

Supplemental information

**Enhancer viruses for combinatorial
cell-subclass-specific labeling**

Lucas T. Graybuck, Tanya L. Daigle, Adriana E. Sedeño-Cortés, Miranda Walker, Brian Kalmbach, Garreck H. Lenz, Elyse Morin, Thuc Nghi Nguyen, Emma Garren, Jacqueline L. Bendrick, Tae Kyung Kim, Thomas Zhou, Marty Mortrud, Shenqin Yao, La' Akea Siverts, Rachael Larsen, Bryan B. Gore, Eric R. Szelenyi, Cameron Trader, Pooja Balaram, Cindy T.J. van Velthoven, Megan Chiang, John K. Mich, Nick Dee, Jeff Goldy, Ali H. Cetin, Kimberly Smith, Sharon W. Way, Luke Esposito, Zizhen Yao, Viviana Gradinaru, Susan M. Sunkin, Ed Lein, Boaz P. Levi, Jonathan T. Ting, Hongkui Zeng, and Bosiljka Tasic

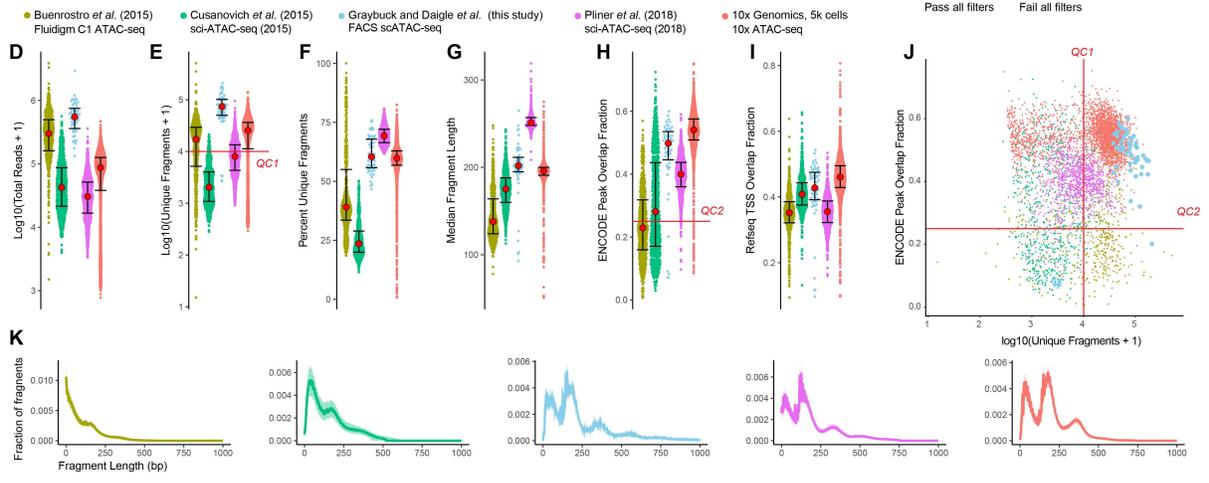
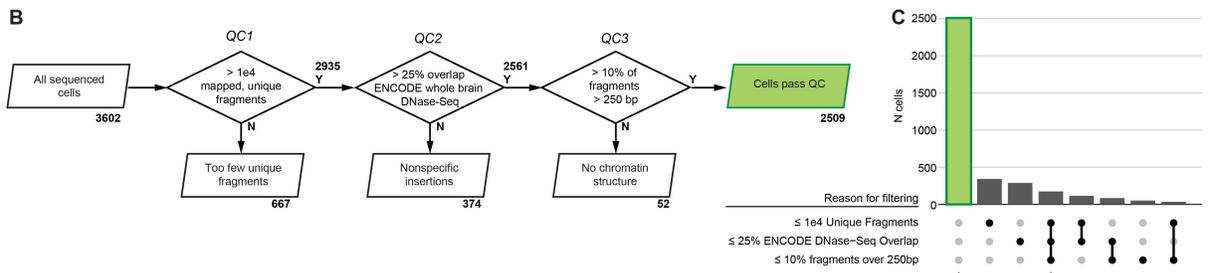
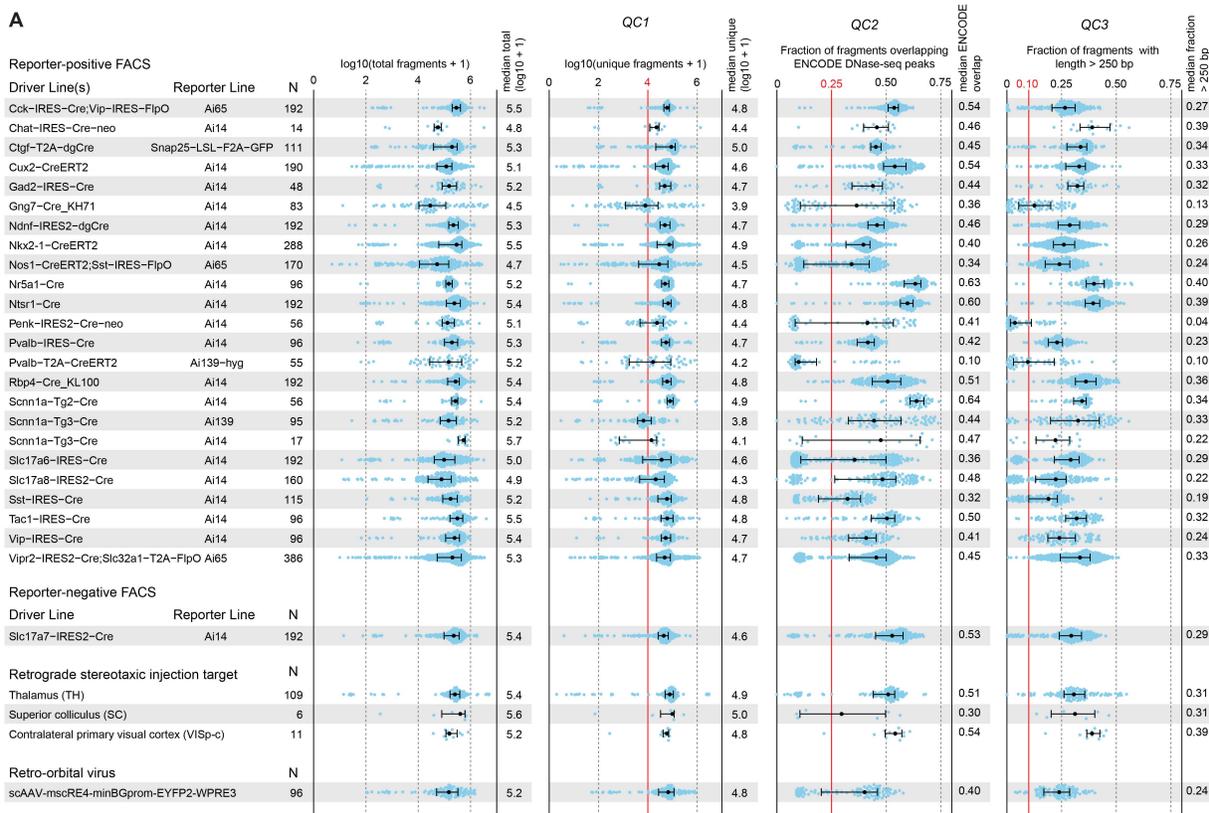


Figure S1. Cell sources, QC Statistics, and comparisons to previously published methods, related to Figure 2. (A) Fragment number and quality for each cell source in the scATAC-seq dataset. See **Table S1** for mouse line details. N, number of cells collected. ‘Total fragments’ is the count of all sequenced fragments for each cell. ‘Unique fragments’ is the number of uniquely-mapped fragments and was used for the first QC cutoff of (> 10,000 unique fragments; QC1). Fraction of fragments overlapping ENCODE DNase-seq peaks was computed for uniquely mapped fragments and was used for the second QC cutoff (> 0.25; QC2). Fraction of fragments with length > 250 bp was computed for unique fragments and was used for the third QC cutoff (> 0.1; QC3). (B) Flow chart showing how many samples were sequentially filtered by these three QC criteria. (C) Set barplot showing how many cells were flagged with each combination of QC criteria. (D-K) Comparison of FACS scATAC-seq libraries to those previously generated using Fluidigm C1 (Buenrostro et al., 2015), sci-ATAC-seq (Cusanovich et al., 2015; Pliner et al., 2018), or droplet-based indexing (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 and was aligned and analyzed using the same methods. For 10x Genomics, aligned fragment locations and metadata were obtained from the 10x genomics website for the “5k 1:1 mixture of fresh frozen human (GM12878) and mouse (A20) cells” dataset. Only samples labeled as GM12878 in sample metadata for each dataset were used for analysis. The color key above panels (D-I) are used throughout the plots (D) 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. (E) Number of uniquely mapped fragments per cell for each dataset. (F) Percent uniquely mapped fragments for each dataset. A high fraction of unique fragments (as in Pliner *et al.*) suggests that deeper sequencing will yield additional useful data. (G) Median fragment length for each dataset. (H) Fraction of reads overlapping ENCODE DNase-seq data generated for GM12878 cells. (I) Fraction of reads overlapping RefSeq TSS regions (TSS ± 5kb) for each dataset. (J) Two-axis QC criteria plot, as in **Fig. 2**, showing the QC1 and QC2 cutoffs used for mouse cortical scATAC-seq data. (K) 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 as shaded regions.

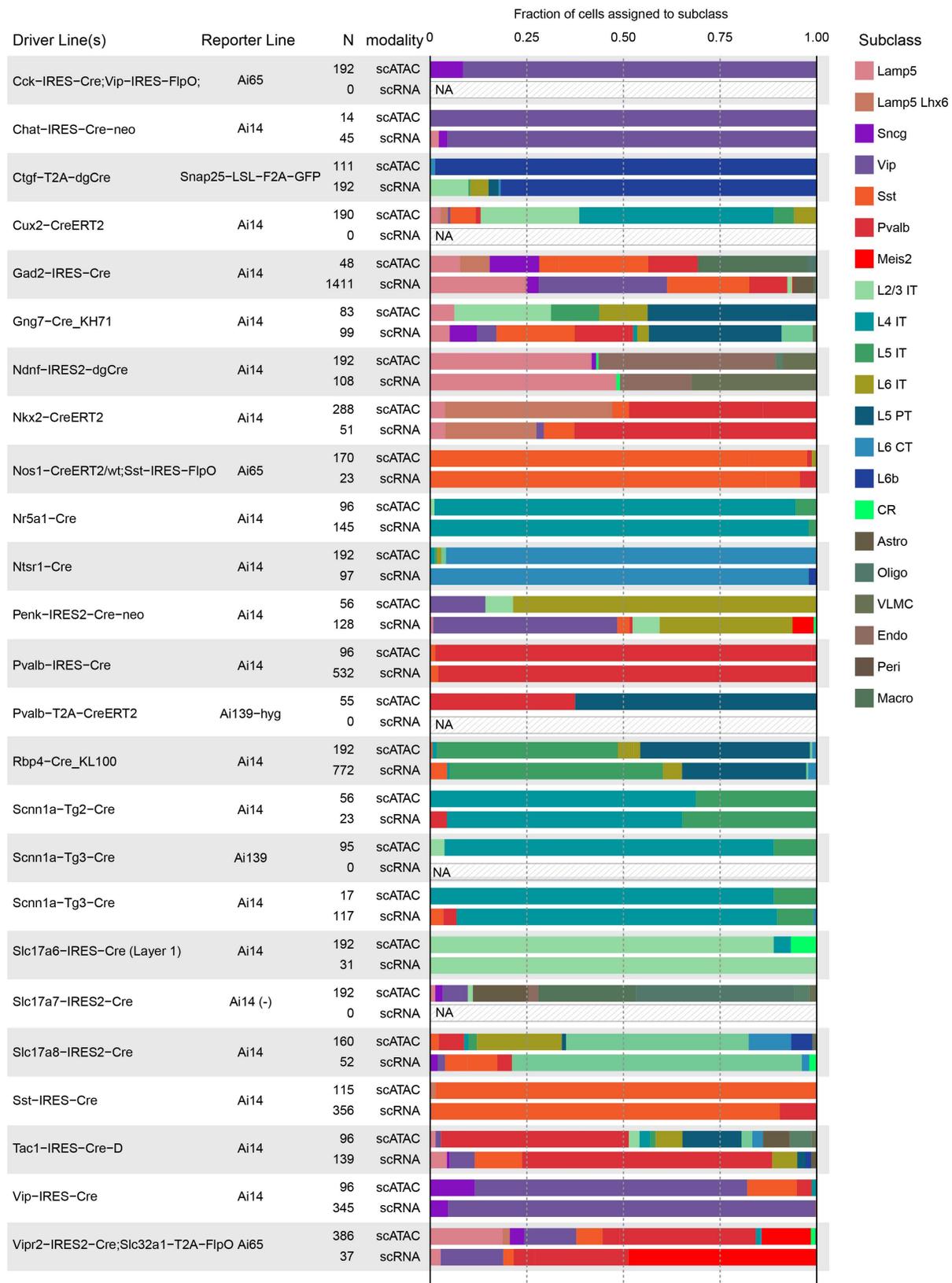


Figure S2. Comparison of scATAC-seq with scRNA-seq for same transgenic lines, related to Figure 2. To assess the accuracy of our scATAC-seq mapping to transcriptomic subclasses, we examined driver lines for which we collected both scATAC-seq and scRNA-seq data (the latter from Tasic *et al.* (Tasic et 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 **Fig. 2** as colored bars for each method. Differences in proportions may be due to differences in dissection and region sampling strategies. Gray hashed bars indicate that no cells were assessed by scRNA-seq for specific driver and reporter line combinations.

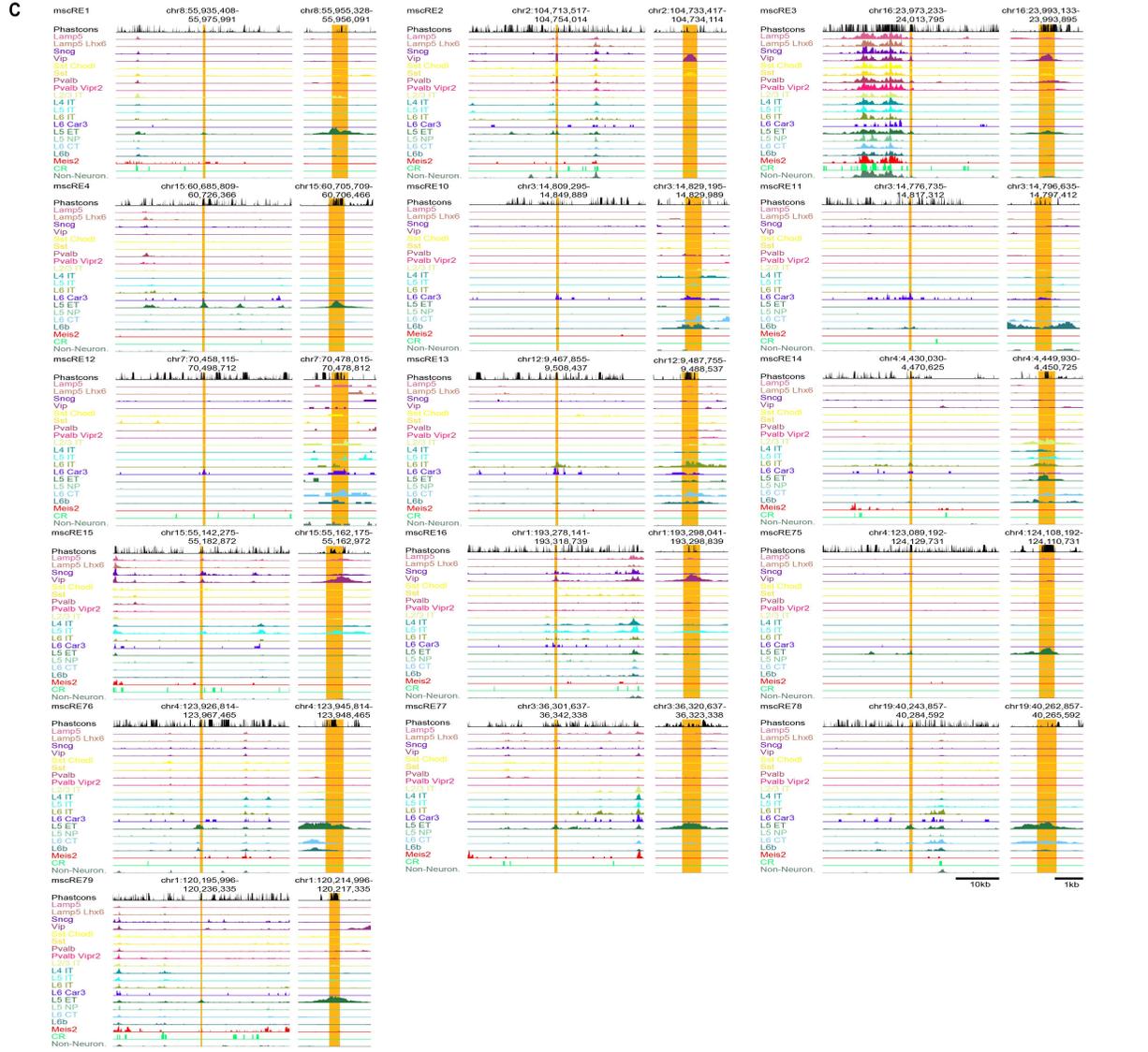
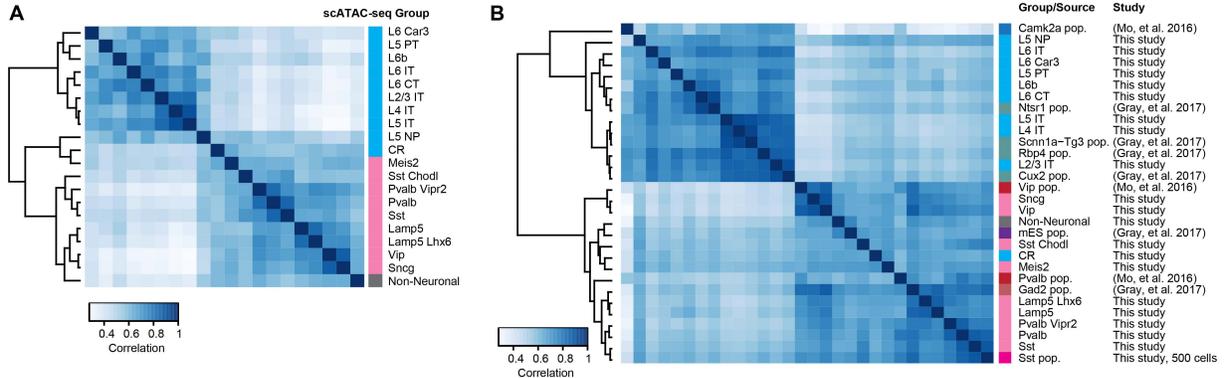
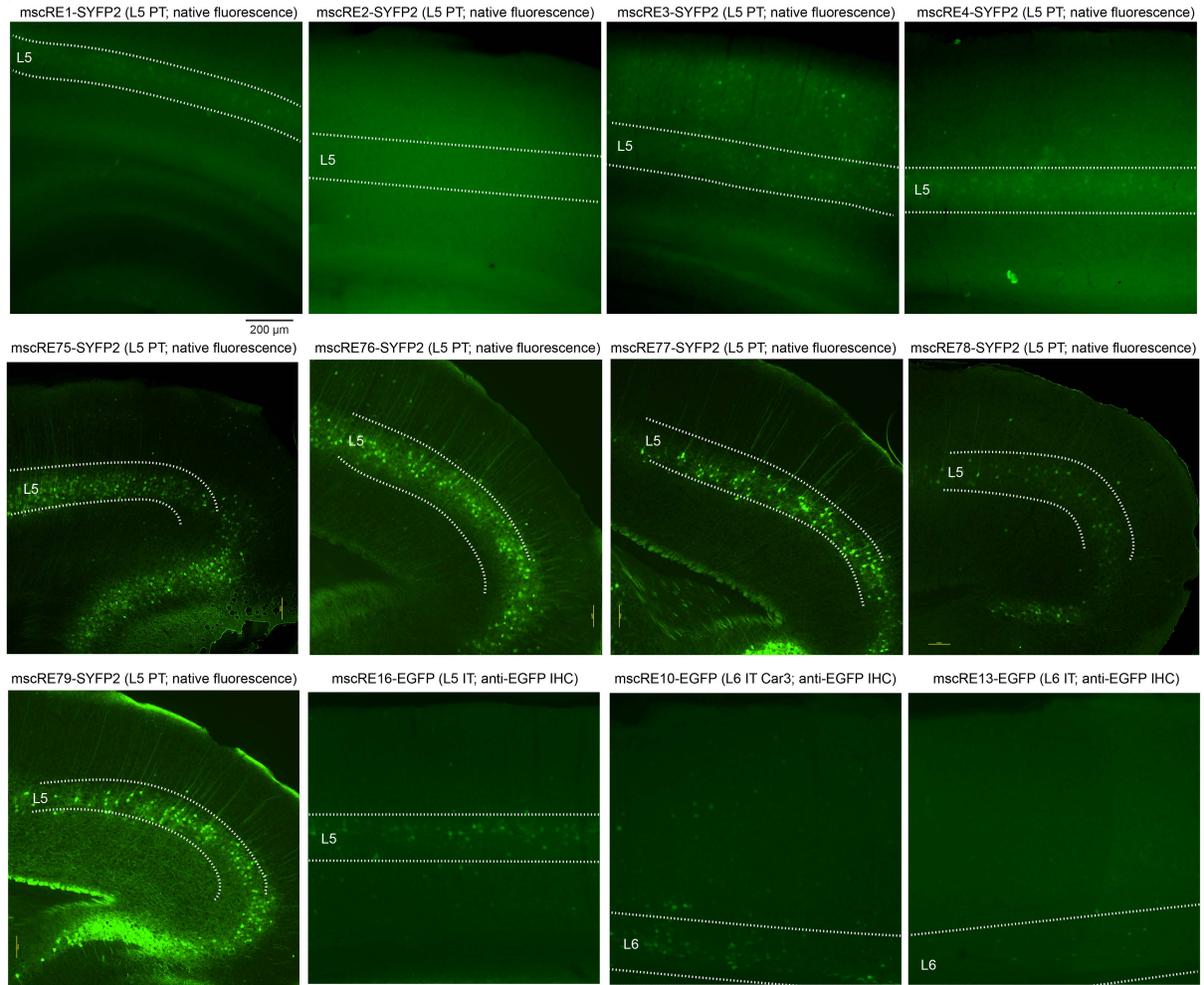


Figure S3. Comparison of our dataset with existing datasets and all mscREs (16 total) examined in this study, related to Figure 3. (A) Pairwise correlations between select cell classes, subclasses, and types defined by clustering scATAC-seq data in this study using the DiffBind package for R (Stark and Brown, 2011) (**STAR Methods**) show expected grouping of GABAergic and glutamatergic cell subclasses and types. The heatmap shows weighted pairwise correlations between each pair of aggregated scATAC-seq cell populations. (B) Approach in (A) was used to compare scATAC-seq clusters with population ATAC-seq datasets generated by Mo *et al.* (Mo *et al.*, 2015), Gray *et al.*, (Gray *et al.*, 2017), or in this study (Sst pop. – “pop.” indicates that the data were experimentally obtained from a population of cells, and not from single cells). Expected divisions and groupings are observed. (C) Chromatin accessibility in clusters based on single cell ATAC-seq data for select genomic regions containing mscREs defined and examined in this study.

A

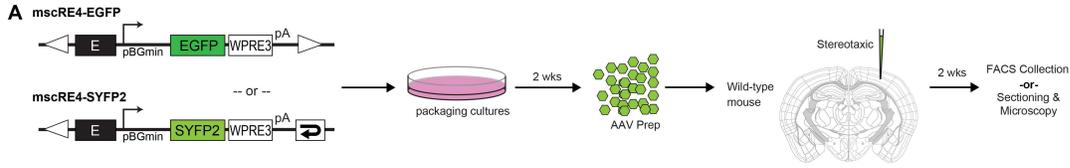


B

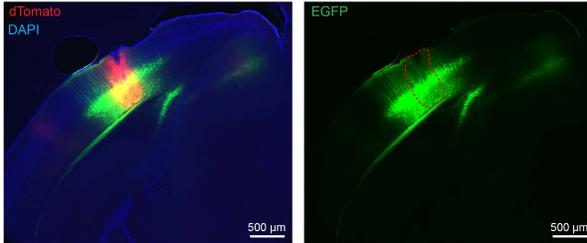
Number	Target cell subclass/type	mscRE	Shown in Fig 54A?	Desired layer enrichment?	Efficiency of desired labeling based on layer enrichment	Desired morphology observed?	Efficiency of desired labeling based on morphology	Transcriptionally confirmed (scRNA-seq and/or RNAscope)?	Efficiency of desired labeling based on transcriptional analysis	Efficiency of desired labeling based on modalities: layer enrichment (L), morphology (M) or transcriptional analysis (T)
1	L5 PT	mscRE1	Y	Y	7/9 (78%)	Y	7/9 (78%)	Y	2/2 (100%)	7/9 (78% for L, M)
2	L5 PT	mscRE2	Y	N		NE				
3	L5 PT	mscRE3	Y	N		NE				
4	L5 PT	mscRE4	Y	Y		Y				
5	L5 PT	mscRE75	Y	Y		Y				
6	L5 PT	mscRE76	Y	Y		Y				
7	L5 PT	mscRE77	Y	Y		Y				
8	L5 PT	mscRE78	Y	Y		Y				
9	L5 PT	mscRE79	Y	Y		Y				
10	L5 IT	mscRE15	N	N	1/2 (50%)	NE	NE	NE	1/1 (100%)	1/2 (50% for L, T)
11	L5 IT	mscRE16	Y	Y	NE	NE	Y			
12	L6 IT Car3	mscRE10	Y	Y	1/3 (33%)	NE	NE	N	0/1 (0%)	0/3 (0% for L,T)
13	L6 IT Car3	mscRE11	N	N	NE	NE	NE			
14	L6 IT Car3	mscRE12	N	N	NE	NE	NE			
15	L6 IT	mscRE13	Y	Y	1/2 (50%)	NE	NE	Y	1/1 (100%)	1/2 (50% for L, T)
16	L6 IT	mscRE14	N	N	NE	NE	NE			

Key
 Y: yes; N: no;
 NE: not evaluated

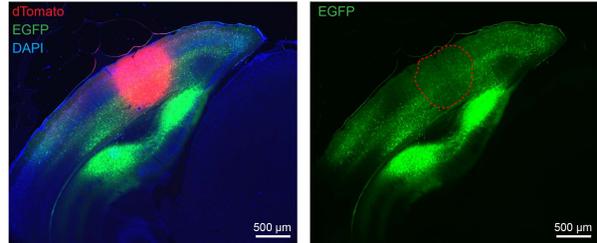
Figure S4. Initial screening of mscREs driving SYFP2 or EGFP, related to Figure 4. (A) Candidate enhancers were first tested by retro-orbital injection of direct fluorophore-expressing viruses. Epifluorescence images are shown of native fluorescence or antibody-enhanced (anti-GFP) fluorescence from VISp of C57BL/6J animals infected with virus. Dashed lines indicate the cortical layer expected to contain virus-labeled cells. mscRE1-4 were screened using self-complementary AAVs (scAAV) driving SYFP2 delivered at $1.0E+11$ GC into P62 mice, whereas mscRE75-79 were screened using rAAVs driving SYFP2 delivered at $6.0-9.0E+11$ GC into P36 mice. Enhancer viruses that exhibited no fluorescence (native or antibody-enhanced) in the initial screen were not included. **(B)** Summary of the results from the primary enhancer virus screen. Success rates were calculated based on the efficiency of labeling for each data modality (layer enrichment, morphology, and transcriptomics) and were grouped by subclass or type. The overall success rates are based on the labeling results obtained from either all three or at least two data modalities (yellow column). For more detailed descriptions of viral constructs, see **Table S5** and the **Key Resources Table**. For a general summary of all results, see **Table S8**.



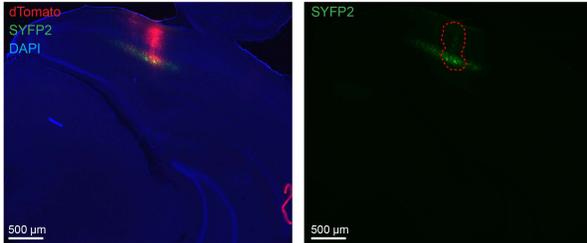
B Stereotaxic Injection (50 nL)
mscRE4-EGFP



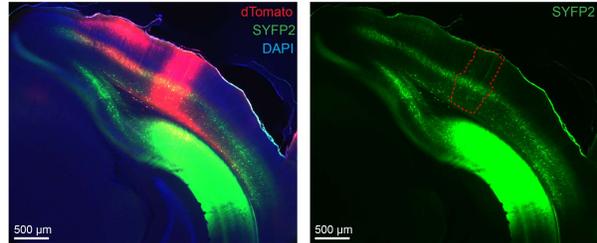
C Stereotaxic Injection (250 nL)
mscRE4-EGFP



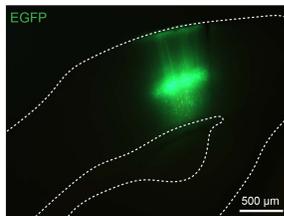
D Stereotaxic Injection (50 nL)
mscRE4-SYFP2



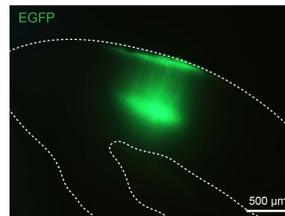
E Stereotaxic Injection (250 nL)
mscRE4-SYFP2



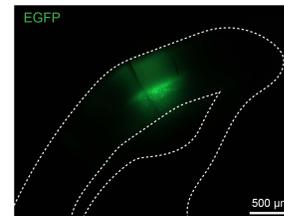
F Stereotaxic Injection (25 nL)
mscRE4-EGFP



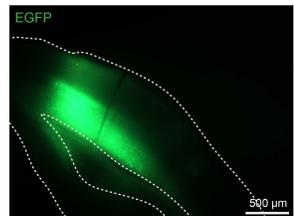
G Stereotaxic Injection (50 nL)
mscRE4-EGFP



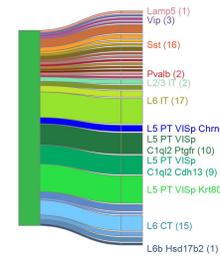
H Stereotaxic Injection (50 nL)
mscRE16-EGFP



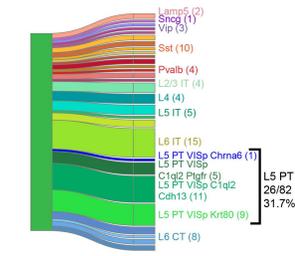
I Stereotaxic Injection (250 nL)
mscRE16-EGFP



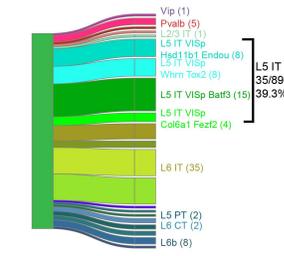
J Stereotaxic Injection (25 nL)
mscRE4-EGFP



K Stereotaxic Injection (50 nL)
mscRE4-EGFP



L Stereotaxic Injection (50 nL)
mscRE16-EGFP



M Stereotaxic Injection (250 nL)
mscRE16-EGFP

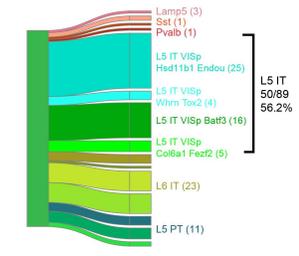


Figure S5. Enhancer virus labeling by stereotaxic injection, related to Figure 4. (A) For stereotaxic injection, enhancer-driven fluorophore viruses were generated by cloning mscRE sequences in rAAV (EGFP viruses) or scAAV (SYFP2 virus) constructs. After packaging, purification, and titering, viruses were stereotaxically targeted to VISp into wild-type mice. (B-C) Native fluorescence imaging of animals with stereotaxic injection of mscRE4-EGFP in VISp at two injection volumes (50 nL and 250 nL). Enhancer-driven viruses were co-injected with a constitutively expressed dTomato virus, rAAVDJ-EF1a-dTomato, at 0.1X of the volumes of the mscRE viruses, to provide injection site location (red outlines). (D-E) Native fluorescence imaging of animals with stereotaxic injection of mscRE4-SYFP2 into primary visual cortex at two injection volumes (50 nL and 250 nL). The same injection site labeling strategy used in (A) was also utilized for these experiments. (F-G) Native fluorescence imaging of stereotaxic injections into the primary visual cortex using mscRE4-EGFP at 25 nL and 50 nL volumes injected into the left and right hemisphere, respectively. (H-I) as in (E-F) using mscRE16-EGFP at 50 nL and 250 nL volumes. (J-M) scRNA-seq mapping results for cells collected from panels above each river plot (e.g., cells in (I) were collected from the injections matching the conditions shown in (E) scRNA-seq data were mapped to a VISp cell type reference from Tasic *et al.* (Tasic et al., 2018). using a centroid classifier, and only cells that mapped to a single cluster in ≥ 80 of 100 bootstrapped classifications were retained for figures: (J) n = 92 of 96; (K) n = 82 of 96; (L) n = 89 of 96; and (M) n = 89 of 96.

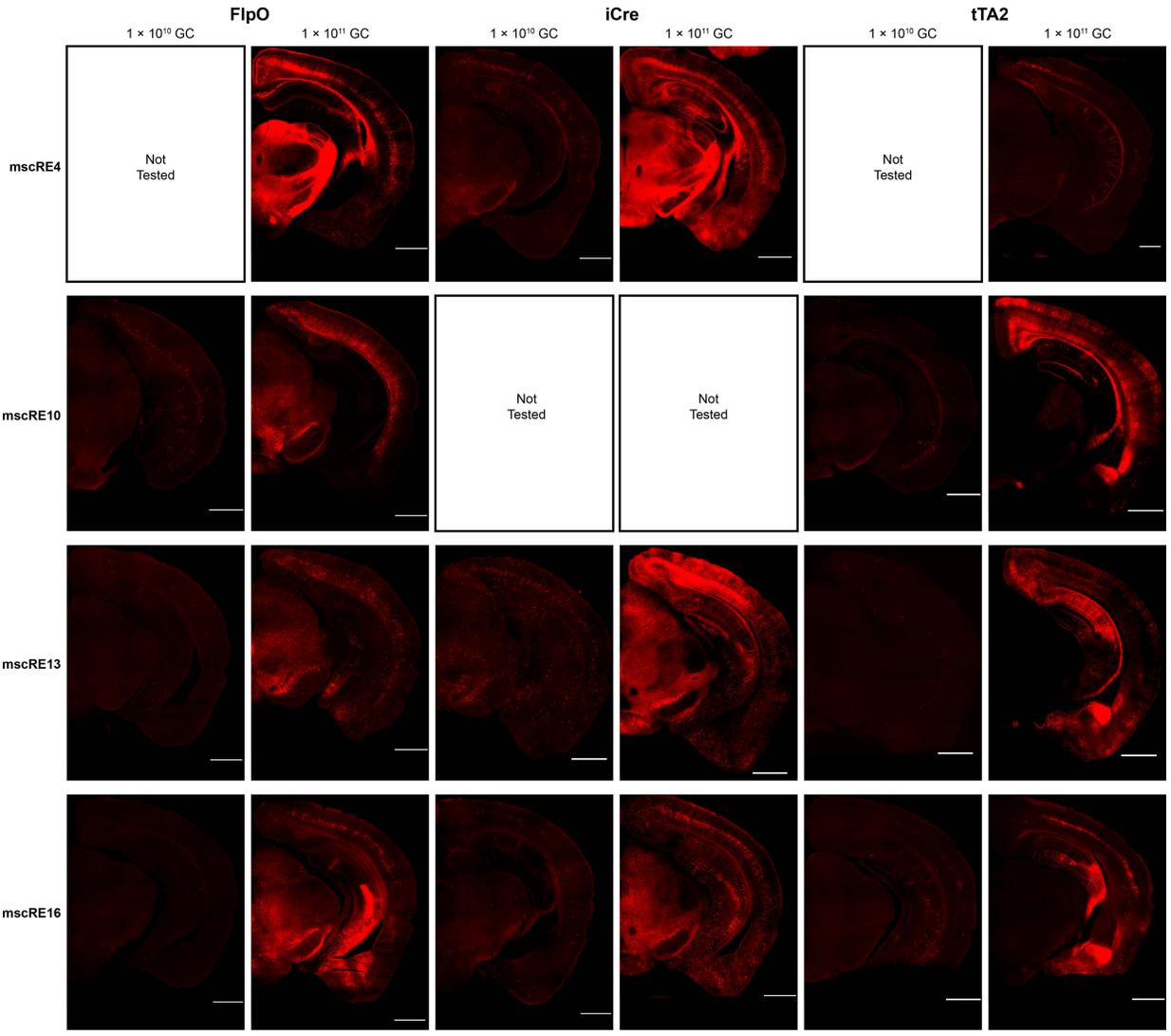
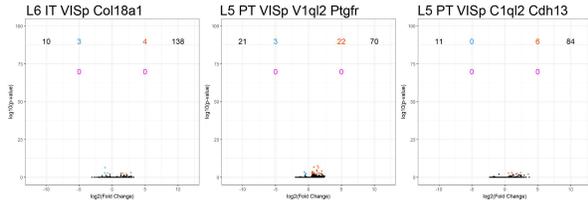
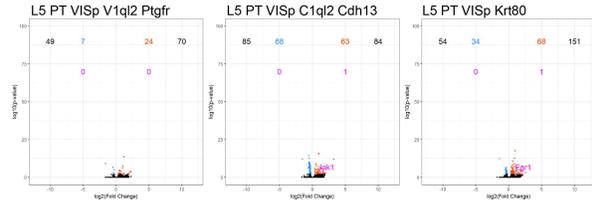


Figure S6. Labeling of retro-orbitally delivered mscRE viruses, related to Figure 5. Native fluorescence of tdTomato imaged in reporter mice injected retro-orbitally with enhancer-driven recombinase or tTA2; 1×10^{10} genome copies (GC) or 1×10^{11} GC of virus was delivered. Each enhancer is shown in a row, and each driver/reporter/dose combination is shown in a column. Reporter mouse lines used: *Ai65F* for FlpO, *Ai14* for iCre, and *Ai63* for tTA2. All scale bars (white) are 500 μm .

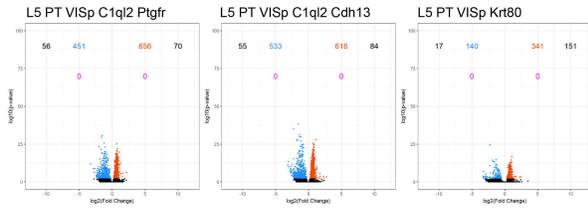
A mscRE1-SYFP2 Retro-orbital in wt



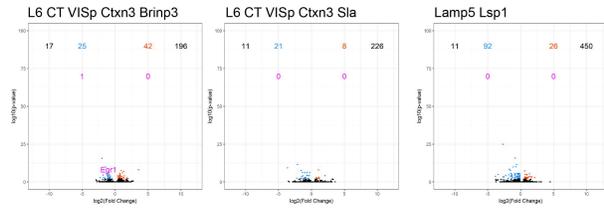
B mscRE4-SYFP2 Retro-orbital in wt



C mscRE4-FlpO Retro-orbital in Ai65F

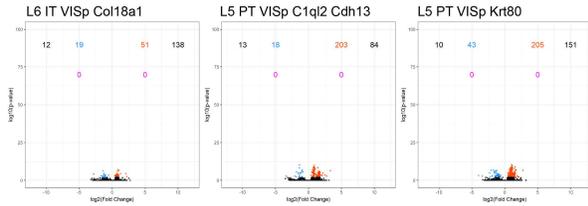


D mscRE10-FlpO Retro-orbital in Ai65F

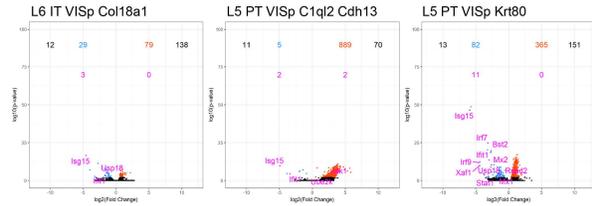


E mscRE16-FlpO Retro-orbital in Ai65F

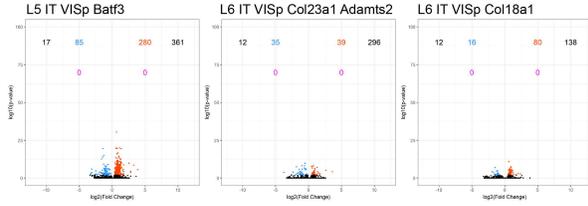
F mscRE4-EGFP Stereotaxic in wt, 25 nL, 4.1 x 10⁸ GC



G mscRE4-EGFP Stereotaxic in wt, 50 nL, 8.2 x 10⁸ GC



H mscRE16-EGFP Stereotaxic in wt, 50 nL, 9.7 x 10⁸ GC



I mscRE16-EGFP Stereotaxic in wt, 250 nL, 4.8 x 10⁹ GC

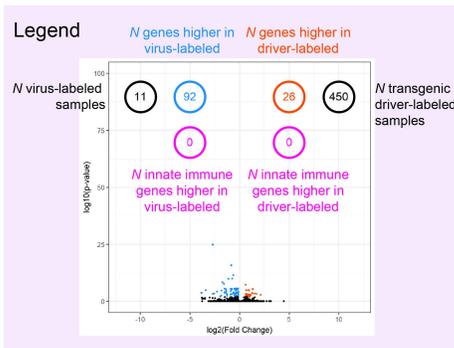
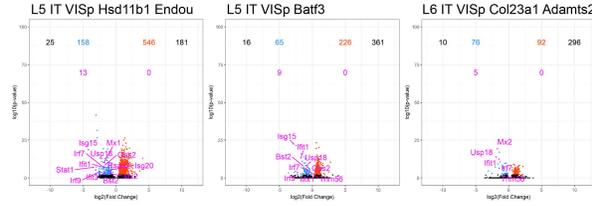


Figure S7. Differential gene expression induced by viral labeling, related to Figure 4 and S5. For each set of scRNA-seq data collected from virally labeled cells (**A–I**), we selected all cell types to which at least 10 cells were mapped. For each selected cell type, we performed pairwise differential gene expression analysis between virally labeled cells and cells labeled by transgenic recombinase driver mouse lines from Tasic *et al.* (Tasic et al., 2018) (**STAR Methods**). A volcano plot is shown for each comparison. Each gene is represented by a point. $\text{Log}_2(\text{fold change})$ is displayed on the x-axis, with genes higher in virally-labeled cells on the left and higher in transgenic mouse lines on the right. $\text{Log}_{10}(\text{adjusted p-value})$ is displayed on the y-axis. Significantly differentially expressed genes (adjusted p-value < 0.01) are highlighted in blue (higher in virally-labeled cells) and orange (higher in transgenic mouse lines). Differentially expressed innate immunity-related genes are highlighted in magenta. Counts for the number of cells in each group and the number of differentially expressed genes and immune-related genes are shown in each plot. The legend below panel (**H**) applies to all plots.

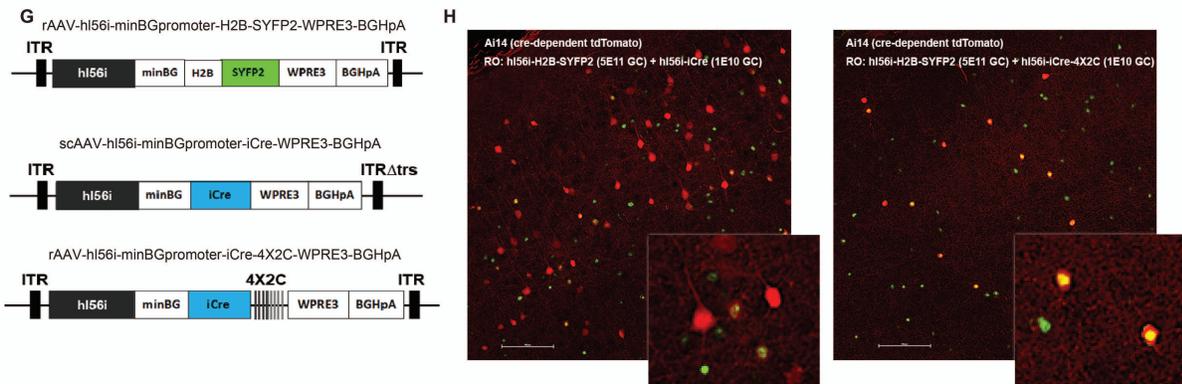
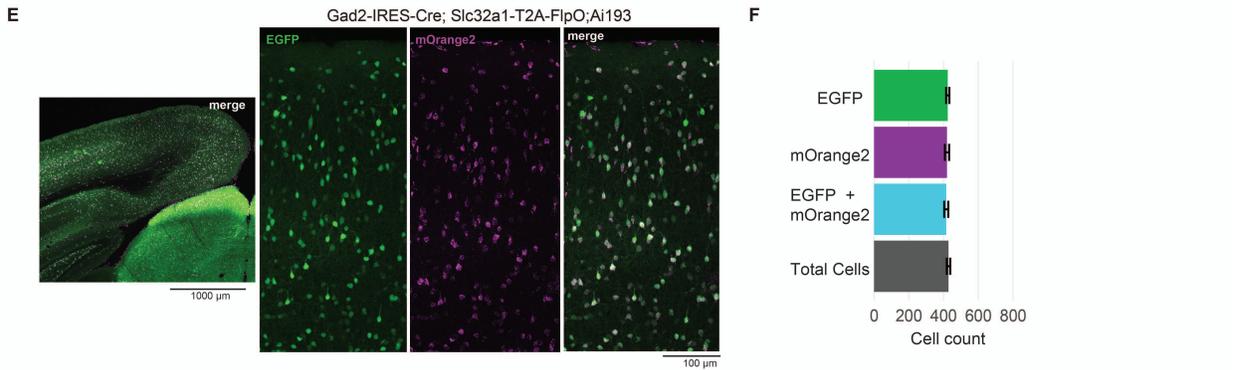
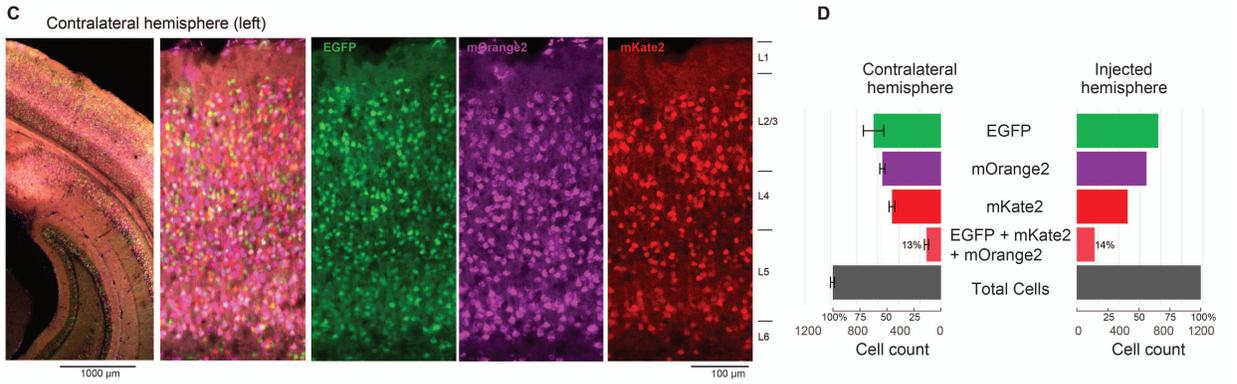
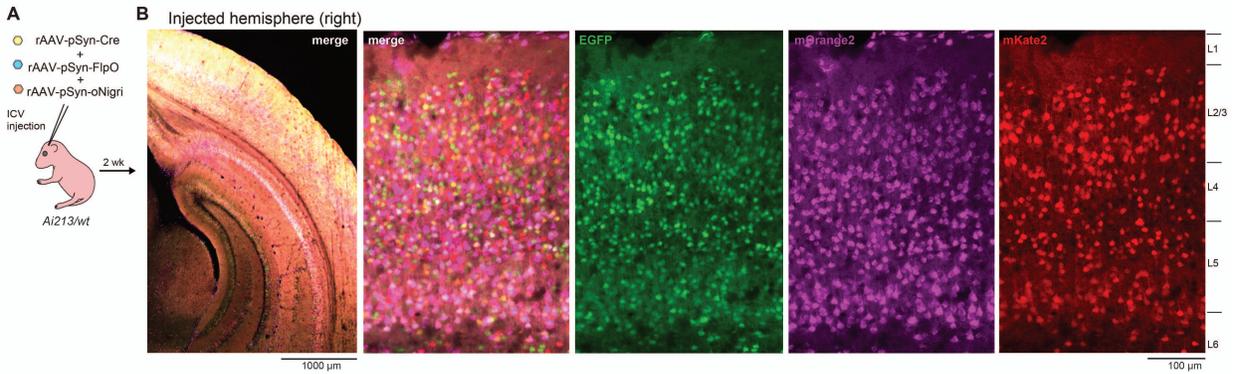


Figure S8. Increased labeling efficiency achieved with ICV injection of viruses or all transgenic strategy and new vector to target forebrain GABAergic neurons, related to Figure 7. (A-C) *Ai213/wt* animals at postnatal day 3 (P3) were injected with a mixture of the three pan-neuronal recombinase viruses each at 2.1×10^{10} GC into the right cerebral ventricle and reporter expression in the brain was analyzed two-weeks post-injection. **(D)** The labeled cells for each fluorophore alone or the combination of all three were counted in confocal images collected from primary visual cortex of ICV injected animals ($n = 2$). Quantification from both hemispheres is presented and shows a similar number of labeled cells in all groups between both regions. **(E)** Representative confocal images from VISp of a *Gad2-IRES-Cre/wt;Slc32a1-T2A-FlpO/wt;Ai213/wt* triple-transgenic animal. Pan-inhibitory labeling patterns were observed for both fluorophores with near perfect overlap of EGFP and mOrange2 when quantified **(F)**. Data in bar graphs are mean \pm S.E.M. **(G)** Diagram of AAV vector designs. Top vector: *DLX* h156i enhancer driving a histone 2B-tagged SYFP2 fluorophore, Bottom left vector: h156i enhancer driving iCre, Bottom right vector: h156i enhancer driving iCre with 3' 4X2C mAGNET sequence to suppress excitatory neuron labeling (Sayeg et al., 2015). **(H)** Viruses were co-delivered (RO) into *Ai14* mice at the doses indicated and native tdTomato (red) and SYFP2 (green) fluorescence was evaluated in micrographs from adult mouse neocortex. Labeling of both excitatory and inhibitory populations was observed with h156i-iCre virus (left panel), compared to exclusively GABAergic cell type labeling with h156i-iCre-4X2C virus (right panel). Scale bars, 100 microns.