ab6556). The following day, sections were washed three times in PBS and incubated in blocking solution containing an Alexa 488 conjugated secondary antibody (1:1500; Invitrogen), washed in PBS, and mounted in Vectashield containing DAPI (H-1500, Vector Labs). Epifluorescence images of native or antibody-enhanced fluorescence were acquired on a Nikon Eclipse Ti microscope or on a TissueCyte 1000 (Tissue Vision) system.

Single cell RNA sequencing and cell type mapping

scRNA-seq was performed using the SMART-Seq v4 kit (Takara Cat#634894) as described previously[6]. In brief, single cells were sorted into 8-well strips containing SMART-Seq lysis buffer with RNase inhibitor (0.17 U/uL), and were immediately frozen on dry ice for storage at -80 C. SMART-Seq reagents were used for reverse transcription and cDNA amplification. Samples were tagmented and indexed using a NexteraXT DNA Library Preparation kit (Illumina FC-131-1096) with NexteraXT Index Kit V2 Set A (FC-131-2001) according to manufacturer's instructions except for decreases in volumes of all reagents, including cDNA, to 0.4x recommended volume. Full documentation for the scRNA-seq procedure is available in the 'Documentation' section of the Allen Institute data portal at http://celltypes.brain-map.org/. Samples were sequenced on an Illumina HiSeq 2500 or Illumina MiSeq as 50 bp paired-end reads. Reads were aligned to GRCm38 (mm10) using STAR v2.5.3[30] in towpassMode, and exonic read counts were quantified using the GenomicRanges package for R as described in Tasic, et al., (2018). To determine the corresponding cell type for each scRNA-seq dataset, we utilized the scrattch.hicat package for R[6]. We selected marker genes that distinguished each cluster, then used this panel of genes in a bootstrapped centroid classifier which performed 100 rounds of correlation using 80% of the marker panel selected at random in each round.

Comparisons to previous scATAC-seq studies

For comparisons to GM12878 datasets, raw data from Cusanovich, et al. (2015) was downloaded from GEO accession GSE67446 and Buenrostro, et al. (2015) from GEO accession GSE65360. Processed 10x Genomics data was retrieved from the 10x Genomics website. Buenrostro, Cusanovich, and our own Gm12878 samples were aligned to the hg38 human genome using the same bowtie pipeline described above for mouse samples to obtain

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

per-cell fragment locations. 10x Genomics samples were analyzed using fragment locations provided by 10x Genomics. For comparison to TSS regions, we used the RefSeq Genes tables provided by the UCSC Genome Browser database for hg19 (for 10X data) and for hg38 (for other datasets). To compare to ENCODE peaks, we used ENCODE Gm12878 DNA-seq HotSpot results from ENCODE experiment ID ENCSR000EJD aligned to hg19 (ENCODE file ID ENCFF206HYT) or hg38 (ENCODE file ID ENCFF773SPT).

For comparisons to previously published mouse cortex datasets, raw FASTQ files were downloaded from GEO accession GSE111586 for Cusanovich, et al. (2018) and from GEO accession GSE100033 for Priessl, et al. (2018). Multiplexed files were aligned to the mm10 genome using bowtie, and were demultiplexed using an R script prior to removal of duplicate location alignments. Only barcodes with > 1,000 mapped reads were retained for analysis. Per-barcode statistics were computed using the same algorithms used for per-cell statistics from our own dataset, and samples from the Cusanovich, et al. (2018) dataset that passed our QC criteria were subjected to the same analysis pipeline as our own data after demultiplexing and duplicate read removal.

Physiology

Coronal mouse brain slices were prepared using the NMDG protective recovery method[31]. Mice were deeply anesthetized by intraperitoneal administration of Advertin (20 mg/kg) and were perfused through the heart with an artificial cerebral spinal (ACSF) solution containing (in mM): 92 NMDG, 2.5 KCl, 1.25 NaH_2PO_4, 30 $NaHCO_3$, 20 HEPES, 25 glucose, 2 thiourea, 5 Na-ascorbate, 3 Na-pyruvate, .5 $CaCl_2{\cdot}4H_2O$ and 10 $MgSO_4{\cdot}7H_2O$. Slices (300 µm) were sectioned on a Compresstome VF-200 (Precisionary Instruments) using a zirconium ceramic blade (EF-INZ10, Cadence). After sectioning, slices were transferred to a warmed (32-34°C) recovery chamber filled with NMDG ACSF under constant carbogenation. After 12 minutes, slices were transferred to a holding chamber containing an ACSF made of (in mM) 92 NaCl, 2.5 KCl, 1.25 NaH_2PO_4, 30 $NaHCO_3$, 20 HEPES, 25 glucose, 2 thiourea, 5 Na-ascorbate, 3 Na-pyruvate,128 $CaCl_2{\cdot}4H_2O$ and 2 $MgSO_4{\cdot}7H_2O$ continuously bubbled with 95/5 O_2/CO_2.

For patch clamp recordings, slices were placed in a submerged, heated (32-34°C) recording chamber that was continuously perfused with ACSF under constant carbogenation containing (in mM): 119 NaCl, 2.5 KCl, 1.25 NaH_2PO_4, 24 $NaHCO_3$, 12.5 glucose, 2 $CaCl_2{\cdot}4H_2O$ and 2 $MgSO_4{\cdot}7H_2O$ (pH 7.3-7.4). Neurons were viewed with an Olympus BX51WI microscope and infrared differential contrast optics and a 40x water immersion objective. Patch pipettes (3-6 MΩ) were pulled from borosilicate glass using a horizontal pipette puller (P1000, Sutter Instruments). Electrical signals were acquired using a Multiclamp 700B amplifier and PClamp 10 data acquisition software (Molecular Devices). Signals were digitized (Axon Digidata 1550B) at 10-50 kHz and filtered at 2-10 kHz. Pipette capacitance was compensated and the bridge balanced throughout whole-cell current clamp recordings. Access resistance was 8-25 MΩ).

Data were analyzed using custom scripts written in Igor Pro (Wavemetrics). All measurements were made at resting membrane potential. Input resistance (R_N) was calculated from the linear portion of the voltage-current relationship generated in response to a series of 1s current

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

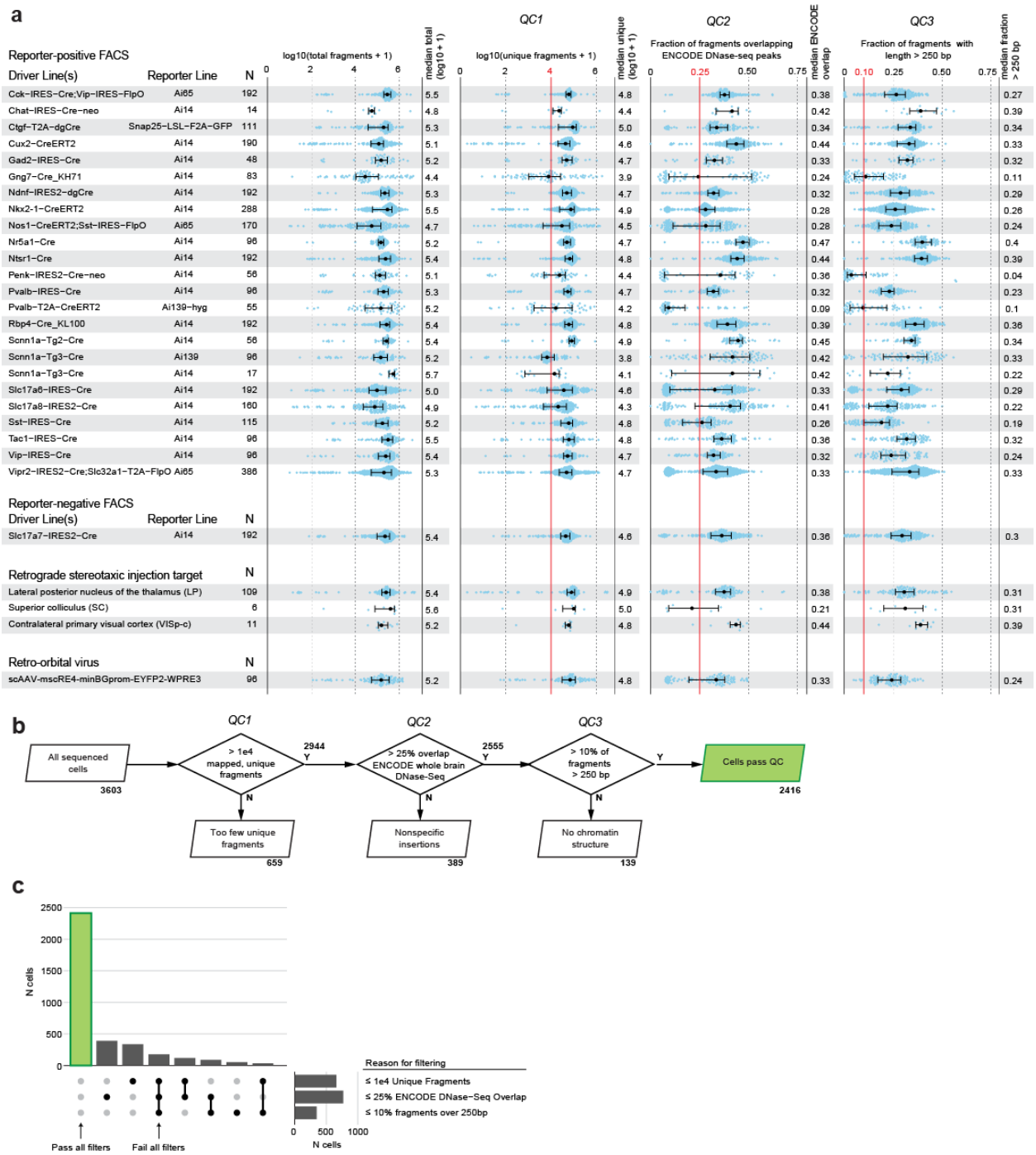
injections. The maximum and steady state voltage deflections were used to determine the maximum and steady state of R_N, respectively. Voltage sag was fined as the ratio of maximum to steady-state R_N. Resonance frequency (f_R) was determined from the voltage response to a constant amplitude sinusoidal current injection that either linearly increased from 1-15 Hz over 15 s or increased logarithmically from 0.2-40 Hz over 20 s. Impedance amplitude profiles were constructed from the ratio of the fast Fourier transform of the voltage response to the fast Fourier transform of the current injection. f_R corresponded to the frequency at which maximum impedance was measured. While the majority of neurons we included in this study were located in primary visual cortex (n=10 YFP+, 10 YFP-), we also made recordings from motor cortex (n=1 YFP+) and primary somatosensory cortex (n=4 YFP). For illustrative purposes, we also compared the properties of YFP+ and YFP- neurons to 32 L5 pyramidal neurons located in somatosensory cortex from an uninfected mouse. To classify these neurons as IT-like or PT-like, we used Divisive Analysis of Clustering (diana) from the cluster package in R. I_h-related membrane properties are known to differentiate IT and PT neurons across many brain regions[16]. As such, features included in clustering were restricted to the I_h- related membrane properties - sag ratio, R_N and f_R. To assess statistical significance of clustering, we used the sigclust package in R.

TissueCyte imaging and analysis

TissueCyte images were collected, registered, and segmented as described previously[32]. After registration, 3D arrays of signal binned to 25 um voxels were analyzed in R by subtraction of background, and averaging the signal in the finest structure in the Allen Brain Atlas structural ontology. To propagate signals from fine to coarse structure in the ontology, we performed hierarchical calculations that assigned the maximum value of child nodes in the ontology to each parent from the bottom to the top of the ontology. We then filtered the ontology to remove very fine structures, and used the taxa and metacodeR packages for R[33] to display the resulting ontological relationships and structure scores.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Fig 1

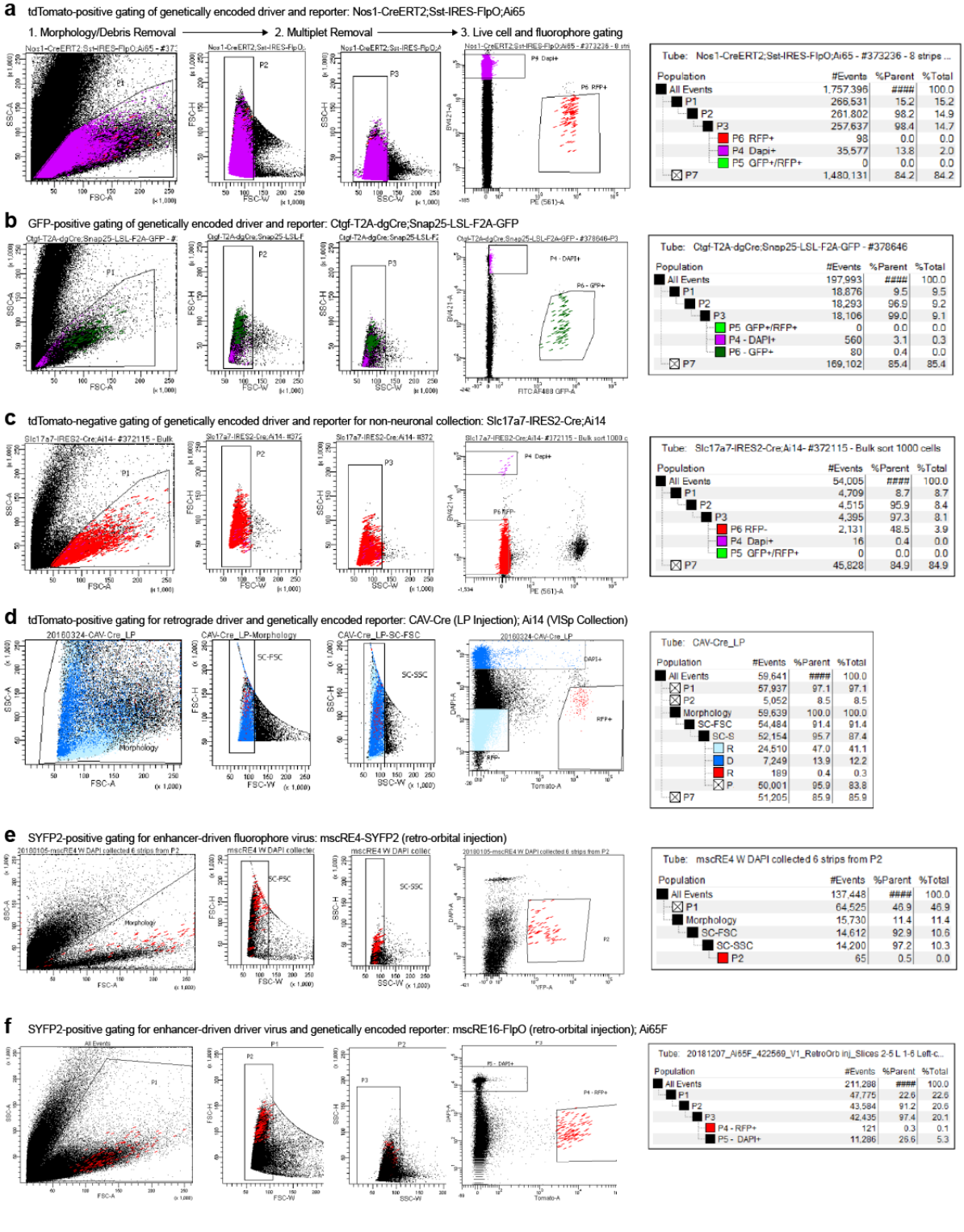


bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 1. Cell sources and QC Statistics.
(a) Dot plots showing sequencing statistics for each cell source in the scATAC-seq dataset. See Supplementary Table 1 for mouse line details. N, number of cells collected. Total fragments is the count of all sequenced fragments for each cell. Unique fragments is the number of uniquely mapped fragments used for analysis, and was used for the first QC cutoff of > 1e4 unique fragments. Fraction of fragments overlapping ENCODE DNase-seq peaks was computed using uniquely mapped fragments, and was used for the second QC cutoff of > 0.25. Fraction of fragments with length > 250bp was computed using unique fragments, and provides the third QC cutoff of > 0.1. **(b)** Flow chart showing how many samples were filtered using these 3 QC criteria, sequentially. **(c)** Set barplot showing how many cells were flagged with each combination of QC criteria.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 2



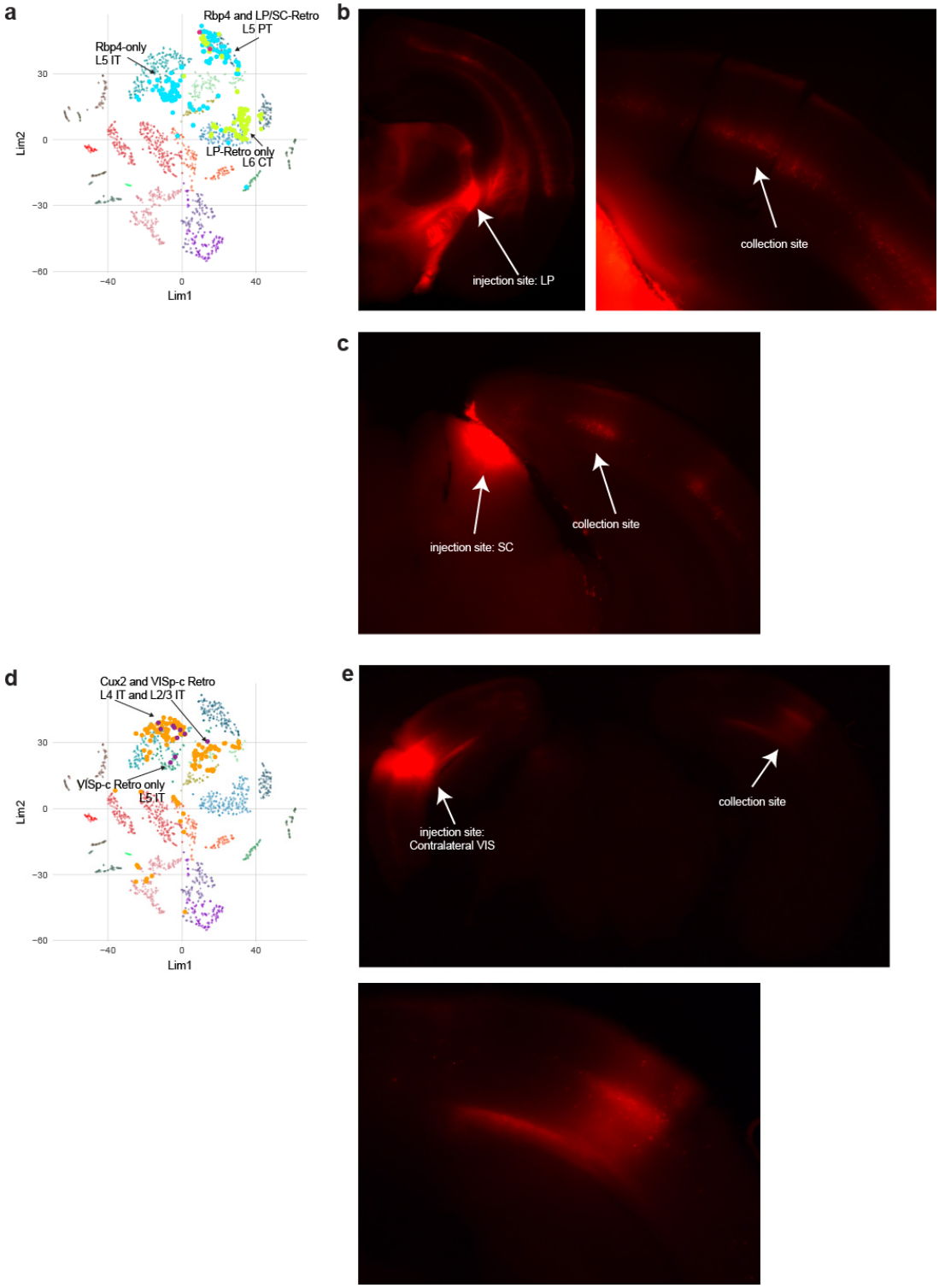
bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 2. FACS Gating examples.

(a) All FACS sorts followed a similar gating strategy: Morphology and debris removal using FSC-A and SSC-A; Removal of doublets/multiplets using FSC-W x FSC-H and SSC-W x SSC-H gating; and selection of live cells with or without fluorescent labels using DAPI and fluorophore signals. This panel shows example filtering for tdTomato-labeled cells from a genetically encoded driver and reporter; **(b)** shows example gating for GFP-positive cells from a genetically encoded driver and reporter; **(c)** shows example gating for tdTomato-negative cells in a genetically encoded driver and reporter line for the collection of non-neuronal cells; **(d)** shows gating for cells labeled by a retrograde driver in a genetically encoded reporter mouse line; **(e)** shows gating for direct fluorophore labeling of cells from injection of mscRE4-SYFP2; and **(f)** shows example gating of cells labeled by a viral driver in a genetically encoded reporter mouse line.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Fig 3



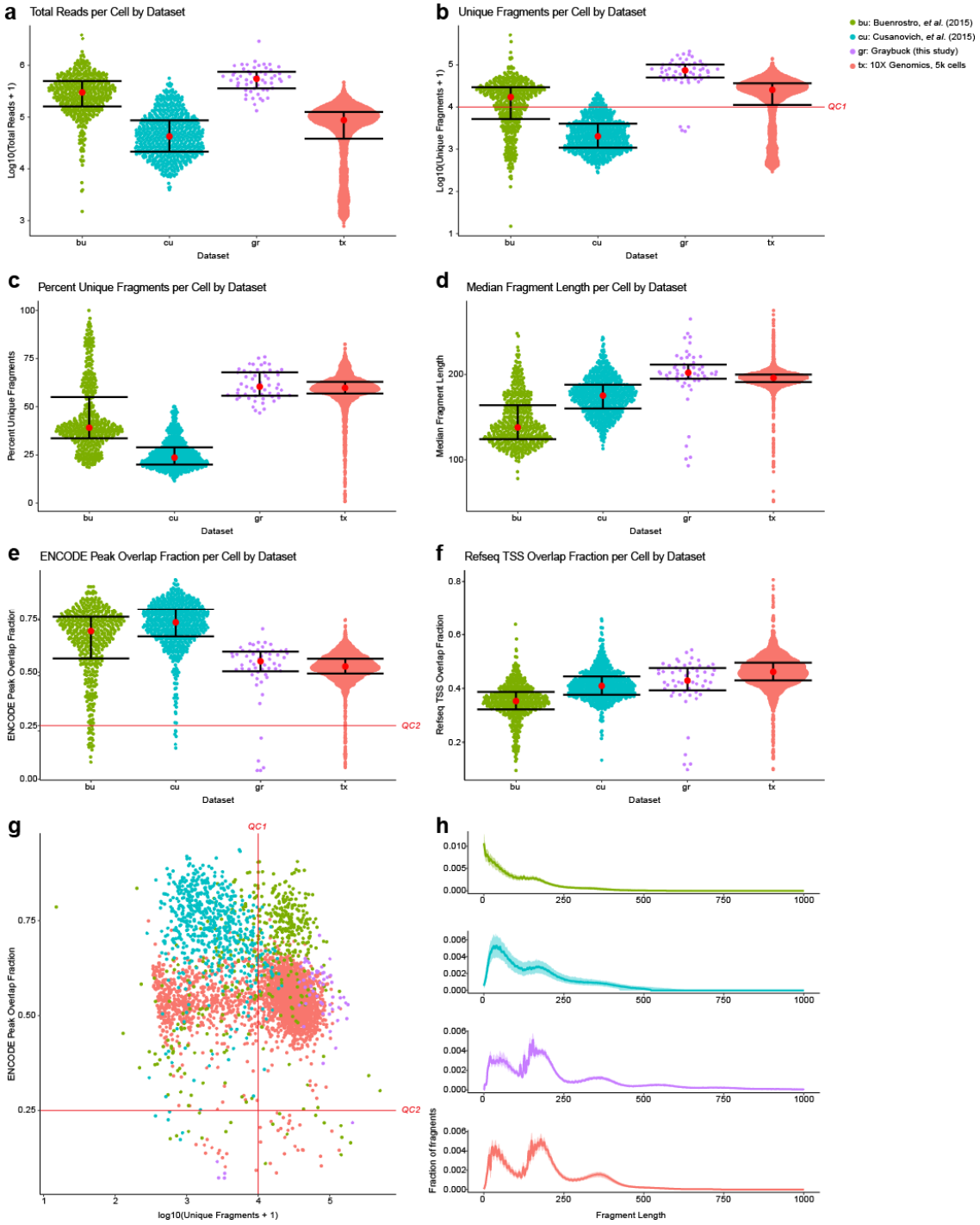
bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 3. Retrograde injections.

(a) Labeling of scATAC-seq cells in t-SNE space to highlight cells collected using Rbp4-Cre (Rbp4, light blue), which labels both L5 IT and L5 PT neurons; using retrograde labeling from the lateral posterior nucleus of the thalamus (LP) to VISp (green); or retrograde labeling from superior colliculus (SC) to VISp (brown). **(b)** Native fluorescence images of a mouse dissected for scATAC-seq sample collection from a LP retrograde experiment. **(c)** Native fluorescence image of a ouse dissected for scATAC-seq sample collection from a SC retrograde experiment. **(d)** As for (a), but highlighting cells labeled using Cux2-CreERT2 (Cux2, orange), which labels both L2/3 IT and L4 IT neurons, and cells with retrograde labeling from VISp in the contralateral hemisphere (brown). **(e)** Native fluorescence images of a mouse dissected for scATAC-seq sample collection from a VISp-contralateral retrograde experiment. In this example, cells were collected only form the hemisphere opposite the single-hemisphere injection site. The lower panel shows a closer view of the collection site.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Fig 4

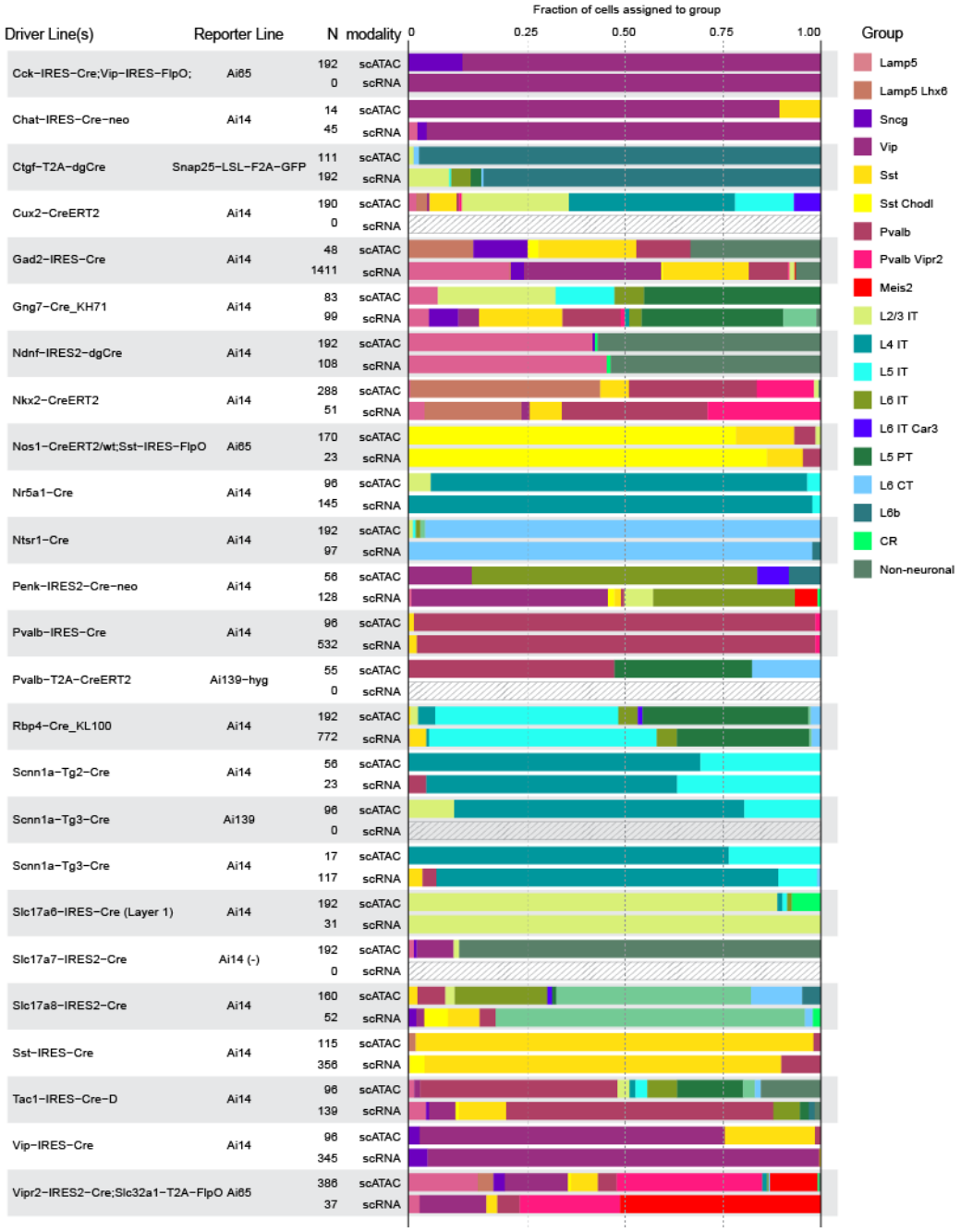


bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 4. Gm12878 platform comparisons. Comparison of FACS-sorted scATAC-seq libraries to those previously generated using Fluidigm C1 (Buenrostro, et al. 2015) or combinatorial indexing (Cusanovich, et al. 2015 and 10X Genomics) for which data using the common cell line of human Gm12878 cells is available. To use in these comparisons, we generated scATAC-seq data using our FACS-based method for 60 Gm12878 cells. For each published dataset, raw data was obtained from GEO or the 10X Genomics website, and analyzed using the same methods. Colors and abbreviations at the top-right are used throughout the plots: bu, Buenrostro, et al. (2015); cu, Cusanovich, et al. (2015); gr, Graybuck, et al. (This study); and tx, 10X Genomics. **(a)** Beeswarm plots showing the total reads per cell. For all similar plots (a-f), each point represents a single cell, with bars showing the 25th and 75th percentiles, and central red dots showing the median values. **(b)** Number of uniquely mapped fragments per cell for each dataset. **(c)** Percent uniquely mapped fragments for each dataset. **(d)** Median fragment length for each dataset. **(e)** Fraction of reads overlapping ENCODE DNase-seq data generated for Gm12878 cells. **(f)** Fraction of reads overlapping RefSeq TSS regions (TSS +/- 5kb) for each dataset. **(g)** Two-axis QC criteria plot, as in Figure 2, showing the QC1 and QC2 cutoffs used for mouse cortical scATAC-seq data. **(h)** Aggregate fragment length frequency plots. Fragment length is shown on the x-axis, and the fraction of reads with fragments of each bp size was calculated for each sample in each dataset. For this analysis, the median fraction at each fragment size is shown as a solid line, with 25th and 75th percentiles shown in the shaded regions.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Fig 5

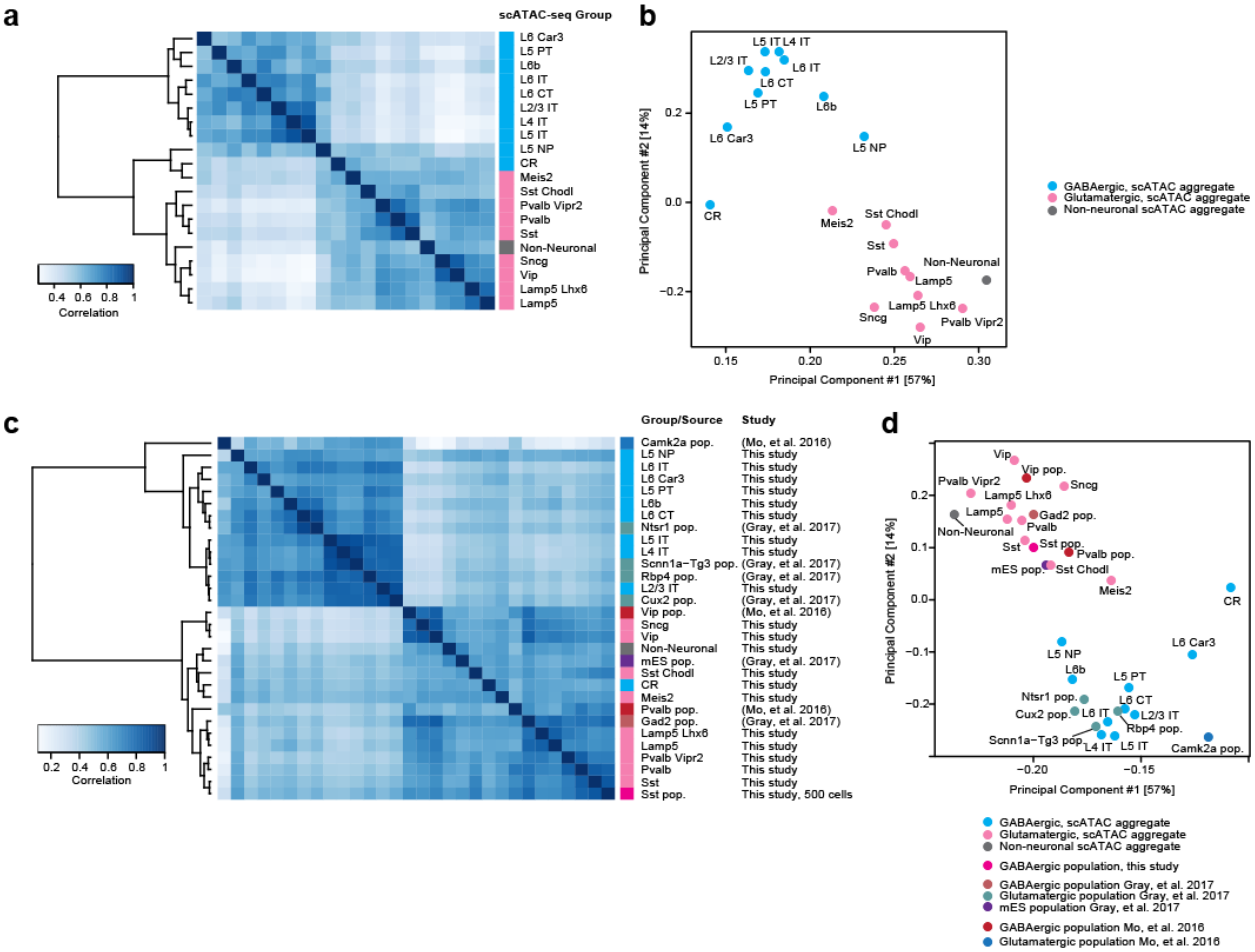


bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 5. Driver line comparison to scRNA-seq. To assess the accuracy of out scATAC-seq mapping to transcriptomic subclasses, we examined driver lines for which we collected both scATAC-seq and scRNA-seq data (the latter from Tasic, et al. (2018)). For each combination of driver and reporter mouse line, we show the number of samples obtained for each method (scATAC-seq or scRNA-seq), and plot the fraction of cells assigned to each group used in Figure 2 as colored bars for each method. Differences in proportions may be due to differences in dissection and region sampling strategies.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

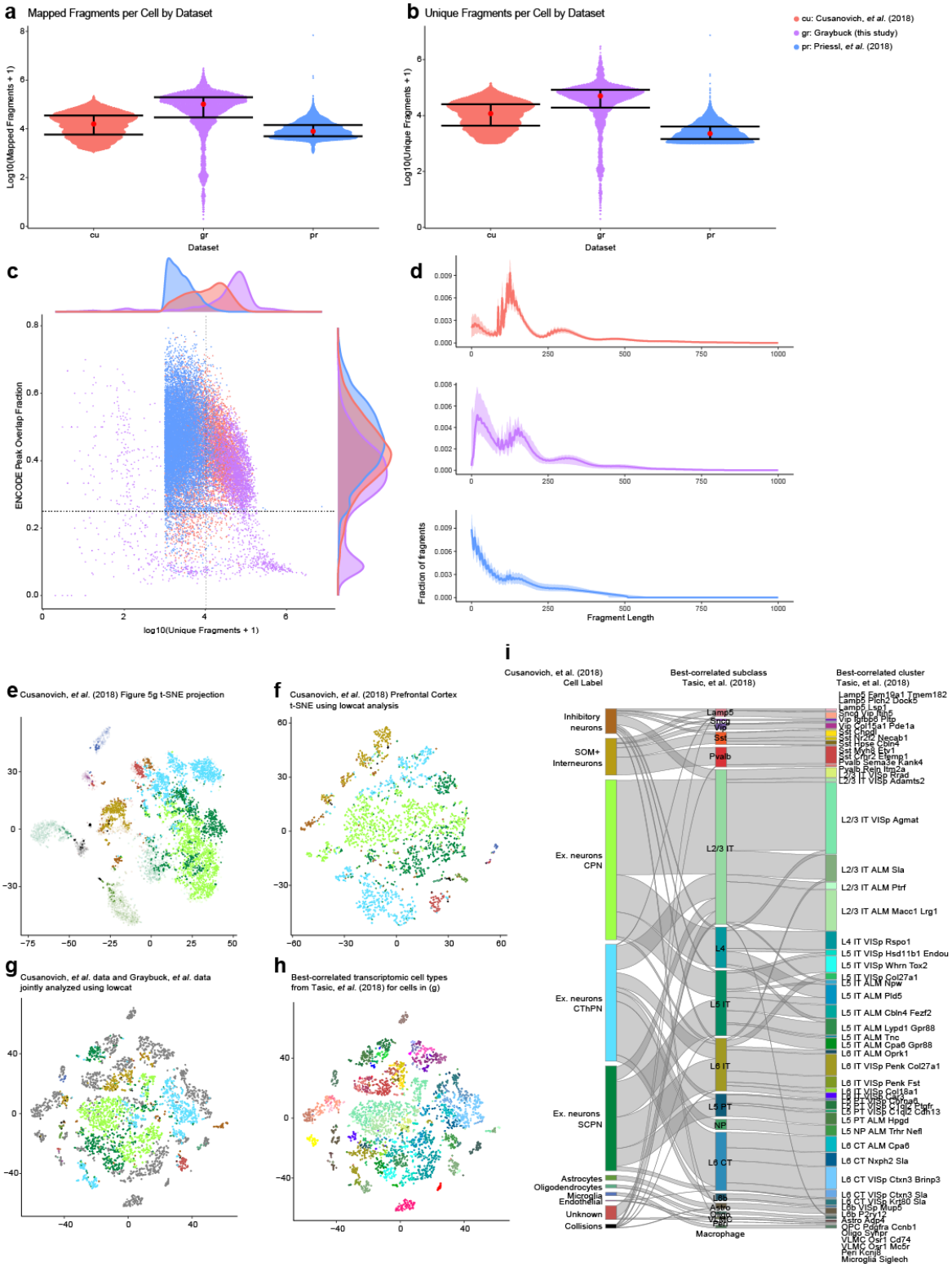
Supplementary Figure 6



Supplementary Figure 6. Diffbind population comparisons. We compared aggregated scATAC-seq datasets to each other and to previously published neuronal ATAC-seq populations using the DiffBind package for R. **(a)** Comparisons of aggregated scATAC-seq datasets shows expected grouping of GABAergic and Glutamatergic cell classes. The heatmap shows weighted pairwise correlations between each pair of aggregated scATAC-seq groups. **(b)** PCA of the aggregated samples showin in (a). **(c)** Comparisons of aggregated scATAC-seq datasets and population ATAC-seq datasets generated by Mo, *et al*. (2016), Gray, *et al*. (2017), or for this study (Sst). **(d)** PCA of the datasets shown in (c).

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 7

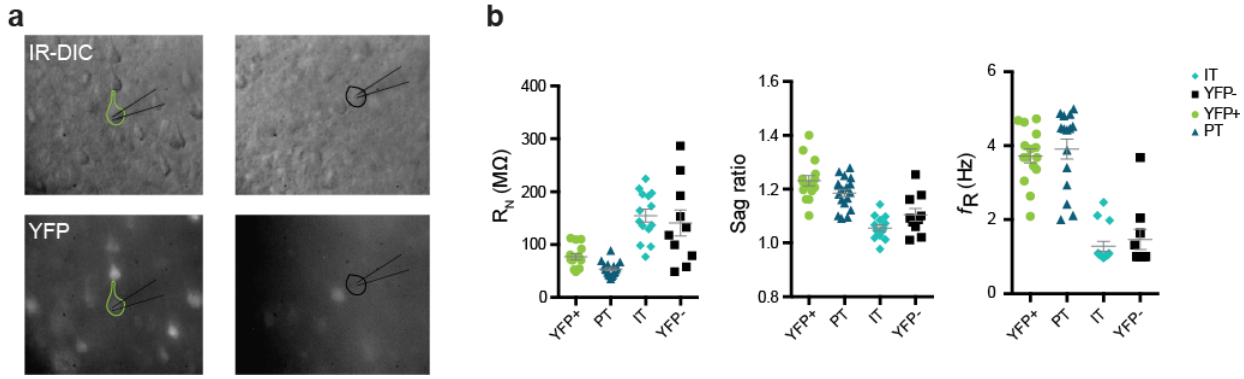


bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 7. Comparisons to neuronal scATAC-seq studies. Comparison of previously published barcode-based scATAC-seq libraries from adult mouse cortex to FACS-sorted scATAC-seq in the present study. Data for the two previous studies, Cusanovich, et al. (2018) (P56 prefrontal cortex sample) and Priessl, et al. (2018) (P56 cortex sample) were obtained from GEO, and were aligned to the mm10 genome using bowtie with the same settings as those for FACS-sorted samples. These samples were then demultiplexed using an R script, and barcodes with >1,000 mapped reads were retained for analysis. Colors and abbreviations at the top-right are used in panels a-d: cu, Cusanovich, et al. (2018); gr, Graybuck, et al. (This study); and pr, Priessl, et al. (2018). **(a)** Beeswarm plots showing the total aligned reads per cell. For all similar plots (a-b), each point represents a single cell, with bars showing the 25th and 75th percentiles, and central red dots showing the median values. **(b)** Beeswarm plot showing the number of uniquely aligned fragments per cell. **(c)** Two-axis QC criteria plot, as in Figure 2, showing the QC1 and QC2 cutoffs used for mouse cortical scATAC-seq data. Marginal plots show the density distributions of each dataset along the two axes. The majority of the samples from the Priessl, et al. (2018) dataset fell below the QC1 criteria of > 10,000 unique fragments, though scoring of ENCODE peak overlaps remained high. **(d)** Aggregate fragment length frequency plots. Fragment length is shown on the x-axis, and the fraction of reads with fragments of each bp size was calculated for each cell in each dataset. For this analysis, the median fraction at each fragment size is shown as a solid line, with 25th and 75th percentiles shown in the shaded regions. **(e-i)** Analysis of high-quality samples from Cusanovich, et al. (2018) using the same criteria and methods used for the Graybuck, et al. dataset. **(e)** t-SNE coordinates presented in Cusanovich, et al. (2018) for prefrontal cortex, with colors showing the major cell types identified in Cusanovich, et al. (2018). Solid points passed Graybuck, et al. QC criteria and were used for subsequent analysis. **(f)** t-SNE projection of Cusanovich, et al. samples processed using fragment-overlap Jaccard similarity analysis as used for Graybuck, et al. cells, and colored according to their original labels. **(g)** t-SNE projection of Cusanovich, et al. and Graybuck, et al. datasets analyzed together. Graybuck, et al. cells are colored gray. **(h)** Mapping of the co-analyzed dataset to scRNA-seq cell types identified in Tasic, et al. (2018). **(i)** River plot shows the mapping of scATAC-seq samples from Cusanovich, et al. on to scRNA-seq subclasses (center) and types (right) from Tasic, et al. (2018). The colors of the original cell labels correspond to the points in plots (e-f).

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

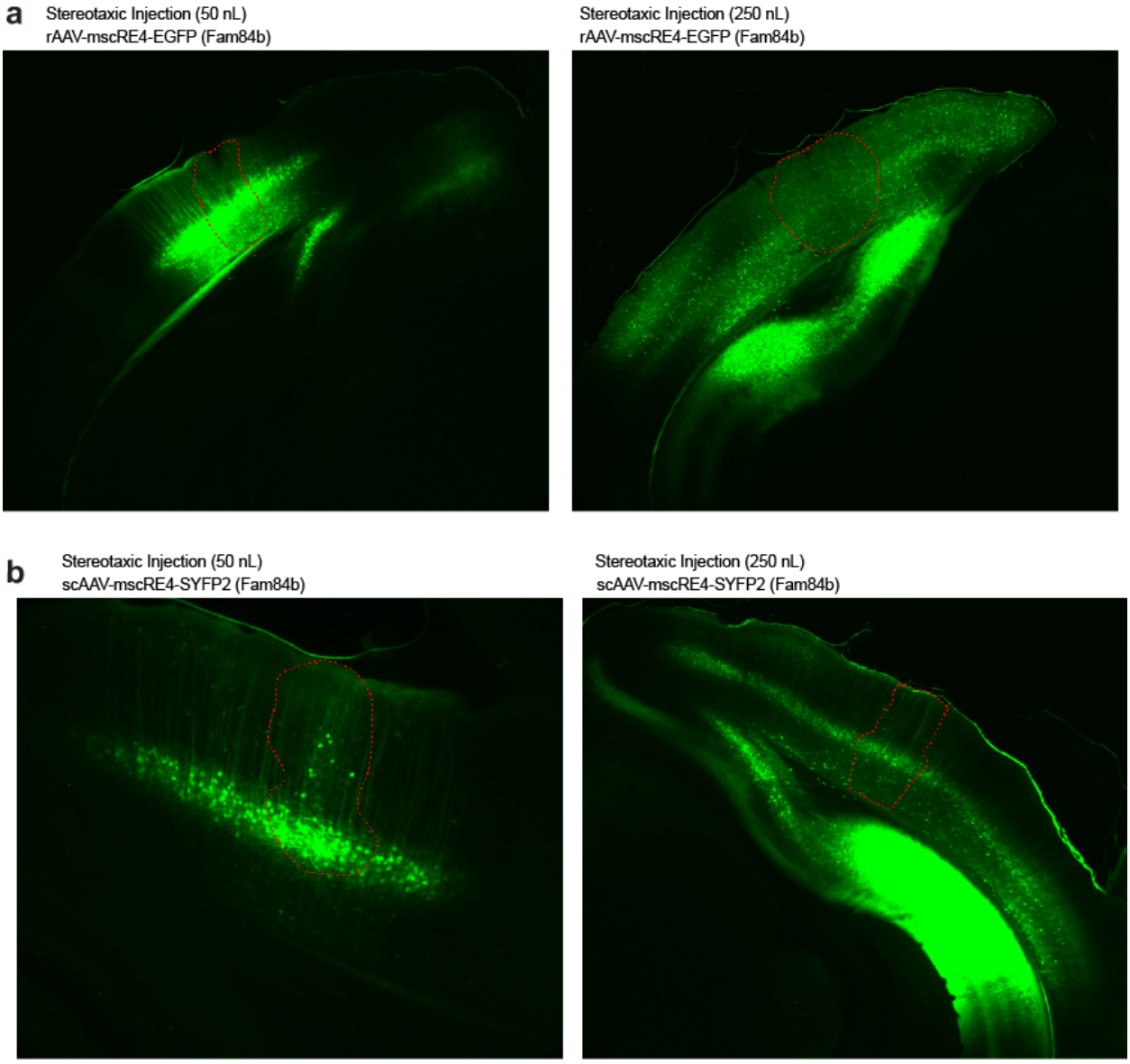
Supplementary Figure 8



Supplementary Figure 8. Additional eletrophysiological characteristics of mscRE4-SYFP2 labeled cells. (a) Microscopy of example cells characterized by patch electrophysiology. Left, a SYFP2-positive cell; right, a SFYP2-negative cell. **(b)** Input resistance, sag ratio, and resonance frequency for the four experimental conditions presented in Figure 3c-d.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

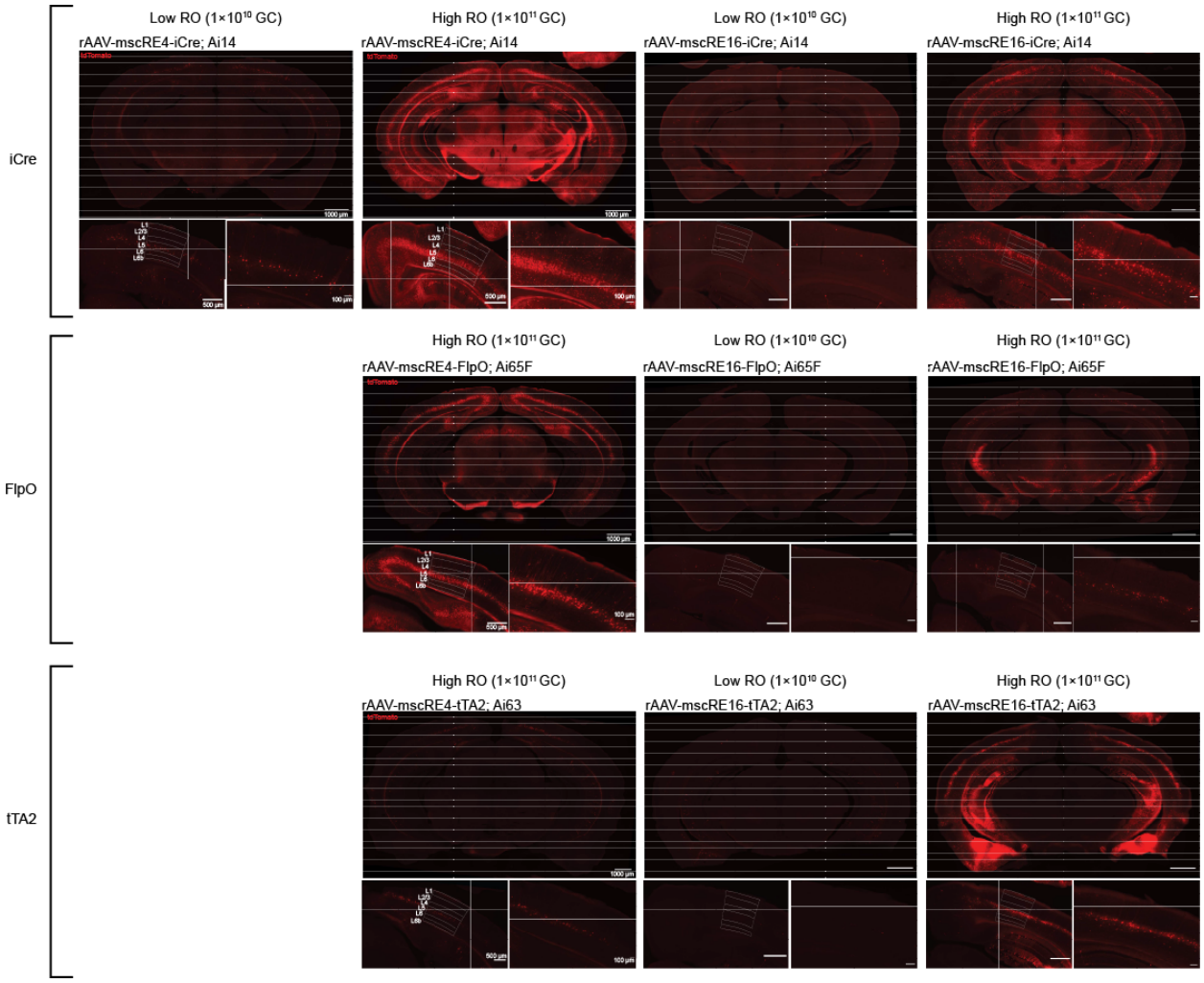
Supplementary Figure 9



Supplementary Figure 9. Stereotaxic labeling using enhancer-driven viruses. (a) Native fluorescence imaging of animals with stereotaxic injection of mscRE4-EGFP in primary visual cortex. Enhancer-driven viruses were co-injected with a constitutive dTomato virus, rAAVDJ-EF1a-dTomato at 0.1X of the volumes of the mscRE viruses, to provide injection site location (red outlines). **(b)** Native fluorescence imaging of animals with stereotaxic injection of mscRE4-SYFP2 into primary visual cortex at the indicated volumes.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

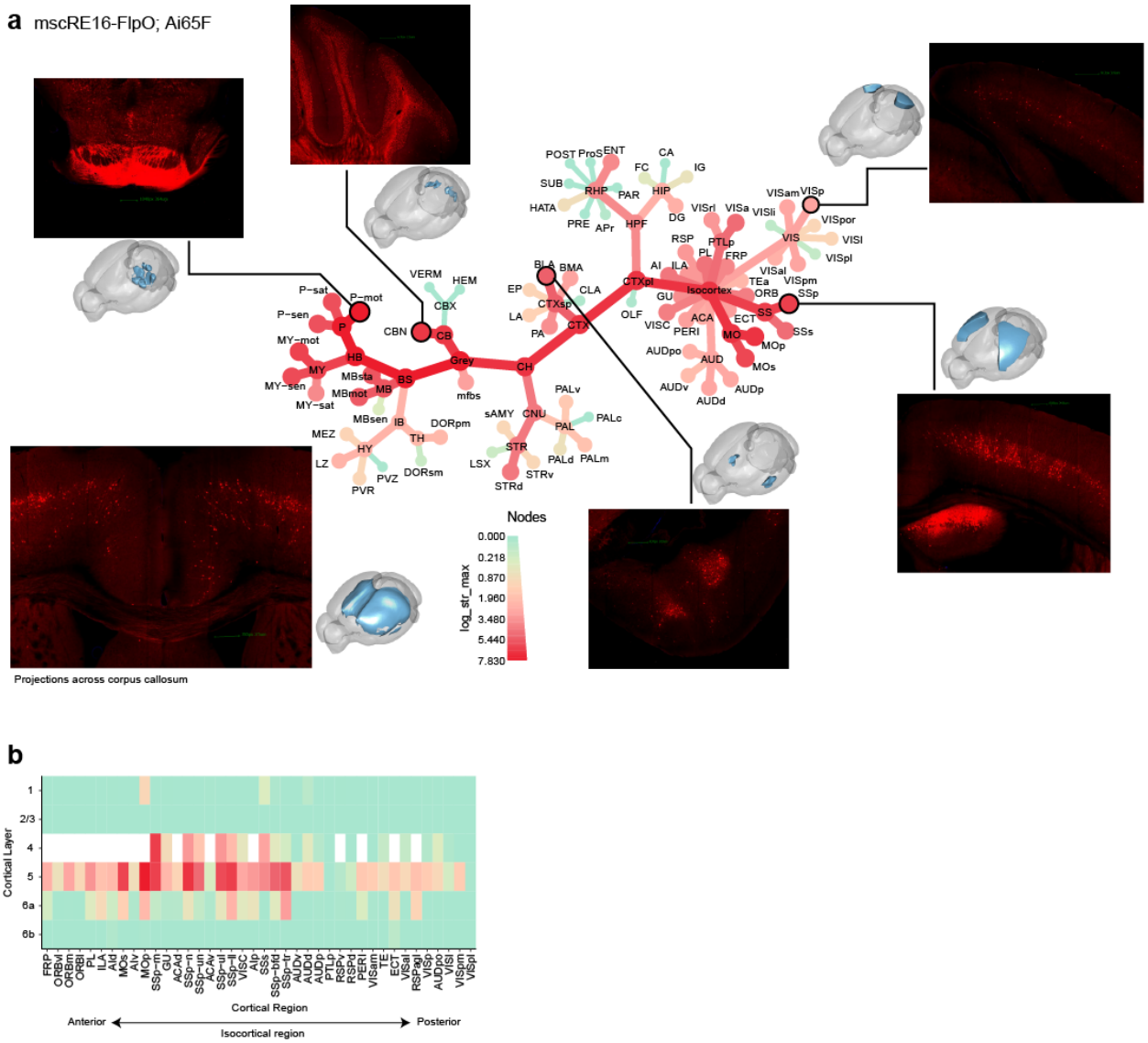
Supplementary Figure 10



Supplementary Figure 10. Retro-orbital mscRE driver screening at multiple titers. Native fluorescence images for reporter mice retro-orbitally (RO) injected with enhancer-driven recombinase viruses at two titers: Low RO, 1×10^{10} genome copies, GC; High RO, 1×10^{11} GC.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Figure 11



Supplementary Figure 11. Whole-brain characterization of mscRE16-FlpO. (a) TissueCyte imaging of an mscRE16-FlpO;Ai65F mouse 2 weeks after retro-orbital injection was registered to the Allen Institute Common Coordinate Framework, and each structure in the adult mouse structural ontology was scored. As for Figure 5, this panel provides a high-level overview of cell labeling throughout the structural ontology. The size and color of nodes represents the maximum signal found among all children of the nodes, which allows one to follow the tree to the source of high signal within each structure. Insets display selected regions of high or specific signal. The inset at the bottom-left shows projection of IT neurons across the corpus callosum. **(b)** Layer quantification for the same TissueCyte image registered to the CCF for all isocortical regions. Agranular regions that lack L4 are shown with a white box in the L4 row. All acronyms correspond to the Allen Institute for Brain Science Adult Mouse 3D atlas.

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

List of Supplementary Tables

Supp Table 1. Mouse lines

Supp Table 2. Driver/reporter donor metadata

Supp Table 3. Stereotaxic and retro-orbital injection donor metadata

Supp Table 4. scATAC-seq cell alignment and QC statistics

Supp Table 5. mscRE locations and cloning primers

Supp Table 6. scATAC-seq clustering and cell type mapping results

Supp Table 7. Virus descriptions and sources

Supp Table 8. scRNA-seq sample alignment and mapping statistics

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgments

We could not have performed this study without the support of the Allen Institute Animal Care Team for mouse husbandry, the Allen Institute Transgenic Colony Management team for colony management, the Allen Institute Laboratory Animal Services team for acute cage preps; Tamara Caspar, Kirsten Chrichton, Matthew Kroll, Josef Sulc, and Herman Tung for tissue processing; Nadiya Shapovalova, Daniel Hirschstein, and Susan Bort for FACS sorting; and Darren Bertagnolli, Michael Tieu, Delissa McMillen, Thanh Pham, Christine Rimorin, Katelyn Ward, Alexandra Glandon, and Amy Torkelson for scRNA-seq processing. The authors also thank Andrew Hill and Darren Cusanovich for assistance in accessing and reusing data published in Cusanovich, *et al*. (2018). The project described was supported by award number 1R01DA036909-01 REVISED from the National Institute on Drug Abuse. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health and the National Institute on Drug Abuse. The authors thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support.

References

1. Tasic, B. Single cell transcriptomics in neuroscience: cell classification and beyond. *Curr. Opin. Neurobiol.* **50,** 242–249 (2018).
2. Zeng, H. & Sanes, J. R. Neuronal cell-type classification: Challenges, opportunities and the path forward. *Nat. Rev. Neurosci.* **18,** 530–546 (2017).
3. Daigle, T. L. *et al.* A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. *Cell* **174,** 465–480.e22 (2018).
4. Taniguchi, H. *et al.* A Resource of Cre Driver Lines for Genetic Targeting of GABAergic Neurons in Cerebral Cortex. *Neuron* **71,** 995–1013 (2011).
5. Gong, S. *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27,** 9817–9823 (2007).
6. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563,** 72–78 (2018).
7. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19,** 335–346 (2016).
8. Gouwens, N. W. *et al.* Classification of electrophysiological and morphological types in mouse visual cortex. *bioRxiv* 368456 (2018). doi:10.1101/368456
9. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19,** (2016).
10. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523,** 486–90 (2015).
11. Cusanovich, D. A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science (80-.).* **348,** 910–914 (2015).
12. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515,** 355–64 (2014).
13. Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162,** 184–197 (2015).
14. Harris, J. A. *et al.* The organization of intracortical connections by layer and cell class in the mouse brain. *bioRxiv* 292961 (2018). doi:10.1101/292961
15. Chan, K. Y. *et al.* Engineered adeno-associated viruses for efficient and noninvasive gene delivery throughout the central and peripheral nervous systems. *Nat. Publ. Gr.* 1–30

bioRxiv preprint doi: https://doi.org/10.1101/525014; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(2017). doi:10.1038/nn.4593

16. Baker, A. *et al.* Specialized Subpopulations of Deep-Layer Pyramidal Neurons in the Neocortex: Bridging Cellular Properties to Functional Consequences. *J. Neurosci.* **38,** 0150-18 (2018).
17. Madisen, L. *et al.* Transgenic Mice for Intersectional Targeting of Neural Sensors and Effectors with High Specificity and Performance. *Neuron* **85,** 942–958 (2015).
18. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13,** 133–40 (2010).
19. Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339,** 1–4 (2013).
20. Shen, S. Q. *et al.* Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res.* **26,** 238–255 (2016).
21. Blankvoort, S., Witter, M. P., Noonan, J., Cotney, J. & Kentros, C. Marked Diversity of Unique Cortical Enhancers Enables Neuron-Specific Tools by Enhancer-Driven Gene Expression. *Curr. Biol.* (2018). doi:10.1016/j.cub.2018.05.015
22. Hartl, D., Krebs, A. R., Jüttner, J., Roska, B. & Schübeler, D. Cis-regulatory landscapes of four cell types of the retina. *Nucleic Acids Res.* **45,** 11607–11621 (2017).
23. Kishi, J. Y. *et al.* SABER enables highly multiplexed and amplified detection of DNA and RNA in cells and tissues. *bioRxiv* 401810 (2018). doi:10.1101/401810
24. Salva, M. Z. *et al.* Design of tissue-specific regulatory cassettes for high-level rAAV-mediated expression in skeletal and cardiac muscle. *Mol. Ther.* **15,** 320–329 (2007).
25. Hnasko, T. S. *et al.* Cre recombinase-mediated restoration of nigrostriatal dopamine in dopamine-deficient mice reverses hypophagia and bradykinesia. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 8858–8863 (2006).
26. Paxinos, G. & Franklin, K. B. J. *The Mouse Brain in Stereotaxic Coordinates Compact Third Edition*. *Academic Press* (Academic Press, 2008).
27. Gray, L. T. *et al.* Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex. *Elife* 1–30 (2017). doi:10.7554/eLife.21883
28. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38,** 576–589 (2010).
29. Mo, A. *et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86,** 1369–1384 (2015).
30. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
31. Ting, J. T., Daigle, T. L., Chen, Q. & Feng, G. Acute brain sclice methods for adult and aging animals: application of targeted patch clamp analysis and optogenetics. *Methods Mol. Biol.* **1183,** 221–242 (2014).
32. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508,** 207–214 (2014).
33. Foster, Z. S. L., Sharpton, T. & Grunwald, N. J. *MetacodeR* : An R package for manipulation and heat tree visualization of community taxonomic data from metabarcoding. *bioRxiv* 071019 (2016). doi:10.1101/071019