A suite of enhancer AAVs and transgenic mouse lines for genetic access to cortical cell types.

Yoav Ben-Simon,[1,4] Marcus Hooper,[1,4] Sujatha Narayan,[1,4] Tanya Daigle,[1,4] Deepanjali Dwivedi,[1] Sharon W. Way,[1] Aaron Oster,[1] David A. Stafford,[2] John K. Mich,[1] Michael J. Taormina,[1] Refugio A. Martinez,[1] Ximena Opitz-Araya,[1] Jada R. Roth,[1] Shona Allen,[2] Angela Ayala,[1] Trygve E. Bakken,[1] Tyler Barcelli,[1] Stuard Barta,[1] Jacqueline Bendrick,[1] Darren Bertagnolli,[1] Jessica Bowlus,[1] Gabriella Boyer,[1] Krissy Brouner,[1] Brittny Casian,[1] Tamara Casper,[1] Anish B. Chakka,[1] Rushil Chakrabarty,[1] Rebecca K. Chance,[2] Sakshi Chavan,[1] Maxwell Departee,[1] Nicholas Donadio,[1] Nadezhda Dotson,[1] Tom Egdorf,[1] Mariano Gabitto,[1] Amanda Gary,[1] Molly Gasperini,[1] Jeff Goldy,[1] Bryan B. Gore,[1] Lucas Graybuck,[1] Noah Greisman,[1] Francoise Haeseleer,[1] Carliana Halterman,[1] Olivia Helback,[1] Dirk Hockemeyer,[2] Cindy Huang,[1] Sydney Huff,[1] Avery Hunker,[1] Nelson Johansen,[1] Zoe Juneau,[1] Brian Kalmbach,[1] Shannon Khem,[1] Rana Kutsal,[1] Rachael Larsen,[1] Changkyu Lee,[1] Angus Y. Lee,[2] Madison Leibly,[1] Garreck H. Lenz,[1] Elizabeth Liang,[1] Nicholas Lusk,[1] Jocelin Malone,[1] Tyler Mollenkopf,[1] Elyse Morin,[1] Dakota Newman,[1] Lydia Ng,[1] Kiet Ngo,[1] Victoria Omstead,[1] Alana Oyama,[1] Trangthanh Pham,[1] Christina A. Pom,[1] Lydia Potekhina,[1] Shea Ransford,[1] Dean Rette,[1] Christine Rimorin,[1] Dana Rocha,[1] Augustin Ruiz,[1] Raymond E.A. Sanchez,[1] Adriana Sedeno-Cortes,[1] Joshua P. Sevigny,[1] Nadiya Shapovalova,[1] Lyudmila Shulga,[1] Ana R. Sigler,[1] La' Akea Siverts,[1] Saroja Somasundaram,[1] Kaiya Stewart,[1] Michael Tieu,[1] Cameron Trader,[1] Cindy T.J. van Velthoven,[1] Miranda Walker,[1] Natalie Weed,[1] Morgan Wirthlin,[1] Toren Wood,[1] Brooke Wynalda,[1] Zizhen Yao,[1] Thomas Zhou,[1] Jeanelle Ariza,[1] Nick Dee,[1] Melissa Reding,[1] Kara Ronellenfitch,[1] Shoaib Mufti,[1] Susan M. Sunkin,[1] Kimberly A. Smith,[1] Luke Esposito,[1] Jack Waters,[1] Bargavi Thyagarajan,[1] Shenqin Yao,[1] Ed S. Lein,[1] Hongkui Zeng,[1] Boaz P. Levi,[1] John Ngai,[2,3] Jonathan Ting,[1] Bosiljka Tasic,[1,*]

[1]Allen Institute for Brain Science, Seattle, WA 98109
[2]University of California, Berkeley, Berkeley, CA 94720
[3]Present affiliation: National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892
[4]equivalent contribution

[*]Correspondence: bosiljkat@alleninstitute.org

Abstract

The mammalian cortex is comprised of cells with different morphological, physiological, and molecular properties that can be classified according to shared properties into cell types. Defining the contribution of each cell type to the computational and cognitive processes that are guided by the cortex is essential for understanding its function in health and disease. We use transcriptomic and epigenomic cortical cell type taxonomies from mice and humans to define marker genes and enhancers, and to build genetic tools for cortical cell types. Here, we present a large toolkit for selective targeting of cortical populations, including mouse transgenic lines and recombinant adeno-associated virus (AAV) vectors containing genomic enhancers. We report evaluation of fifteen new transgenic driver lines and over 680 different enhancer AAVs covering all major subclasses of cortical cells, with many achieving a high degree of specificity, comparable with existing transgenic lines. We find that the transgenic lines based on marker genes can provide exceptional specificity and completeness of cell type labeling, but frequently require generation of a triple-transgenic cross for best usability/specificity. On the other hand, enhancer AAVs are easy to screen and use, and can be easily modified to express diverse cargo, such as recombinases. However, their use depends on many factors, such as viral titer and route of administration. The tools reported here as well as the scaled process of tool creation provide an unprecedented resource that should enable diverse experimental strategies towards understanding mammalian cortex and brain function.

Introduction

To understand how the brain works, we need to establish the structure-function relationships between the units that comprise it and the outputs of its activity. One approach is to gain experimental access to each unit (e.g., a cell type) and manipulate it to observe effects on an animal's physiology and behavior[1–3]. Historically, the definition of building blocks and 'tags' for those building blocks have relied on non-systematic cell-type definition and non-systematic marker gene or genomic enhancer discovery. Many useful tools were created and used to gain insight into cell type function in the brain[1,2], but they may not have always corresponded to cell types[4].

Advances in next generation nucleic acid sequencing and machine learning have transformed cell classification based on single-cell genomics (transcriptomics, epigenomics, etc.)[5–11]. These new methods enable measurements of tens of thousands of molecular properties in individual cells and their subsequent classification into molecularly defined cell types. Single-cell genomics also enables cell type-congruent discovery of marker genes and genomic regulatory elements that can be used to make tools for precision cell type access[6,12,13], and then use them to test the correspondence of molecularly defined cell types to 'functional' cell types (**Figure 1A**).

Here, we generated mouse adult cortex single-nucleus Multiome data (snMultiome 10x, consisting of joint snRNA-seq and snATAC-seq from each nucleus) and used this, along with an existing human cortical dataset[14] to create a transcriptomic and epigenomic cross-species taxonomy consistent with previously published datasets[11,15]. We employed cell type definition in both transcriptomic and epigenomic space to discover corresponding 'tags': marker genes and putative enhancers to build and characterize a large suite of tools for cortical cell types. We screened >680 putative enhancers by systemic introduction (retro-orbital injection) of individual enhancer AAVs, pseudotyped with the blood-brain-barrier penetrating PHP.eB capsid[16–18], expressing a fluorescent protein, and evaluated their expression in the cortex[17,19–25]. We used epifluorescence imaging of brain sections for an initial evaluation of the labeling patterns. Promising vectors were further evaluated for their brain-wide labeling pattern with serial two-photon tomography (STPT), and the cell-type specificity was determined by single-cell RNA sequencing (scRNA-seq) of labeled visual cortex cells (**Figure 1B**). We also optimized expression from these vectors by enhancer core bashing and concatenation, or expression of recombinase versions that can be combined with reporter mice[20,26–28]. Last, we generated new transgenic driver lines which can be used alone or in combination with enhancer AAVs to genetically access specific cortical populations at the finest taxonomical categories (supertype/cluster-level; **Figure 1C**).

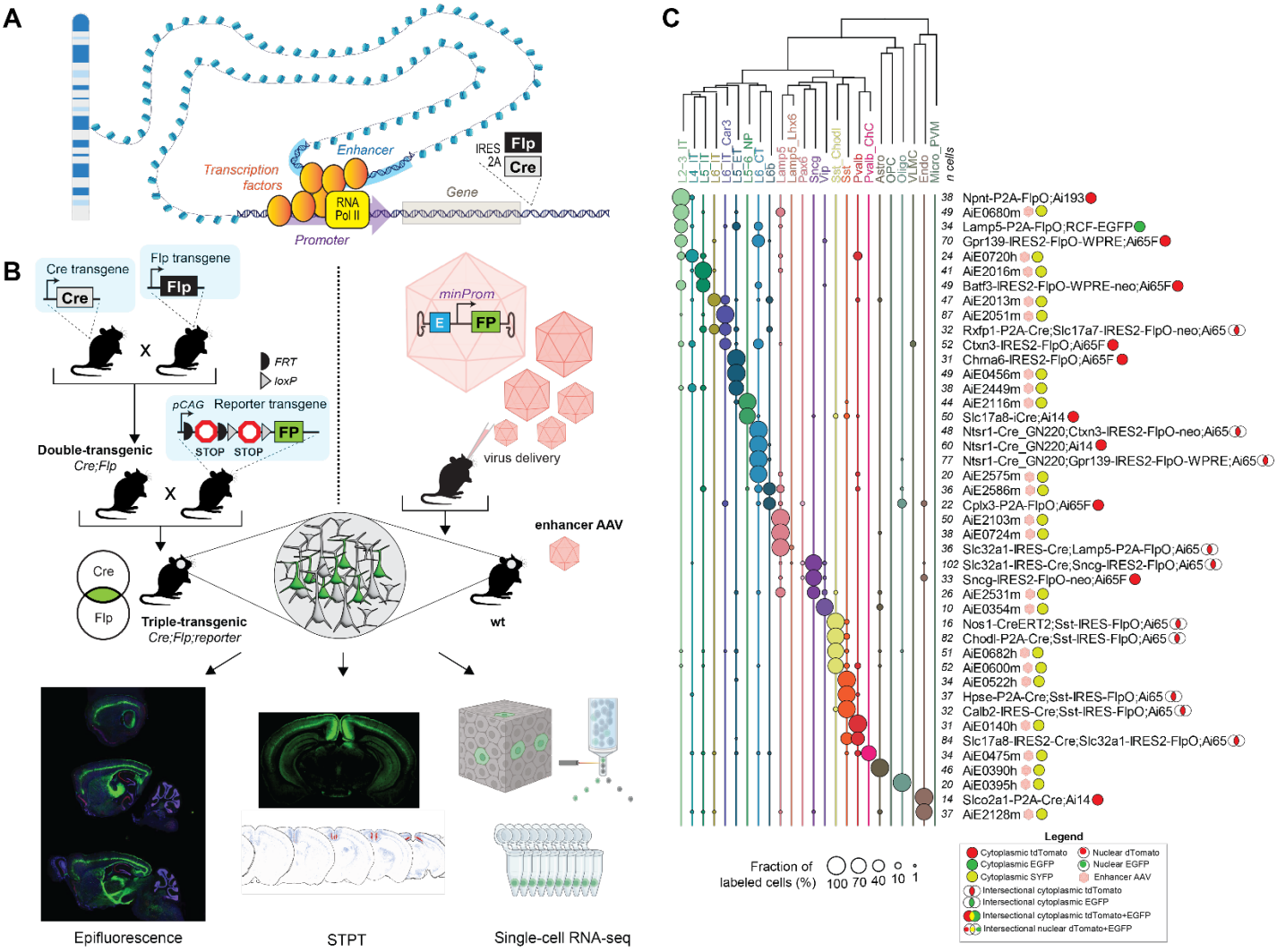


Figure 1: Genetic tool design based on cell type-specific genomic elements. A. A schematic illustration of gene regulation and creation of genetic tools. **B.** Diagrams describing the use of transgenic mouse lines (left), which utilize genomic insertions of recombinases or reporters to specific genomic locations, taking advantage of differentially expressed genes or viral vectors (right), which utilize enhancers to achieve a tool specificity. The three modalities used to evaluate new viral genetic tools are shown at the bottom of the panel: epifluorescence on select sections, serial-two-photon tomography (STPT) on whole brain, and single-cell RNA-seq on visual cortex. **C.** Single-cell RNA-seq data for some of the best tools reported in the study. Cortical taxonomy at the subclass level is on top and fraction of cells labeled per tool is represented by circles. Total number of cells per experiment (***n cells***) is represented in front of each tool's name. Viruses are represented by pink hexagons. Other tools are transgenes. Artwork in (A) and (B) generated using BioRender©.

In summary, we created and characterized a suite of genetic tools and compared them to many existing ones using a common battery of imaging and molecular techniques to provide detailed examination of their specificity and approaches to use (**Figure 1**). We provide all tools through public repositories (Addgene and Jackson Labs) and created a public web resource (**Supplemental information**, in preparation) for scientists to access the information on these tools and select the best tool for their use.

Results

<u>Selection of putative enhancer sequences from mouse and human multiome data</u>

In the present study, we primarily used two new cortical snMultiome datasets: one generated from adult mouse visual, somatosensory, and motor cortices (82,654 nuclei post-QC; **Figure 2A**, left), and one from a recently-published adult human middle temporal gyrus[14] (MTG, 83,977 nuclei post-QC; **Figure 2A**, right). For mouse, we mapped nuclei to an early version of the whole mouse brain taxonomy (AIT21) using Hierarchical approximate nearest neighbor mapping[29]. For the human MTG data, nuclei were mapped to the great apes taxonomy using scANVI[30,31]. We compared the great ape taxonomy and a previously published mouse cortical

taxonomy[10] to generate a common list of cell types that occur in both species. This list was used to re-annotate mapped mouse and human nuclei at an intermediate cell type resolution (subclass). This resulted in comparable cell populations across mouse and human datasets; similar to what is described in these two studies[30,32]. As an example of correspondence across datasets, we show the gene expression data for Nos1 and the accessibility of a successful enhancer targeting Sst_ChodI near this gene in both the mouse and human datasets. (**Figures 2A and 2B**). This annotation was performed on the transcriptomic part of the multiome datasets. Each nucleus's resulting transcriptome-based annotation was used to label its respective ATAC-seq profile for downstream analysis.

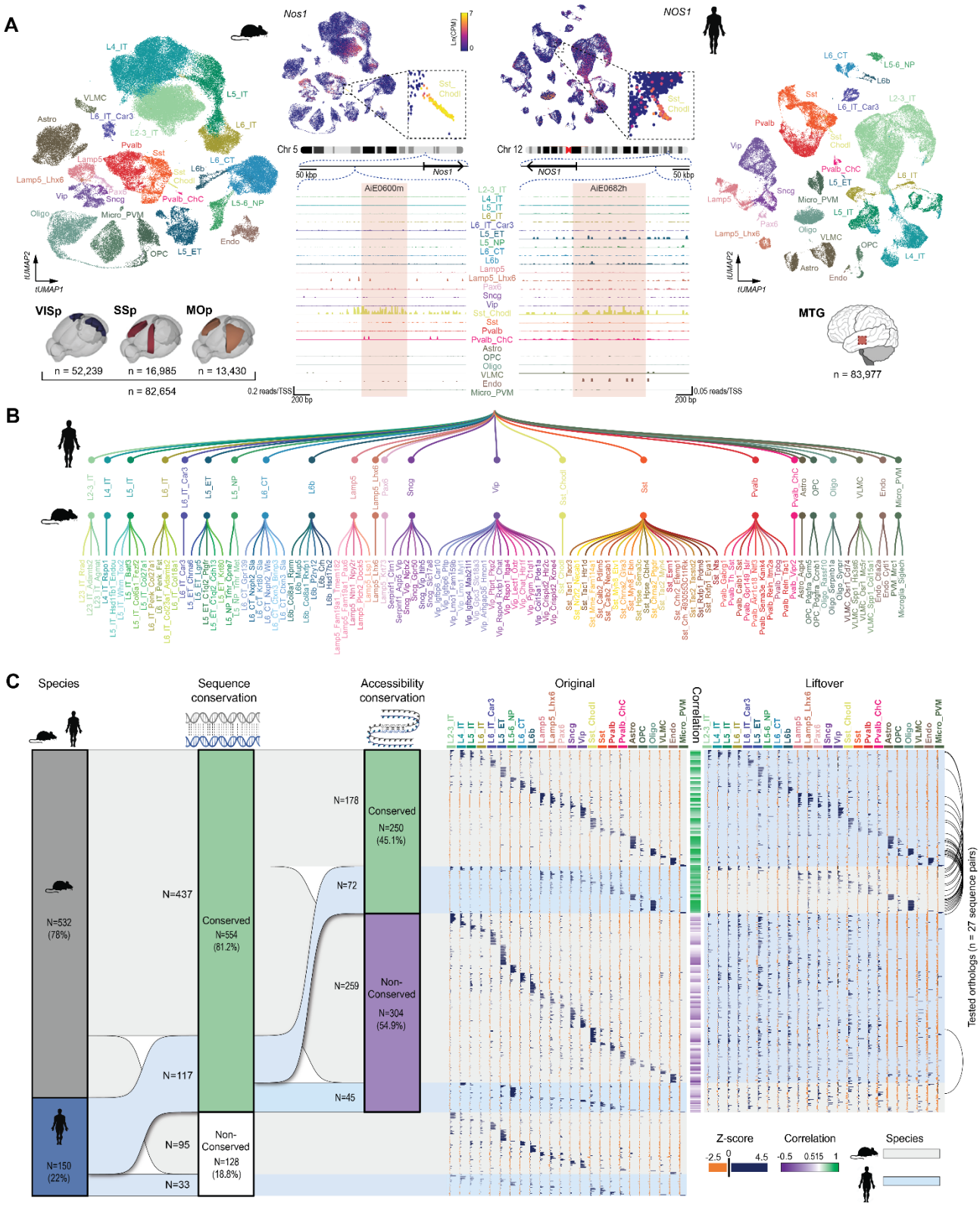


Figure 2: Selection of putative enhancer sequences. A. Mouse and human single-nucleus transcriptomes obtained from single-nucleus multiomes represented in a UMAP and labeled according to cortical cell subclasses. Mouse nuclei were collected from the primary visual (VISp), somatosensory (SSp) and motor (MOp) cortices; and human nuclei were collected from the medial temporal gyrus (MTG). Numbers of nuclei included in the analysis are represented by 'n'. **B.** Simplified representation of the unified mouse-human taxonomy of cortical cell subclasses, along with the cluster-level taxonomy for mouse only. **C.** Summary of all putative enhancer sequences tested, divided by the genome of origin, followed by cross-species conservation of sequence and accessibility (left). The 'Original' and 'Liftover' plots show relative accessibility of all individual sequences in each subclass in their respective species, alongside the relative accessibility of its orthologous liftover sequence in the other species, respectively. Correlation between the original and liftover accessibility data is shown as a green/purple heatmap in the middle of these plots. Black arcs on the very right indicate instances where orthologs from both species were tested.

These new cortical snMultiome datasets and two previously-published sc/snATAC-seq datasets[20,22] were used to select putative mouse and human enhancer sequences for virus production. Pseudobulk ATAC coverage data were generated using ArchR and were used for downstream analysis and enhancer selection [33]. MACS2 and ArchR were used to identify differentially-accessible peaks across subclasses[33,34]. Enhancer selection was based on 1) high and differential chromatin accessibility across the cortical cell subclasses and 2) proximity to marker genes.

In total, we selected 532 sequences from the mouse genome (average length ± SD = 583 ± 256 bp; range = 133-2096 bp, **Figure S1A**) and 150 sequences from the human genome (average length ± SD = 484 ± 176 bp; range = 162-1070 bp) (**Figures 2C and S1A**). Each putative enhancer sequence was assigned a unique ID, consisting of the initials AiE (Allen institute Enhancer), followed by a four-digit numeral and a species indicator ("m" or "h" for mouse or human, respectively). Additionally, each enhancer was associated with one or two target cell populations (TCP). These TCPs were determined by scaling ATAC signal across subclasses and nominating the subclasses with accessibility Z-scores higher than 2 as a TCP.

We evaluated the degree of cross-species sequence conservation using the UCSC genome browser LiftOver tool [35], followed by BLAST, and correlated the chromatin accessibility pattern of the orthologous pairs across the cortical subclasses (**Figure 2C**). Most of the candidate enhancers showed conservation of sequence between the two species (88.9% and 78% of mouse and human enhancers, respectively; **Figures 2C and S1A–S1B**). However, while the degree of sequence conservation was found to be correlated with the degree of accessibility conservation (**Figures S1C–S1E**), most sequence-conserved mouse sequences showed no conservation of accessibility (59.2%) while the opposite was true for the human sequences, where only a minority lacked conserved accessibility (39.0%; **Figure 2C**). This difference is by design, since candidate human sequences with conserved accessibility were prioritized for testing in the mouse under the assumption that these are more likely to work across species[36], but mouse sequences were not prioritized based on this criterion. This dataset also included 27 orthologous pairs from the two species (black arcs in **Figure 2C**).

Enhancer AAV screening *in vivo*

Candidate enhancer sequences were amplified from the respective genome and cloned into the previously described AAV plasmid backbone [20,22], upstream of a minimal promoter (BetaGlobin, Rho or CMV) driving expression of SYFP2[37]. The plasmids were packaged into blood-brain-barrier-penetrating PHP.eB-pseudotyped AAV vectors[17,18] and delivered retro-orbitally (RO) to 30 day-old wild-type mice (average age ± SD = 30.8 ± 5 days, range = 27-70 days; **Methods**)[38]. About four weeks after virus delivery (average duration ± SD = 27.2 ± 1.5 days, range = 23-39 days; **Methods**), the brains were extracted, fixed, and sectioned along the sagittal plane, with five selected sections spanning the mediolateral axis mounted and imaged (**Figure 3A**). We evaluated each experiment by visual inspection and categorized labeled cell populations (LCPs) in the cortex into 1 of 11 visually distinguishable categories. Each LCP was further visually assessed for labeling density relative to the expected density for that target cell population (high vs. low density), as well as relative labeling brightness (high vs. low brightness; **Figure 3B**). We ranked a combination of brightness and density into five categories, with strong yet sparse labeling more highly ranked than weak and dense for any cell population, with the idea that it could potentially be the result of specific targeting of a particular cell type taxon (e.g., cluster) within a given cell type subclass (**Figures 3C and 3D**).

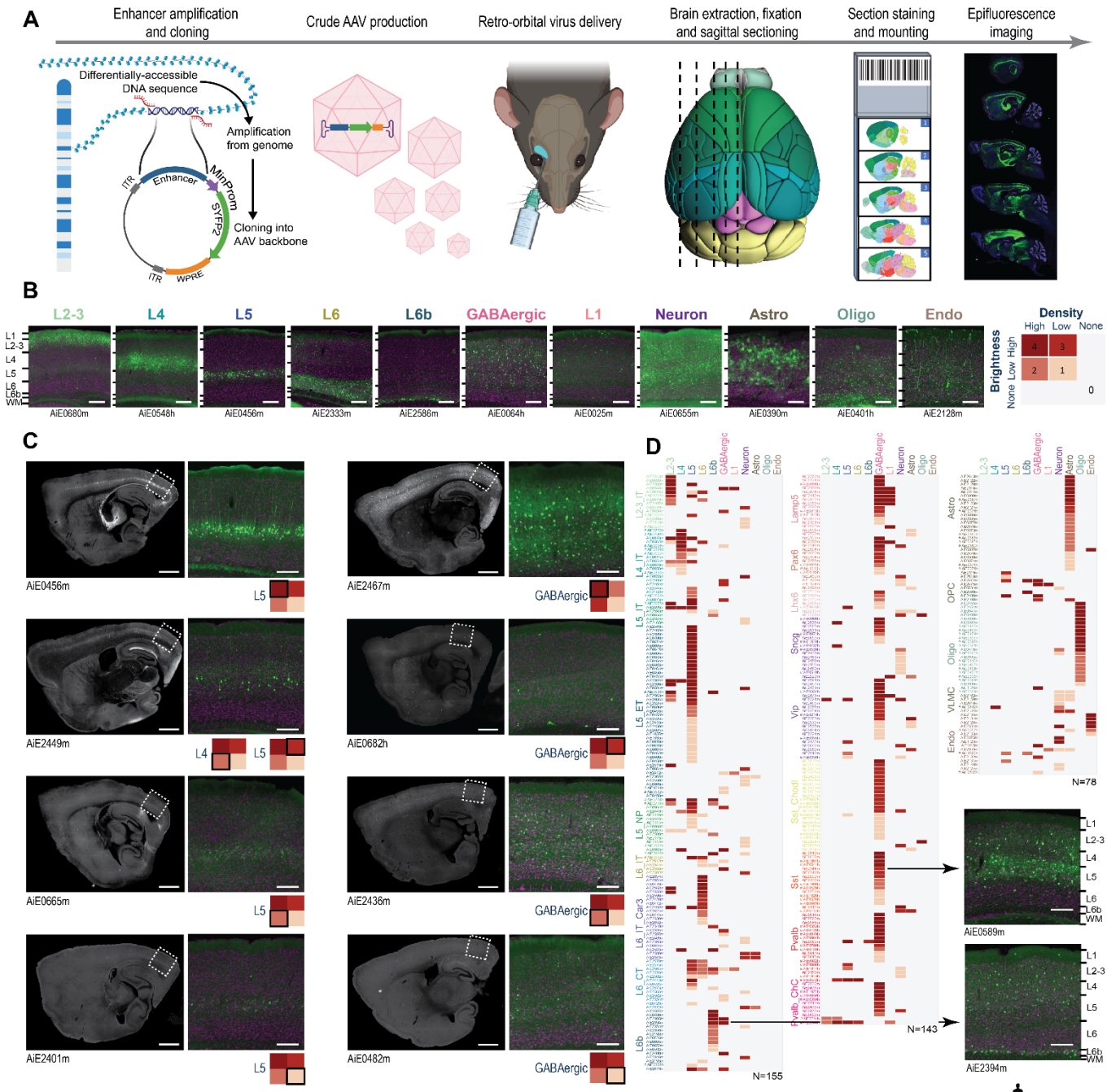

A
Enhancer amplification and cloning
Crude AAV production
Retro-orbital virus delivery
Brain extraction, fixation and sagittal sectioning
Section staining and mounting
Epifluorescence imaging
Differentially-accessible DNA sequence
Amplification from genome
Cloning into AAV backbone
ITR
Enhancer
MinProm
SYFP2
WPRE
B
L2-3
L4
L5
L6
L6b
GABAergic
L1
Neuron
Astro
Oligo
Endo
Density
High
Low
None
Brightness
AiE0680m
AiE0548h
AiE0456m
AiE2333m
AiE2586m
AiE0064h
AiE0025m
AiE0655m
AiE0390m
AiE0401h
AiE2128m
C
AiE0456m
AiE2467m
AiE2449m
AiE0682h
AiE0665m
AiE2436m
AiE2401m
AiE0482m
D
N=155
N=143
N=78
AiE0589m
AiE2394m
E
Micro_PVM
L2-3_IT
L4_IT
L5_IT
L5_ET
L5-6_NP
L6_IT
L6_IT_Car3
L6_CT
L6b
Lamp5
Lamp5_Lhx6
Pax6
Sncg
Vip
Sst_Chodl
Sst
Pvalb
Pvalb_ChC
Astro
OPC
Oligo
VLMC
Endo
GABA-ergic
Neuron
Target cell population
Labeled cell population
Specificity
On-target
Mixed
Off-target
None
Brightness
High
Low
None
Species
Specificity
Brightness
N=532
N=150
N=143
N=59
N=72
N=23
N=68
N=11
N=249
N=57
On-target
N=202
Mixed
N=95
Off-target
N=79
No signal in cortex
N=306
N=111
N=67
N=30
N=91
N=28
N=49
No signal in cortex
N=306
F
Fraction of experiments (%)

Figure 3: A pipeline for systemic enhancer screening in the mouse brain. A. Graphical summary of the different steps in the screening pipeline. **B.** Representative images of the visual cortex, showing the 11 categories of visually distinguishable cortical populations used to evaluate the labeling pattern of each enhancer AAV (left), along with a scoring matrix to determine the brightness of the fluorescent signal, and the labeling density, compared to the average density observed for each category (right). Individual enhancer IDs are shown beneath each image. **C.** Representative sagittal images from eight experiments where either L5 (left) or GABAergic (right) cells were labeled, demonstrating the scoring according to each of the four labeling categories shown in (B). **D.** Summary plot of all enhancers tested which produced any labeling in the neocortex, showing the identity of the labeled population, and the labeling quality, according to the schemes shown in (B). Each enhancer was assigned a target population, according to the subclass where its accessibility was highest (human enhancers are preceded by "#"). **E.** Schematic describing the approach used to determine target specificity, according to the alignment between the target subclass and labeled population (top left), a matrix used for classifying all enhancers, based on a combination of their target specificity and signal brightness (bottom left) and a summary plot showing performance of all tested enhancers according to the categories (right). **F.** Proportion of enhancers in each of the categories specified in (E), according to their genome of origin (top) and the conservation of their sequence and accessibility across the two species. For images, scale bars for full section and expanded views are 1 and 0.2 mm, respectively. Artwork in (A) generated using BioRender©.

Since the morphological properties and distribution across cortical layers for all TCP is known[10], we matched each observed LCP with the expected pattern of the TCP to derive putative target specificity. Following this TCP-LCP alignment, we assigned each enhancer to 1 of 7 categories, indicating whether it had putative on-target, mixed, or off-target labeling, or whether no signal was detected; each of the first three categories was further divided according to the labeling brightness (**Figure 3E**). Multiple TCPs can belong to a single LCP, for example different subclasses of GABAergic neurons. For this reason, this level of analysis can only provide putative on-target specificity for groups clearly differentiable by histological examination. Nonetheless, this scoring system allowed us to independently examine the two main factors associated with enhancer efficacy (**Figures 3E and 3F**) and compare how various factors, such as species of origin, target class, or enhancer directionality can affect enhancer performance. We observed that the distribution of scores varies according to target class and sequence conservation only within the human, but not mouse, sequences and that a higher hit rate was observed for GABAergic and non-neuronal populations for sequences originating in the human genome (**Figure S2A and S2B**). We also found that enhancers which had higher mouse/human sequence homology led to more on-target and strong labeling [36] (**Figure S2C and S2D**) but no differences were observed between tested pairs of orthologs (**Figure S2E and S2F**). Similarly, no differences were found between the 11 pairs of tested reverse complement sequences (**Figures S2F and S2G**). We describe here a screening pipeline for putative enhancer sequences in the mouse brain and conclude that sequences derived from mouse and human genome can perform comparably well, when evaluated in the mouse brain, with roughly one third of them showing on-target specificity as determined by our manual scoring.

Enhancer AAV secondary characterization

Enhancers that showed promising results in our initial screen were nominated for further evaluation with single-cell RNA-seq (Smart-seq v4, SSv4) and/or serial two photon tomography (STPT)[39,40] (**Figure 4A**). To better evaluate target specificity and labeling brightness, we delivered each vector with the same parameters used for the primary screen, and 4-8 weeks later (average duration ± SD = 45 ± 10 days; range = 21-91 days), dissected the visual cortex, dissociated the cells and sorted SYFP2(+)/DAPI(-) cells. These cells were subjected to scRNA-seq and mapped to the mouse VISp cortical taxonomy (**Methods**). We then proceeded to quantify the fraction of cells belonging to any cell type including the TCP in each experiment. In addition, we determined the median SYFP2 mRNA count per million transcripts, normalized to the median count measured following delivery of hSyn1-promoter/enhancer driven SYFP2 (**Figure 4B, S3A and S3B**). This dataset allowed us to highlight enhancers which showed a high degree of specificity, alongside strong SYFP2 expression, as potentially useful tools for targeting their respective TCP. However, this approach does depend on FACS gate delineation. We biased our collection towards more strongly labeled cells to avoid false-positive cells (cells that are negative but may have some SYFP-positive cytoplasm associated with them). This is especially noticeable when two or more populations are labeled and the more abundant cells are either dimmer than or do not survive the dissociation process, as well as the smaller population. To further interrogate the labeling patterns, we generated whole-brain STPT image sets, which are more comprehensive and have better resolution than the primary screen datasets

and can provide us with additional information about brain-wide labeling patterns to complement the SSv4 data (**Figure 4C**).

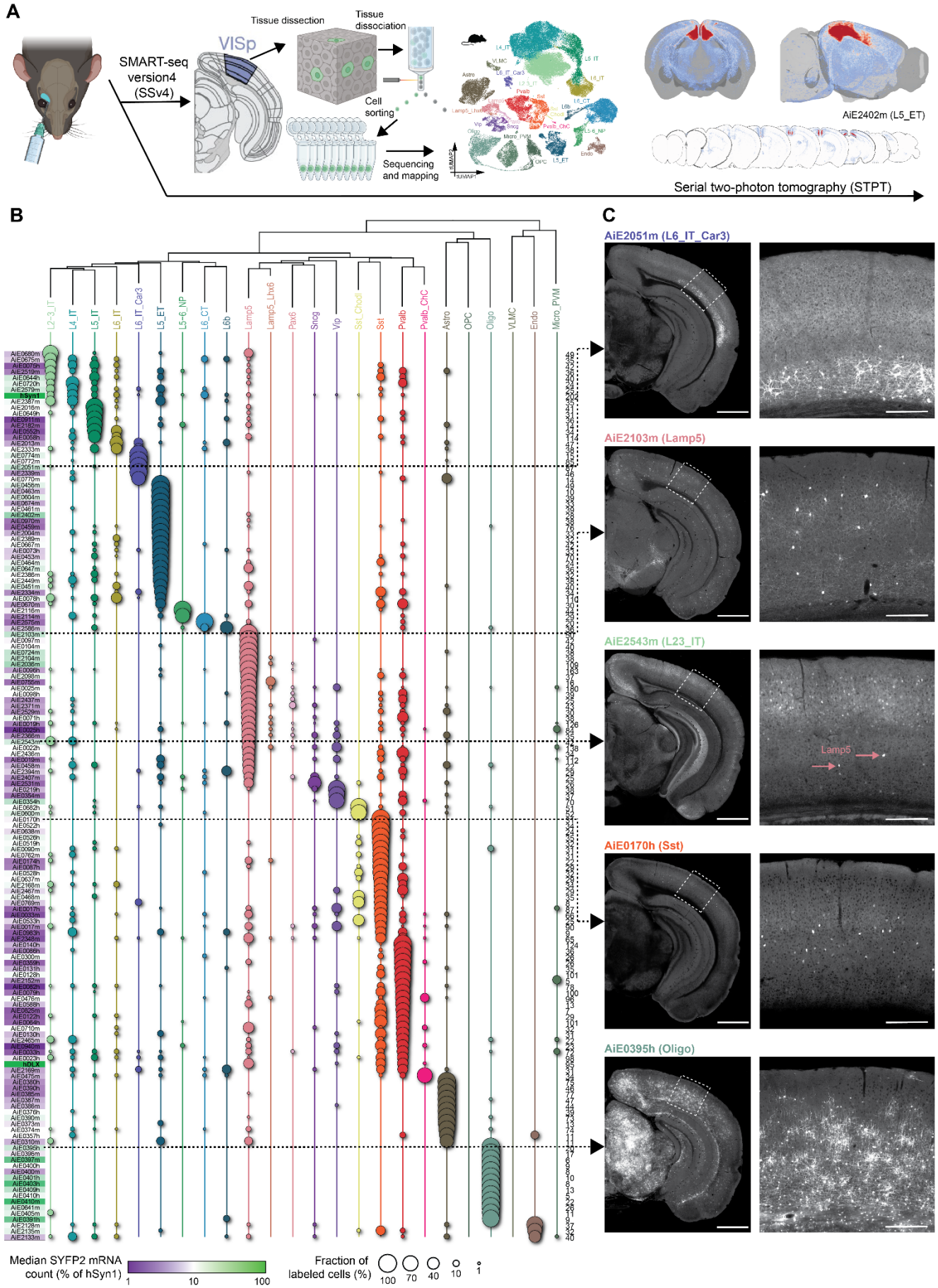


Figure 4: Secondary validation of target specificity, using whole brain imaging and scRNA-seq. A. Schematic of methods for secondary validation. **B.** scRNA-seq analysis using SSv4, of FACS-sorted SYFP2^{+} cells from the mouse visual cortex, following systemic administration of the enhancer AAV. The fraction of cells mapped to each of the cortical subclasses is demonstrated by circle size, and the median SYFP2 mRNA count in each experiment, relative to the hSyn1 promoter, is denoted by a purple-to-green color gradient. The number of sequenced cells for each experiment is shown to the right of the table. **C.** Representative STPT images for 5 individual enhancers, with an expanded view of the visual cortex displayed to the right of the Full-sized hemisphere. Dashed arrows connect each image set to its corresponding SSv4 data. Pink arrows in the 3rd image set point to sparse, yet strong Lamp5 interneurons, which skew the labeling distribution. For images, scale bars for full section and expanded view = 1 and 0.2 mm, respectively. Artwork in (A) generated using BioRender©.

Additional analyses of enhancer AAVs by SSv4 revealed that more than half of screened enhancers exhibited target specificity of >70% and 28% were >90% specific. However, we note that SYFP2 transcript count was low for all, when compared with the pan-neuronal hSyn1 promoter (average ratio ± SD = 13.3 ± 15.3% of hSyn1; **Figures S3C and S3D**). We also note that 29% of enhancers exhibited labeling distribution patterns that were not correlated with the distribution of accessibility across subclasses, suggesting that they might be labeling significantly different populations than the TCPs based on chromatin accessibility (**Figure S3E**). However, this result could also be partially explained by the previously mentioned potential biases in the cell-sorting and sequencing process.

We cross-correlated the specificity of each enhancer, measured as the maximal fraction of labeled cells, with both the SYFP2 transcript count and cross-species conservation of sequence and accessibility. We observed that enhancer specificity was positively correlated with the SYFP2 transcript count, suggesting that specific enhancers also tend to drive stronger cargo expression. We also found a significant positive correlation with the mouse-human sequence homology for the 130 enhancers for which the sequence was conserved, but not with the degree of accessibility conservation across the two species. (**Figure S3F**). It should be noted that this dataset includes only enhancers which demonstrated putative on-target specificity in our primary screen, and therefore is not representative of the entire collection. These data provide detailed information about individual enhancer specificity and strength, which should be considered in their future experimental use.

Enhancer optimization and use diversification

Previous studies have shown that the active site of a putative enhancer sequence can be short and can be discovered through an approach called “bashing”, where shorter fragments of the originally tested sequence are individually examined for their ability to drive cargo expression[41–44]. We proceeded to fragment select enhancers into three putative “cores”, overlapping by 50 bp (C1, C2, and C3; average length ± SD = 212 ± 75 bp) that tile the originally examined sequence[20]. To try to enhance labeling brightness and specificity, we cloned a 3x concatenated version of each core (3xCore, 3xC) into the AAV backbone (**Figure 5A**). We hypothesized that the 3xC enhancers, which contain the most important transcription factor (TF) binding sites, could increase reporter brightness[22,45]. In addition, this approach may also increase labeling specificity in cases where the complete enhancer sequence contains several independent elements, each driving expression in a different population of cells[46,47].

We core-bashed a total of 82 enhancers, creating 3xCores of all three cores or only the middle core (C2), which usually aligns with the peak of chromatin accessibility. For each original enhancer, at least one, and in some cases two bashed enhancers displayed labeling in the original TCP. Most also resulted in a marked increase in brightness (**Figure 5B**), particularly those core-bashed enhancers designed using the middle (C2) core. However, core-bashed versions frequently labeled additional cell populations which were not labeled by the original enhancer (**Figures 5C and 5D**). It is unclear if the concatenation of the cores produced novel TF syntax, which led to *de novo* labeling, or whether these populations were labeled by the full-length (original) enhancer but at levels below detection threshold. We collected SSv4 scRNA-seq data for the best bashed version for each enhancer to compare the change in both specificity and mRNA expression levels with the original enhancer. We observed a nearly 3-fold increase in SYFP2 mRNA count (average ratio ± SD = 274 ± 217% of full-length

enhancer) with a minimal reduction in the maximal fraction of labeled cells (average ratio ± SD = 91 ± 30% of full-length enhancer; **Figures 5E and 5F**).

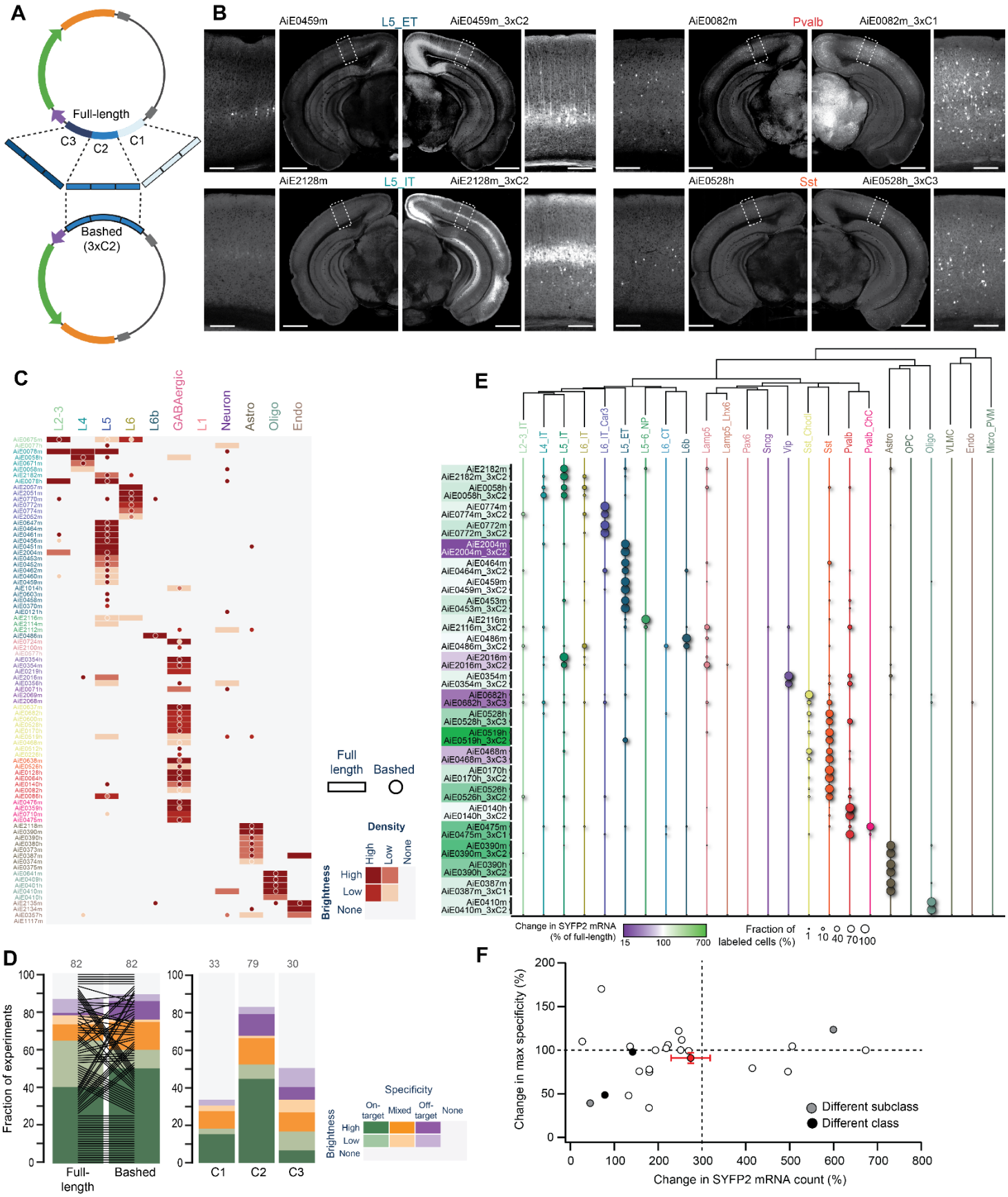


Figure 5: Optimization of enhancer activity through core bashing. A. Schematic representation of the core bashing approach for enhancer optimization (C = Core). **B.** Representative STPT images of coronal hemisections, showing labeling pattern for 4 individual full-length enhancer (left) and their best bashed version (right), which was selected according to a combination of brightness and specificity. A magnified view of VISp is shown alongside each hemisection. **C.** Heat map showing the scoring data of the full-length enhancer (rectangles) alongside its best bashed version (circles). **D.** Dot plot of the scoring data, broken down according to brightness and specificity of the full-length vs. the bashed enhancer (top) and of the different cores tested (bottom). **E.** Summary plot of SSv4 data for full-length enhancers and their bashed counterpart, with circle size denotes the fraction of cells mapped to each of the cortical subclasses in each experiment and the tinted background colors for each pair denote the relative change in SYFP2 transcript count of the bashed-, relative to the full-length enhancer (n=24 pairs). **F.** Summary plot showing the change in specificity as a function of the change in SYFP2 transcript count for all enhancer pairs in (E). Average and SEM are shown in red, along with individual experiments where the bashed version labeled a different subclass (grey) or class (black) than the full-length enhancer. For images, scale bars for full section and expanded view are 1 and 0.2 mm, respectively.

To diversify the use of our enhancer-based genetic toolkit, we replaced SYFP2 with a Cre recombinase, which would enable expression of a variety of tools from existing or new Cre-dependent AAVs or mouse transgenes. We employed iCre or a mutated iCre(R297T) with reduced recombination efficacy[48] and removed the WPRE sequence in a subset of enhancer vectors. The latter two choices were implemented to limit recombination and reduce background labeling. We evaluated the recombination-mediated reporter labeling and compared it to the original pattern defined by direct fluorescent protein labeling (**Figure 6A**). Several enhancers labeling various TCPs showed highly similar labeling patterns following replacement of the fluorophore with a recombinase. Others produced broad labeling patterns (**Figures 6B – 6D**) that could be overcome by using the mutated version of Cre, and/or by titrating the amount of virus delivered (**Figure 6E**).

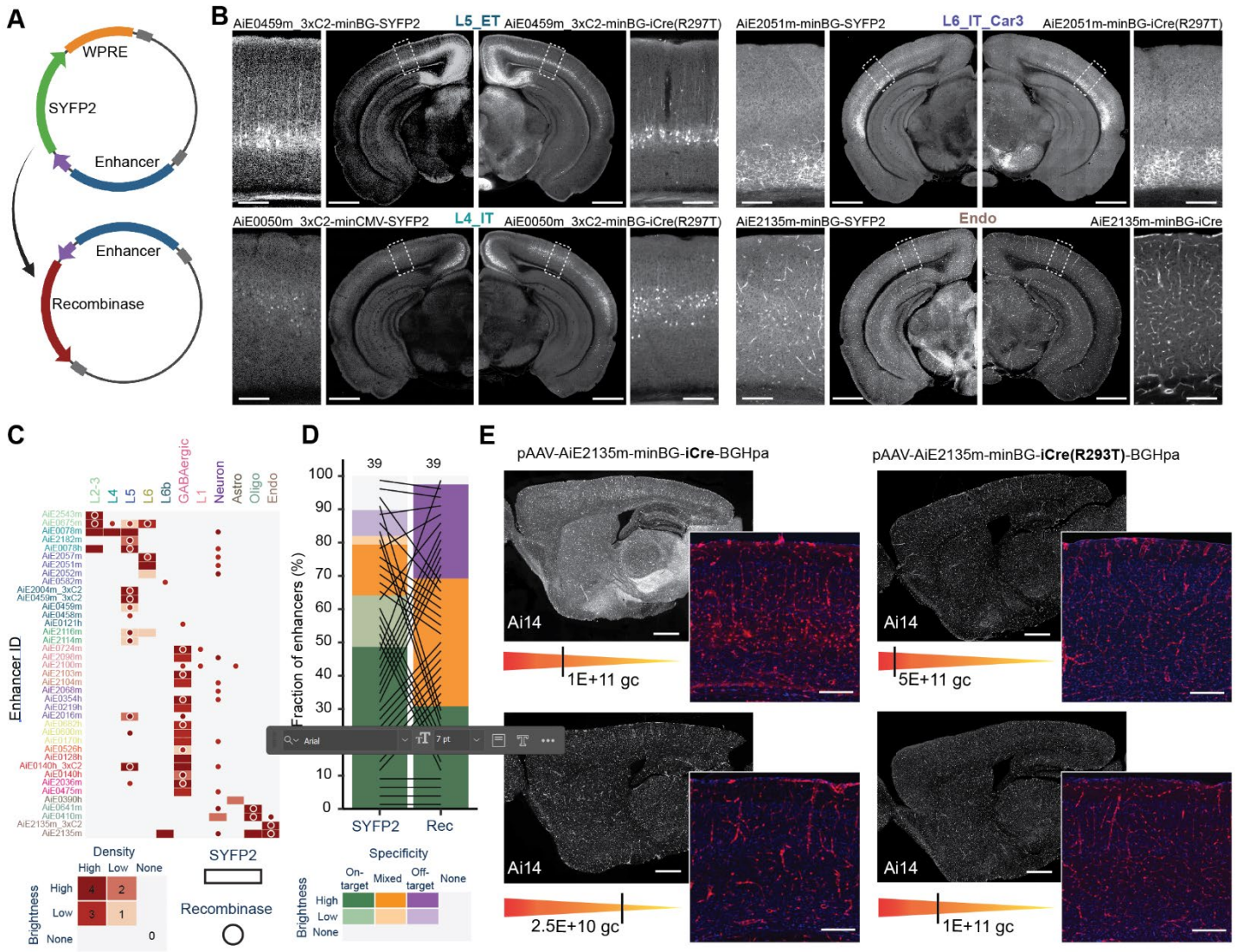


Figure 6: Recombinase-expressing enhancer-AAVs. A. Schematic representation of the vector design. **B.** Representative STPT images of coronal hemisections, showing labeling pattern for 4 individual enhancers expressing SYFP2 delivered to a wt mouse (left) and the same enhancers driving iCre(R297T), delivered to the *Ai14* reporter mouse (right). A magnified view of VISp is shown alongside each hemisection. **C.** Heat map showing the scoring data of the enhancers driving SYFP2 (rectangles) or a recombinase (circles). **D.** Summary plot of the scoring data, broken down according to brightness and specificity of the SYFP or recombinase expression. **E.** Specificity of Cre-dependent recombination in endothelial cells with the Ai2135m enhancer is reduced in iCre compared to the mutated version iCre(R297T) and is more sensitive to the viral dose. Scale bars in B and E, 1 and 0.2 mm for full section and expanded view, respectively.

To further test the applicability of these viral tools, we evaluated their performance under different delivery methods. We observed that intracerebroventricular (ICV) delivery of vectors into the lateral ventricle of newborn mouse pups can lead to specific and widespread labeling, comparable with RO injections, but with substantially brighter fluorescent signal[49] (**Figures S4A – S4C**). However, in several cases, we observed a non-uniform signal distribution and an increase in non-specific labeling, which may have been present below detection levels when the virus was RO-delivered[49]. Interestingly, we found that some populations, particularly endothelial cells, could only be labeled when the virus was delivered RO, but not by ICV administration (**Figure S4C**), which is consistent with previous observations where alternative approaches were used to target this population [17,50]. In addition, we delivered several purified enhancer AAVs stereotaxically to the visual cortex and noted that the fluorescent signal was substantially brighter, restricted to the injection site, and mostly maintained specificity, consistent with previous observations[51](**Figures S5A, S5B**). We observed that even enhancers that exhibited weak fluorescence when delivered RO had a very bright signal when delivered stereotaxically, likely as a result of the substantially higher multiplicity of infection (MOI) at the target region.

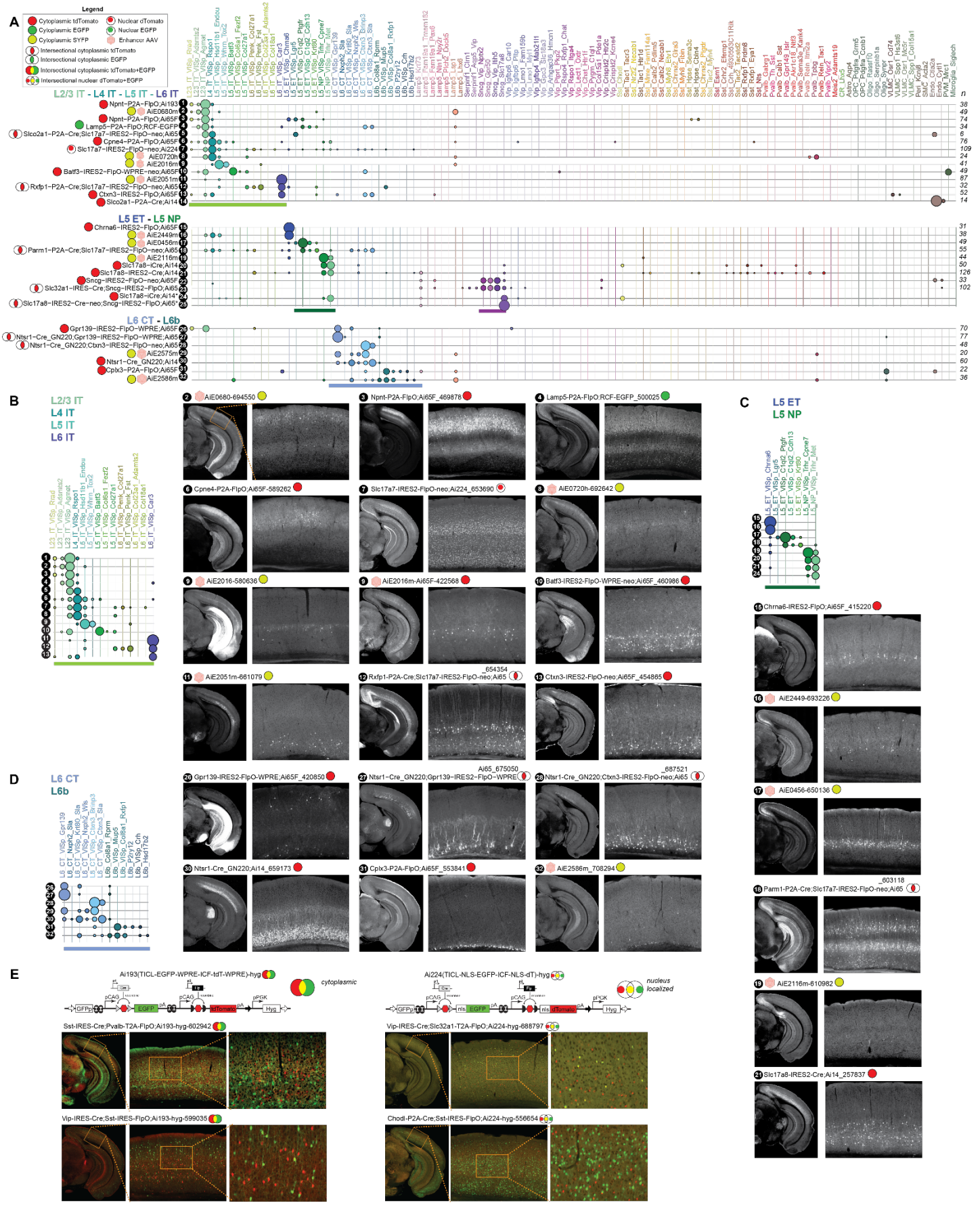
New transgenic lines targeting cortical cell types

Based on the transcriptomic taxonomies of cell types in the mouse cortex[10,15,52], we selected marker genes that could label specific cell type taxons. We targeted regions of the taxonomy for which tools did not exist. Based on select marker genes, we generated a total of 15 new transgenic driver lines: 12 preferentially for glutamatergic and 3 preferentially for GABAergic cell types (**Figure S6A**). We also made two new reporter lines that address the unmet need for AND/OR reporters: *Ai193* and *Ai224* (**Figure S6B**). Unlike *Ai65*[53], which requires both Cre and Flp to express a fluorescent protein (FP, tdTomato in this case), these lines have separate transcription units that independently report presence of Cre or Flp, resulting in GFP expression for Cre, tdTomato (or dTomato) expression for Flp, and both FPs when both Cre and Flp present. *Ai193* expresses cytoplasmic FPs, whereas *Ai224* expresses nuclear FPs (note imperfect nuclear localization for the GFP protein, **Figure S6C**).

The driver lines were crossed to a fluorescent reporter line (*Ai14* for Cre, *Ai65F* for Flp transgenes, respectively)[54,55] and examined by STPT. We examined labeling of cortical cells for expected cell distribution (**Table S7**). Most driver lines produced expected labeling based on the marker gene and cell types for that marker gene. These lines were further analyzed by single-cell transcriptomic profiling (SSv4) of fluorescently labeled FACS-isolated cells. The single cell transcriptomes were then mapped to the VISp taxonomy[10] (**Figures 7 and 8**). Most driver lines labeled the expected cell type to varying extents (Table S7). Other driver lines produced very broad labeling and did not warrant further characterization with SSv4.

Lines from both categories were examined by intersectional crosses to showcase that additional specificity can be achieved to the level of single transcriptomic cell type. The intersectional crosses were performed by first making double recombinase transgenics (Cre and Flp; **Figure 1A**) and then crossing to a dual reporter (e.g., *Ai65* or *RCFL-H2B-GFP*)[53,56]. The triple transgenic mice were examined by STPT and SSv4 (**Figures 7 and 8**). In most cases, the triple transgenic further refined expression and resulted in greater specificity **(Table S7).** For comparison, we present equivalent data from enhancer viruses that were RO-injected into either wild type or reporter mice. In some cases, the viruses were further validated by injecting them into triple transgenic lines that already showed target cell labeling **(Figure 8C13-16).**

In general, the success in targeting various cell type taxons with selected markers and enhancer viruses depends on the discreteness of marker expression or enhancer activity. Some cell type taxons, such as the ITs, are



difficult to target with specificity (**Figure 7A**), and markers selected for individual IT types tend to more continuously expressed across related IT types or subclasses. This is true also for the enhancer viruses that target these populations. A notable exception is the L6_IT_Car3 type, which is labeled by several unique markers

Figure 7: Characterization of new transgenic driver lines, targeting glutamatergic subclasses and types. A. scRNA-seq (SSv4) data showing distribution of labeled excitatory cells mapped to mouse VISp taxonomy at the cluster level. Thirty-one different lines (1-31), both transgenic driver: reporter lines, and wild-type animals that received a systemic delivery of enhancer viruses (marked by pink hexagons), were grouped into panels according to cell type. Representative sections showing a hemi-brain with VISp magnified (box, inset) for: **B.** Lines 1-13 that label IT neurons from Layer 2/3, L4, L5 and L6; **C.** Lines 14-24 that label ET and NP neurons in Layer 5. **D.** Lines 25-31 that label L6 CT and the L6b neurons. **E.** Schematic and representative sections from two reporter lines: *Ai193* (TICL-EGFP-WPRE-ICF-tdT-WPRE)-hyg and *Ai224* (TICL-NLS-EGFP-ICF-NLS-dT)-hyg, crossed with different double transgenic driver lines. In both reporters, cells that express the Cre recombinase are EGFP-positive (green) and those that express FlpO are tdTomato-positive (red). Cells that express both are labelled yellow. Both fluorophores are cytoplasmic in *Ai193* and nucleus-localized in *Ai224*.

(**Figures 7A12-13 and 7B12-13)** in addition to being targeted by several good enhancer viruses (**Figures 4, 7A11 and 7B11**).

In contrast, it was easier to target the ET and NP types at the subclass and cluster level. For example, the *Chrna6* gene labels the unique L5_ET_Chrna6 cell type and the driver line that we generated to target this type, *Chrna6-IRES2-FlpO*, labels the L5_ET_Chrna6 cells (**Figures 7A15** and **7C15**), as well as a small population of L6b Chrna6-expressing cells (L6b_Col8a1_Rprm).

For L6_CT and L6b types, we show that judicious use of crosses can result in nearly single cell-type labeling **(Table S7).** For example, *Ctxn3-IRES2-FlpO* labels an assortment of glutamatergic cell types including L2-3, L4, L5, L6 IT cells as well as L6_CT types (**Figure 7A13 and 7B13**). Crossing this line with a L6_CT-specific line, *Ntsr1-Cre_GN220*[57] (**Figure 7A30 and 7D30**), and then to the intersectional reporter *Ai65*, produces a triple transgenic that precisely labels the only two *Ctxn3*-expressing L6_CT types (L6_CT_VISp_Ctxn3_Sla and L6_CT_VISp_Ctxn3_Brinp3; **Figures 7A-28 and 7B-28; Table S7**). Similarly, we find we can generate a tool that exhibits a remarkably L6_CT specific pattern of expression. *Gpr139-IRES2-FlpO*, a line based on *Gpr139*, is expressed in L2-3, in L6 and in a specific L6_CT type. When this line is crossed with *Ntsr1-Cre_GN220* and *Ai65*, we find that it almost exclusively labels the L6_CT_VISp_Gpr139 cluster (**Figures 7A26-27 and 7D26-27**).

For GABAergic cell types, we selected *Lamp5*, *Sncg* and *Chodl* genes for targeting. Although *Lamp5* is an excellent marker for the Lamp5 subclass of GABAergic cells, it also labels glutamatergic cells in L2-3, L5 and L6 (**Figures 7A4, 7B4, 8A1 and 8B1**). To exclude expression in glutamatergic cells, we generated the triple transgenic by including the pan-GABAergic line *Slc32a1-IRES-Cre.* We show that *Lamp5-2A-FlpO;Slc32a1-IRES-Cre*;*Ai65* exclusively labels the Lamp5 subclass of GABAergic interneurons in the cortex (**Figures 8A-4 and 8B-4**). We also find that this subclass can be very specifically labeled by many enhancer viruses (**Figures 8A2-3 and 8B2-3**). Another subclass of MGE derived-GABAergic cells is labeled by the gene *Sncg*. *Sncg-IRES-FlpO;Ai65* labels the *Sncg* population of neurons as well as endothelial cells (**Figures 8A-6 and 8B-6**). We used the same approach as we did for the Lamp5 subclass and crossed this line with the pan-GABAergic line *Slc32a1-IRES-Cre*; thus excluding the endothelial cells and labeling only Sncg GABAergic cells (**Figures 8A-7 and 8B-7**). Interestingly, we were able to use the Sncg line to label the Sncg_Slc17a8 interneuron type, a very specific Sncg type enriched in the frontal cortex [10]. We crossed this line with the *Slc17a8-IRES-Cre* driver line followed by *Ai65* and observed exquisitely specific labeling in the Anterior Lateral Motor (ALM) cortex of the triple transgenic *Sncg-IRES-FlpO;Slc17a8-IRES-Cre*;*Ai65* mice (**Figures 8A8-9 and 8B9**).

To gain access to the Sst_Chodl cell type, we made a Cre line that targets the gene *Chodl* (*Chodl-P2A-Cre*). We found that this line also labels other cells (**Figure 8C**), so we generated the triple transgenic *Chodl-P2A-Cre;Sst-IRES-Flpo;Ai65* that exhibits very specific Sst_Chodl cell type labeling (**Figure 8C12**)**.** This cell type is also quite specifically labeled by two enhancer viruses (**Figures 4B,C; 8C13**). Interestingly, we find that the gene in closest proximity to one of the enhancers that targets the Chodl population is *Nos1*. The *Nos1-CreERT2* also labels Sst_Chodl, but not cleanly (**Figure 8C10-11**). However, *Nos1-CreERT2;Sst-IRES-FlpO;Ai65* quite specifically labels the Sst_Chodl cell type (**Figure 8C10-11**).

We tested if similar level of specificity can be achieved for the Sst_Hpse clusters. The *Hpse-P2A-Cre* line labels the Sst_Hpse clusters as well as a few cells of almost every other type in this subclass. Although the triple

transgenic *Hpse-P2A-Cre*;*Sst-IRES-FlpO*;*Ai65* resulted in an enrichment of the *Hpse*$^{+}$ population, it did not completely eliminate other Sst cell type labeling (**Figure 8C18-19;Table S7**). Therefore, each cross needs to be tested before assuming that based on the marker genes it will provide expected labeling.

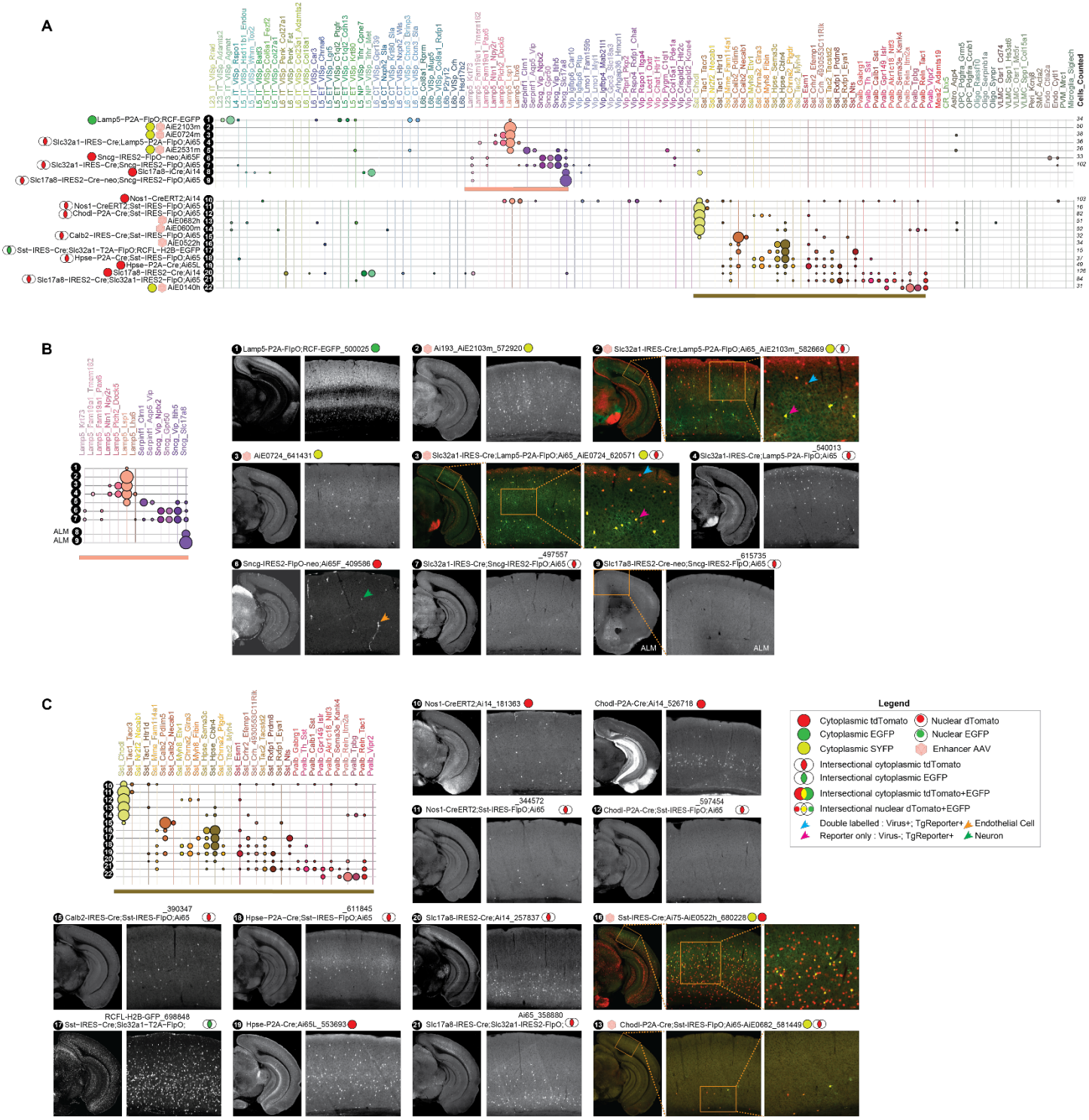


Figure 8: Characterization of new transgenic driver lines, targeting GABAergic subclasses and types. A. scRNA-seq (SSv4) data showing distribution of labeled inhibitory cells mapped to mouse VISp taxonomy at the cluster level. Twenty different lines (1-20), both transgenic driver: reporter lines, and wild-type or transgenic animals that received a systemic delivery of enhancer viruses (marked by red hexagons), are sorted according to cortical cell type. **B. Focused representation of RNA-seq data for reagents labeling** Lamp5 and Sncg inhibitory cortical interneuron subclasses. All data (tools 1-7) are from VISp, except for tools 8 and 9 that were characterized in ALM. **C.** Lines 8-20 that label both the Sst and Pvalb subclasses of inhibitory interneurons that arise from the caudal ganglionic eminence during development.

We note that the exact components used to create a transgenic line affect specific expression pattern. This has previously been observed with lines targeting the parvalbumin gene (*Pvalb*), where *Pvalb-2A-Cre* expresses in L5_ET types as well as thalamus excitatory cells, while *Pvalb-IRES-Cre* line does not[55]. Consistent with previous observations, we note that compared to IRES, the 2A peptide usually produces broader recombinase and therefore reporter expression. This effect is likely a consequence of 2A fusion better capturing the lower end of the range of marker gene expression than IRES. Similarly, the addition of WPRE increases the overall expression level of the recombinase from the targeted locus. For example, *Gpr139-IRES2-FlpO-WPRE*, is expressed in a broader range of cells including L2-3, when compared to both *Gpr139-IRES2-FlpO and Gpr139-IRES2-FlpO-neo* (**Figure S7A5-6; S7B4-6**).

As observed before[10], presence of the Neomycin selection cassette (Neo) can affect transgene expression. Some of the lines we include in this paper were generated by non-selection-based targeting, whereas others contained Neo. It is a common practice to remove Neo after the line is established[55]. However, we have observed that in many cases, the expression patterns are different with and without Neo, and these differences may be useful depending on the application. In most cases, the removal of Neo results in an expansion of expression as in the case of *Slc17a8-IRES2-Cre* (**Figure S7B6-7**). It does not appear to make a difference for the *Gpr139-IRES2-FlpO* line (**Fig S8B4-5**). In contrast, *Chrna6-IRES2-FlpO-neo* labels a small but well-defined portion of L6b cells (L6b_VISp_Col8a1_Rxfp1 and L6b_Col8a1_Rprm) in addition to the intended L5_ET cells. However, when the Neo cassette is removed, its expression is restricted to a single cell type: L5_ET_VISp_Chrna6 (**Figure S7A1-2**).

Discussion

The definition of cell types and their taxonomy in the central nervous system and genetic access to each one are essential for our understanding of how they contribute to nervous system function and dysfunction[2,58,59]. We utilized single-cell transcriptomic and epigenomic data to identify marker genes for generation of new transgenic lines, as well as putative enhancers for creation of enhancer AAVs.

Currently, the use of transgenic mouse lines is the leading approach for gaining access to molecularly identified cellular populations. Knock-in transgenic mice are created by inserting an exogenous gene (e.g., fluorophore, recombinase or transcription factor) next to the marker gene in the genome[54,60]. This approach takes advantage of the endogenous gene regulatory elements *in situ* to enable selective expression of various transgenes for cell labeling, monitoring and/or perturbation. These knock-in transgenes usually produce expected cell-type labeling but their efficiency, specificity, and strength for the same locus can vary depending on the exact components inserted[61]. This contrasts with random transgenes where the endogenous elements have been taken out of context and can be either inserted randomly in the genome or provided in non-genome-integrated context. Depending on the elements, insertion sites, etc., they work with varied success. All the transgenic driver lines described in this paper were generated by knocking in a recombinase into the endogenous gene locus. When we assessed the extent to which each line labeled the 'target' population, we found that several driver lines labeled multiple cell types. In these cases, we find that an effective mechanism of generating cell type specificity is by constructing a crossing scheme to exclude the 'non-target' cell types.

The success of this approach varies and needs to be established for every specific application since it appears to correlate with discreteness of cell type and the marker gene/enhancer used to access it. For example, for glutamatergic IT types, we had more difficulty identifying marker genes and enhancers that were congruent with cell types: there is a lot of continuity in gene expression and cell type landscape. We did our best to select markers that could parcellate this landscape, sometimes trying to target a unique cell type, sometimes a set of related types, and in other cases, we relied on enhancer viruses.

Transgenic driver lines enable reliable and reproducible labeling of cell types based on differential gene expression. These lines, when coupled with reporter mouse lines or viral vectors, are responsible for the bulk of our current understanding of the functional properties of individual cell types[15,53,54]. However, it is not practical to completely rely on transgenic lines for every experiment. Generation, validation, and subsequent maintenance of new transgenic lines is expensive and takes time. Transgenes are integrated into the genome and there is little scope of modifying them once they are made. Moreover, expanding this approach to other species depends on generating transgenic animals in that other species and in many other species transgenesis is either not feasible or not easily achievable. While use of enhancer AAVs can potentially circumvent all these obstacles, there are other caveats that should be considered when applying these tools: 1) the degree of specificity and expression strength heavily depend, among other factors, on the concentration and delivery method of the vectors, which can lead to larger variability within and across experiments; 2) access to some populations is more difficult, due to their decreased susceptibility to viral transduction or difficulty in finding usable enhancer sequences; 3) much remains unknown about enhancer function and how they drive gene expression in a cell type-specific manner. Many potential caveats which could influence their usability might still be unknown.

In this study, we provide a detailed and comprehensive view of how these tools can be developed, screened, and applied to label most subclasses of cortical cells. We demonstrate that the species of origin has little effect on its performance, suggesting that many of these sequences can potentially be equally useful when delivered into other mammalian species, and possibly even in non-mammalian vertebrates. We also found that while the degree of cross-species conservation of sequence contributes to its overall performance, this contribution remains marginal and many human sequences which produced strong and specific labeling in the mouse do not align with any region of the mouse genome. This observation implies that the basic units which confer an enhancer with its properties might make up a small fraction of the cloned sequence. This observation is further supported by our experiments with bashed enhancers, which show that the properties of enhancers can remain intact within fragments as short as 200 bp (**Figure 5**).

To evaluate enhancer specificity at a higher resolution, we performed scRNA-seq on isolated cells labeled by individual enhancer AAVs. We confirmed that many enhancer AAVs designated as ‘promising’ in primary screen showed high specificity for a single cell type taxon either at the subclass or cluster level (**Figure 4**). While measurements of enhancer AAV-driven SYFP2 transcript counts were much lower than those driven by the pan-neuronal hSyn1 promoter, which could impede their use in driving functional cargo such as effectors and indicators, this limitation can potentially be overcome by optimization of enhancers through core bashing and concatenation (**Figure 5**), expression of a recombinase (**Figure 6**) or by delivering viral vectors using alternative routes, such as ICV or stereotaxically (**Figures S4 and S5**). We demonstrate that all these approaches can lead to a marked increase in fluorescent signal, often with minimal loss of specificity.

In summary, we provide a suite of transgenic lines, characterized crosses and comparisons with viruses to inform the user on the best choice for their experiment. For any untested cross or combination of transgene and virus, we advise the user to characterize it before assuming specificity based on the marker genes or enhancers used. The tools reported here as well as the scaled process of tool creation provide an unprecedented resource that should enable diverse experimental strategies for understanding mammalian cortex function. The combined toolbox of deeply validated genetic labeling tools we describe here should enable diverse experimental strategies for dissecting mammalian cortex function, including access to several previously inaccessible cortical cell subclasses and types. Furthermore, these tools can be readily applied to future investigations aimed at deciphering the contribution of genetically defined cell types controlling specific behaviors in diverse mammalian species.

Materials and Methods

Animals

Mice were housed in the Allen Institute Vivarium and all animal procedures were conducted in accordance with the Institutional Animal Care and Use Committee (IACUC) protocol #2105. Mice were housed in a 14/10 h day/night cycle with food and water supplied *ad libitum*. A maximum of 5 animals were housed per cage. Both male and female mice were used for the study.

10x snMultiome data

Mice were anaesthetized with 2.5–3% isoflurane and transcardially perfused with cold, pH 7.4 HEPES buffer containing 110 mM NaCl, 10 mM HEPES, 25 mM glucose, 75 mM sucrose, 7.5 mM $MgCl_2$, and 2.5 mM KCl. Following perfusion, brains were isolated quickly, frozen for 2 minutes in liquid nitrogen vapor. Frozen brain samples were sectioned on a cryostat. Mice were anesthetized using 2.5-3.5% isoflurane 300 µm sections were collected at the appropriate region. Micro knives were used to obtain the area of interest according to the Allen reference atlas. Images were collected pre- and post-collection to document which regions were profiled. Dissectates were placed in 12-well plates in the cryostat after collection.

Following dissection, nuclei were isolated using the RAISINs (RNA-seq for Profiling Intact Nuclei with Ribosome-bound mRNA) protocol # PF0334 [62]. Briefly, tissue samples were placed in CST buffer (see protocol) and single cell suspensions were obtained by chopping tissue using spring scissors for 10 minutes in the buffer. Cell suspensions were centrifuged 500 rcf and resuspended in lysis buffer. Nuclei were counted, resuspended, and processed according to the 10x multiome protocol from 10x Genomics. Short-read sequencing was done with Illumina. Fastq generation and alignment to mm10 was done with Cell Ranger ARC (version 2.0.0). Downstream data analysis was performed with scanpy (version 1.9.8).

Human MTG data were collected by subsetting healthy control data from the SEA-AD study[14]. Briefly, brain specimens were donated for research to the University of Washington BioRepository and Integrated Neuropathology (BRaIN) laboratory from participants in the Adult Changes in Thought (ACT) study and the University of Washington Alzheimer's Disease Research Center (ADRC). 10x multiome library preparation was performed as per the 10x multiome protocol (10x Genomics). Sequencing was performed using a NovaSeq 6000, using either a NovaSeq-X or S4 flow cell. Fastq generation and data alignment to GRCh38 was done with Cell Ranger ARC. Downstream data processing was done using the scanpy python package (version 1.9.1).

snMultiome: Mouse snRNA-seq data analysis

Cell clustering and filtering was performed using the standard scanpy workflow. Following this, individual datasets across cortical regions were integrated using SCVI, using individual donors as the batch key. Final cell type annotations for both species were derived from de novo clusters (scanpy). Low quality clusters (containing doublets) were removed after clustering. This was done iteratively. Low quality cells (fewer than 2,000 genes detected for neurons, and fewer than 1,000 genes detected for non-neuronal clusters) were removed. For each de novo cluster, the predominant cell type was used to label the cluster with the appropriate cross-species subclass label. Known marker genes as identified from the Tasic et al 2018 cortical taxonomy were used to check cell type identities. Gene expression data were plotted using scanpy.pl.umap.

snMultiome: snATAC-seq data analysis

snMultiomic ATACseq data were analyzed using ArchR[33]. Pseudobulk coverage data were used for identification of peaks in VISp subclasses using Macs2 as implemented by ArchR (version 1.0.2) [33,34]. Statistically significant peaks were identified using ArchR Bigwig files were used for enhancer selection, data visualization and for determination of accessibility at putative enhancer sites. For downstream analyses, enhancer accessibility was obtained by summing ATAC signal in subclass-specific bigwig files in all bins overlapping enhancer regions. This

was done using the GenomicRanges package in R. Accessibility was scaled across subclasses to obtain z-scores for enhancer accessibility at each enhancer site. This measure was used to characterize enhancer target populations in an unbiased way.

Cross-species analysis

Enhancer sequences were obtained from mm10 using genomic coordinates using the Biostrings package in R. Mouse sequences were lifted over to hg38 and human sequences were lifted over to mm10. BLAST was performed in a similar manner, using rBLAST using the following arguments: word size = 10; reward = 2; penalty = 3; gapopen = 5; gapextend = 2; dust = no; soft masking = false. Enhancers that returned a match using both liftover and BLAST were considered to have conserved sequence overall. Alignment statistics from BLAST output were retained for each enhancer and were used to calculate a percent of bases in the original enhancer sequences that overlapped with the aligned region. Accessibility correlation between mouse and human was done for each enhancer, using signal from bigwig files across all cortical subclasses.

Enhancer cloning

Short enhancer sequences were nominated from the mouse or human genome based on selective chromatin accessibility in either of the cortical subclasses, with preference for sequences found in proximity to marker genes for that subclass. These sequences were subsequently amplified from purified genomic material of wt C57BL6/J mice, cloned into an AAV2-minPromoter-SYFP2-WPRE-bGH backbone upstream of the minimal promoter (BetaGlobin, CMV or Rho) along with a bovine growth hormone polyA (BGHpA), and a woodchuck post-transcriptional regulatory element (WPRE or WPRE3) [63]. Cloning was carried out either using the Gibson assembly method (NEB; Catalog# E2621L) or restriction digestion and ligation. The recombinant plasmids were verified with Sanger sequencing. Select plasmids have been submitted or are in the process of being submitted to Addgene for distribution (**Supplemental Material; Full plasmid list**).

AAV Production

Verified plasmids were packaged into AAV vectors by transient transfection of HEK 293T cells (ATCC CRL-11268). The cells were seeded at 2×10^7 cells per 15-cm dish to achieve ~70% to 80% confluency before transfection. Cells were maintained in DMEM (Thermo Fisher Scientific Cat#1995-065) with 10% Fetal Bovine Serum (FBS; VWR Cat#89510-184) and 1% Antibiotic-Antimycotic solution (Sigma Cat#A5955). Each enhancer AAV vector was mixed with pAdenoHelper and PHP.eB rep-cap plasmids in a ratio of 30:15:15 µg in 1.35 ml of OptiMem (Thermo Fisher Scientific Cat#51985-034), which was then supplemented with 150 µl of 1 mg/mL Polyethylenimine (PEI; Polysciences Cat#23966), incubated for 10 minutes and then added to a single 15 cm plate of fully confluent cells. Twenty-four hours post transfection, the cell medium was changed to media containing 1% FBS and 1% Antibiotic-Antimycotic in DMEM and 72 hours later, cells were harvested into a 50 ml-tube, subjected to three 20-minute long freeze-thaw cycles to lyse cells and release adeno-associated virus (AAV) particles, and the resulting lysate was incubated with benzonase for an additional 30 minutes (Sigma-Aldrich Cat#E1014) at 37°C, to remove non-encapsidated nucleic acids. The crude AAV-containing suspension was centrifuged at 3000*xg* for 10 minutes to remove residual cell debris, and the supernatant was concentrated using an Amicon Ultra-15 centrifugal filter (Sigma Cat # UFC910024) by centrifuging at 3000*xg*, until the volume was reduced to below 150 µl. This crude AAV prep was then aliquoted and kept at -80°C until use.

For vectors intended for stereotaxic delivery, the transfected cells were pelleted and resuspended in 1 ml lysis buffer containing 150 mM NaCl and 50 mM TrisHCl (pH=8), instead of in their growth media, and following the freeze-thaw cycles and Benzonase treatment, the lysate was passed through a 0.22 µm filter, to remove any large debris which might clog the capillary.

AAV titer determination with dd-PCR

For measuring virus titers, we used ddPCR (Bio Rad; QX 200 Droplet Digital PCR System). We used primers against AAV2 ITR for amplification. Seven serial dilutions with the factor of 10 ranging from $2.5x10^{-2}$ to $2.5x10^{-8}$ were used for the measurement. Serial dilutions of $2.5x10^{-5}$ to $2.5x10^{-8}$ were used for fitting the dynamic linear range. Viral titer was calculated by averaging virus concentration of two dilutions within the dynamic linear range. A positive control of a known viral titer, and a negative control with no virus was also run along with all the samples.

Retroorbital (RO), Intracerebroventricular (ICV) and Stereotaxic (STX) Virus Injections

RO injections were performed according to a previously described protocol[38]: Male and female C57BL/6J mice, aged P27–P33, were anesthetized using isoflurane. Each mouse received an injection of $5x10^{11}$ genome copies (GC) diluted to 90 µl with PBS, administered into the retro-orbital sinus. Intracerebroventricular (ICV) injections were performed according to the previously described protocol [64]: P0-2 mouse pups were anaesthetized by placing them on aluminum foil on ice and injected with 5×10^{10} GC diluted to 5 µl with PBS, targeting the lateral ventricles. The injection was made 1 mm lateral to the midline and approximately 1 mm posterior to bregma. Mice were euthanized after a four-week incubation period.

For stereotaxic (STX) injections, filtered PHP.eB-pseudotyped AAV was injected unilaterally or bilaterally into the primary visual cortex (VISp) using the following coordinates (in mm): anterior/posterior (A/P) -3, medial/lateral (M/L) ±2, and dorsal/ventral (D/V) 0.45/0.65. A total volume of 300 nl containing $1.5x10^{9}$ GC/ml virus was delivered at a rate of 50 nl per pulse with Nanoinject II. Before incision, the animal was injected with Bupivacaine (2-6 mg/kg) and post injection, the animal was injected with ketofen (2-5 mg/kg) and Lactated Ringer's Solution; LRS (up to 1 ml) to provide analgesia. Mice that underwent STX injections were euthanized after 25-31 days, transcardially perfused, and the brains were dissected for further analysis.

Brain tissue preparation and image acquisition

Mice were anaesthetized with isoflurane and perfused transcardially with 10 ml of 0.9% saline, followed by 50 ml of 4% PFA. The brain was removed, bisected along the midsagittal plane, placed in 4% PFA overnight and subsequently moved to a 30% sucrose in PBS solution until sectioning. From the left hemisphere, 30 µm sections were obtained along the entire mediolateral axis using a microtome. Five sections, roughly 0.5, 1, 1.5, 2.3 and 3.5 mm from the midline, were collected, stained by DAPI and/or PI to label nuclei and cellular RNA, and after drying for 24 hours at 37 ºC, mounted on a barcoded slide. Once the mounting medium hardened, the slides were scanned with Aperio VERSA Brightfield epifluorescence microscope (Leica) in the UV, green, and red channels, illuminated with a metal halide lamp. After passing QC, digitized images were analyzed by manual scoring.

Primary screen scoring

Each enhancer vector was scored based on the labeling pattern it produced in the neocortex. First, each region of the brain where labeling of cell somata was observed, was manually scored based on the labeling brightness and density, classifying each into either low or high. In addition, we created 11 categories of different cell populations within the neocortex, which could be visually distinguished one from the other, and whenever cortical labeling was observed, in one or more of these populations, each was individually evaluated based on its own brightness and density. Whereas brightness was classified based on whether the labeling was stronger or weaker than the common brightness observed across all experiments, density was evaluated based on the expected density of cells for each of the scored regions or populations, using the nuclear markers as reference. To determine target specificity, we aligned each target cell population with the labeled population which best matches it's known anatomical location, distribution, and morphological characteristics. We determined an enhancer to be "On-target" if the target population aligned with the labeled population, "Mixed" if labeling was observed in populations other populations, in addition to the target one, "Off-target" if labeling was observed

exclusively in population/s other than the target one, and "no labeling" if no labeling was observed in the neocortex, regardless of whether labeling was observed in other brain regions.

SMART-Seq sample preparation and analysis

Sample preparation for SMART-Seq was performed using the SMART-Seq v4 kit (Takara Cat#634894) as described previously[10]. In brief, single cells were sorted into 8-well strips containing SMART-Seq lysis buffer with RNase inhibitor (0.17 U/µL; Takara Cat#ST0764) 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 Cat#FC-131-1096) with NexteraXT Index Kit V2 Set A (Illumina Cat#FC-131-2001) according to manufacturer's instructions except for decreases in volumes of all reagents, including cDNA, to 0.4 x 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 as 50 bp paired end reads. Reads were aligned to GRCm38 (mm10) using STAR v2.5.3[65] with the parameter "twopassMode," 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[11] (https://github.com/AllenInstitute/scrattch.hicat). 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. For plotting, we retained only cells that were assigned to the same cluster in ≥ 80 of 100 rounds. Cells that did not map to the taxonomy confidently were excluded from analysis and further data processing. Mapping results and scRNA-seq sample metadata, including the most-frequently assigned cell type and the fraction of times each cell was assigned to that type, are included in supplemental data.

For experiments involving enhancer AAVs, mice were injected retro-orbitally with the indicated AAV. One month post injection smart seq samples were collected from FACS-sorted cell populations collected from cortical regions. In most cases this was mouse visual cortex, cells were also collected from claustrum. In most cases, all VISp layers were isolated prior to FACS. In others, multiple collections were made using an "upper" (layers 1-4) and a "lower" (layers 5-6) dissection strategy. For these cases, the total number of cells and number of cells per population were summed across collections, and these were used to calculate percentages, indicated by *. For all transgenic lines, cells were summed across all layer collections, which may include single layers or combinations.

Mouse breeding and husbandry

Adult mice were housed under the Institutional Care and Use Committee protocols 1508, 1802, 1806, and 2105 at the Allen Institute for Brain Science, with no more than five animals per cage, maintained on a 14/10 h day/night cycle, with food and water provided *ad libitum*. Both male and female mice were used for experiments and the minimal number of animals were used for each experimental group. Animals with anophthalmia or microphthalmia were excluded from experiments. Animals were maintained on a C57BL/6J genetic background[66]. At the University of California, Berkeley, experiments with mice were conducted under the campus's Animal Care and Use Committee Animal Use Protocol # AUP-2016-08-9032.

Generation of driver lines at UC, Berkeley

To generate mouse lines bearing in-frame genomic insertions of P2A-FlpO or P2A-Cre, we engineered double-strand breaks at the stop codons of the targeted genes using ribonucleoprotein (RNP) complexes composed of SpCas9-NLS protein and in vitro transcribed sgRNA for the following gene targetings:

Cplx3-P2A-FlpO (sgRNA: GGGGAAGTGGTCACATGATA);
Cpne-P2A-FlpO (sgRNA: ATATGAATCGTCCCGGACAC);

Npnt-P2A-FlpO (sgRNA: GATGATGTGAGCTTGAAAAG);
Lamp5-P2A-FlpO (sgRNA: CCAGTACAAGCACATGGGCT);
Chodl-P2A-Cre (sgRNA: ATGGAGGTATAATAATGAAC);
Hpse-P2A-Cre (sgRNA: TCATATACAAGCAGCGATTT);
Parm1-P2A-Cre (sgRNA: CGTTAAGAGTCATCGTAGAG);
Rxfp1-P2A-Cre (sgRNA: ACTCAATTCTTATTCGTAAC);
Slco2a1-P2A-Cre (sgRNA: CAGTCTGCAGGAGAATGCCT).

The RNP complexes were nucleofected into 10^6 v6.5 mouse embryonic stem cells (C57/BL6;129/sv; a gift from R. Jaenisch) along with repair constructs in which P2A-FlpO or P2A-Cre was flanked with homologous sequences 5' and 3' to the target site, thereby enabling homology-directed repair. Colonies grown from transfected cells were screened by PCR for successful integration; proper insertion of the transgene was confirmed by DNA sequencing. Cell lines with normal karyotypes were aggregated with albino morulae and implanted into pseudopregnant females to produce germ line competent chimeric founders, which were then bred to C57BL/6J mice for further analysis.

Generation of driver lines at Allen Institute

Knock-in driver lines contained components that were previously described[53–55]. Targeting of the transgene cassettes into an endogenous gene locus was accomplished via CRISPR/Cas9-mediated genome editing using circularized targeting vector in combination with a gene-specific guide vector (Addgene plasmid #42230). The129S6B6F1 ES cell line, G4, was utilized directly for driver line targeting. Targeted ES cell clones were subject to standard antibiotic selection and correctly targeted ES cells were identified using standard screening approaches (PCR, qPCR, and Southern blots) and injected into blastocysts to obtain chimeras and subsequent germline transmission. Resulting mice were crossed to the Rosa26-PhiC31 mice (JAX Stock # 007743) to delete the pPGK-neo selection marker cassette, and then backcrossed to C57BL/6J mice and maintained in C57BL/6J congenic background. Only mice heterozygous for both reporter and driver transgenes were used for experiments.

Generation of TIGRE3.0 reporter lines

To target multiple transgene expression units into the *TIGRE* locus[67] we employed a recombinase-mediated cassette exchange (RMCE) strategy similar to that previously described[53], but instead of using Flp recombinase for targeting, we used Bxb1 integrase[68] so that Flp recombinase could later be used for transgene expression control. A new landing pad mouse embryonic stem (ES) cell line was generated by taking the 129S6B6F1 cell line, G4[69], and engineering it to contain the components from 5' to 3' Bxb1 AttP-PhiC31 AttB-PGK promoter-gb2 promoter-Neomycin gene-PGK polyA-Bxb1 AttP-splice acceptor-3' partial hygromycin gene-SV40 polyA-PhiC31 AttP within the *TIGRE* genomic region. Southern blot, qPCR and junctional PCR analyses were performed on genomic DNA (gDNA) samples from modified ES cell clones to confirm proper targeting, copy number, and orientation of the components within the *TIGRE* locus. A Bxb1-compatible targeting vector with three independent and conditional expression units was then generated by standard molecular cloning techniques. The vector contained the following components from 5' to 3': gb2 promoter-Neo gene-Bxb1 AttB-partial GFP-2X HS4 Insulators-CAG promoter-LoxP-stop-LoxP-EGFP-WPRE-BGH polyA-2X HS4 Insulators-CAG promoter-FRT-stop-FRT-mOrange2-HA-WPRE-BGH polyA-PhiC31 AttB-WPRE-BGH polyA-2X HS4 Insulators-CAG-nox-stop-nox-mKate2-P2A-WPRE-PGK polyA-PhiC31 AttB-PGK promoter-5' hygromycin gene-splice donor-Bxb1 AttB. The sequence and integrity of the targeting vector was confirmed by Sanger sequencing, restriction digests and *in vitro* testing performed in HEK293T cells. The targeting vector (30 µg of DNA) was then co-electroporated with a plasmid containing a mouse codon optimized Bxb1 gene under the control of the cytomegalovirus (CMV) promoter (100 µg of DNA) into the Bxb1-landing pad ES cell line and following hygromycin drug selection at 100-150 µg/ml for 5 days, monoclonal populations of cells were hand-picked and expanded. gDNA was prepared from the modified ES cell clones using a kit (Zymo Research Cat#D4071) and it was screened by qPCR and

junctional PCR assays to confirm proper targeting into the *TIGRE* locus. Correctly targeted clones were injected into fertilized blastocysts at the University of Washington Transgenic Research Program (TRP) core to generate high percentage chimeras and then the chimeras were imported to the Institute, bred to C57BL/6J mice to produce F1 heterozygous reporter mice, and subsequently maintained in a C57BL/6J congenic background.

Statistical analysis

All values were shown as mean and error bars as ± SEM in Figures and reported as mean ± SD in the main text. Statistical significance was tested with a 1-way ANOVA, followed a Tukey test for post-hoc comparisons, or by the Chi square test for analysis of differences in group proportions. All p-values reported were corrected for multiple comparisons using the Bonferroni correction. All calculations were performed in Microsoft Excel or R. Statistical differences with $p < 0.05$ were considered significant. In Figures, a single asterisk (∗), double asterisks (**), and triple asterisks (***) indicate $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

Acknowledgments

We are grateful to the In Vivo Sciences, Molecular Biology, Histology, Imaging, and Neurosurgery & Behavior teams at the Allen Institute for their technical support. We thank Robert Hunter for coordinating transgenic mouse production. This work was funded by the Allen Institute for Brain Science in addition to NIH grants U19MH114830 to H.Z. and RF1MH121274 to B.T. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/IBC, or the U.S. Government. The authors wish to thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement, and support.

Declaration of Interests

H.Z. is on the scientific advisory board of MapLight Therapeutics, Inc. The other authors declare no competing interests.

References

1. DeNardo, L., and Luo, L. (2017). Genetic strategies to access activated neurons. Curr. Opin. Neurobiol. *45*, 121–129.

2. Luo, L., Callaway, E.M., and Svoboda, K. (2018). Genetic dissection of neural circuits: a decade of progress. Neuron *98*, 256–281. https://doi.org/10.1016/j.neuron.2018.03.040.

3. Peng, H., Xie, P., Liu, L., Kuang, X., Wang, Y., Qu, L., Gong, H., Jiang, S., Li, A., Ruan, Z., et al. (2021). Morphological diversity of single neurons in molecularly defined cell types. Nature *598*, 174–181.

4. Rudy, B., Fishell, G., Lee, S., and Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. Devel Neurobio *71*, 45–61. https://doi.org/10.1002/dneu.20853.

5. Bakken, T.E., Hodge, R.D., Miller, J.A., Yao, Z., Nguyen, T.N., Aevermann, B., Barkan, E., Bertagnolli, D., Casper, T., Dee, N., et al. (2018). Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PLoS One *13*, e0209648. https://doi.org/10.1371/journal.pone.0209648.

6. Li, Y.E., Preissl, S., Miller, M., Johnson, N.D., Wang, Z., Jiao, H., Zhu, C., Wang, Z., Xie, Y., Poirion, O., et al. (2023). A comparative atlas of single-cell chromatin accessibility in the human brain. Science *382*, eadf7044. https://doi.org/10.1126/science.adf7044.

7. Liu, H., Zhou, J., Tian, W., Luo, C., Bartlett, A., Aldridge, A., Lucero, J., Osteen, J.K., Nery, J.R., Chen, H., et al. (2021). DNA methylation atlas of the mouse brain at single-cell resolution. Nature *598*, 120–128. https://doi.org/10.1038/s41586-020-03182-8.

8. Luo, C., Liu, H., Xie, F., Armand, E.J., Siletti, K., Bakken, T.E., Fang, R., Doyle, W.I., Stuart, T., Hodge, R.D., et al. (2022). Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. Cell Genom *2*, 100106. https://doi.org/10.1016/j.xgen.2022.100107.

9. Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular diversity and specializations among the cells of the adult mouse brain. Cell *174*, 1015–1030. https://doi.org/10.1016/j.cell.2018.07.028.

10. Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. Nature *563*, 72–78. https://doi.org/10.1038/s41586-018-0654-5.

11. Yao, Z., van Velthoven, C.T.J., Nguyen, T.N., Goldy, J., Sedeno-Cortes, A.E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., et al. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell *184*, 3222-3241.e26. https://doi.org/10.1016/j.cell.2021.04.021.

12. Shima, Y., Sugino, K., Hempel, C.M., Shima, M., Taneja, P., Bullis, J.B., Mehta, S., Lois, C., and Nelson, S.B. (2016). A Mammalian enhancer trap resource for discovering and manipulating neuronal cell types. eLife *5*, e13503. https://doi.org/10.7554/eLife.13503.

13. Zu, S., Li, Y.E., Wang, K., Armand, E.J., Mamde, S., Amaral, M.L., Wang, Y., Chu, A., Xie, Y., Miller, M., et al. (2023). Single-cell analysis of chromatin accessibility in the adult mouse brain. Nature *624*, 378–389. https://doi.org/10.1038/s41586-023-06824-9.

14. Gabitto, M.I., Travaglini, K.J., Rachleff, V.M., Kaplan, E.S., Long, B., Ariza, J., Ding, Y., Mahoney, J.T., Dee, N., Goldy, J., et al. (2023). Integrated multimodal cell atlas of Alzheimer's disease. Res Sq, rs.3.rs-2921860. https://doi.org/10.21203/rs.3.rs-2921860/v1.

15. Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci *19*, 335–346. https://doi.org/10.1038/nn.4216.

16. Challis, R.C., Kumar, S.R., Chen, X., Goertsen, D., Coughlin, G.M., Hori, A.M., Chuapoco, M.R., Otis, T.S., Miles, T.F., and Gradinaru, V. (2022). Adeno-Associated Virus Toolkit to Target Diverse Brain Cells. Annual Review of Neuroscience *45*, 447–469. https://doi.org/10.1146/annurev-neuro-111020-100834.

17. Chan, K.Y., Jang, M.J., Yoo, B.B., Greenbaum, A., Ravi, N., Wu, W.-L., Sánchez-Guardado, L., Lois, C., Mazmanian, S.K., Deverman, B.E., et al. (2017). Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat. Neurosci. *20*, 1172–1179.

18. Deverman, B.E., Pravdo, P.L., Simpson, B.P., Kumar, S.R., Chan, K.Y., Banerjee, A., Wu, W.-L., Yang, B., Huber, N., Pasca, S.P., et al. (2016). Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat Biotechnol *34*, 204–209. https://doi.org/10.1038/nbt.3440.

19. Dimidschstein, J., Chen, Q., Tremblay, R., Rogers, S.L., Saldi, G.-A., Guo, L., Xu, Q., Liu, R., Lu, C., Chu, J., et al. (2016). A viral strategy for targeting and manipulating interneurons across vertebrate species. Nat. Neurosci. *19*, 1743–1749.

20. Graybuck, L.T., Daigle, T.L., Sedeño-Cortés, A.E., Walker, M., Kalmbach, B., Lenz, G.H., Morin, E., Nguyen, T.N., Garren, E., Bendrick, J.L., et al. (2021). Enhancer viruses for combinatorial cell-subclass-specific labeling. Neuron *109*, 1449-1464.e13. https://doi.org/10.1016/j.neuron.2021.03.011.

21. Hrvatin, S., Tzeng, C.P., Nagy, M.A., Stroud, H., Koutsioumpa, C., Wilcox, O.F., Assad, E.G., Green, J., Harvey, C.D., Griffith, E.C., et al. (2019). A scalable platform for the development of cell-type-specific viral drivers. Elife *8*. https://doi.org/10.7554/eLife.48089.

22. Mich, J.K., Graybuck, L.T., Hess, E.E., Mahoney, J.T., Kojima, Y., Ding, Y., Somasundaram, S., Miller, J.A., Kalmbach, B.E., Radaelli, C., et al. (2021). Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. Cell Rep. *34*, 108754.

23. Nair, R.R., Blankvoort, S., Lagartos, M.J., and Kentros, C. (2020). Enhancer-Driven Gene Expression (EDGE) Enables the Generation of Viral Vectors Specific to Neuronal Subtypes. iScience *23*, 100888.

24. O'Toole, S.M., Oyibo, H.K., and Keller, G.B. (2023). Molecularly targetable cell types in mouse visual cortex have distinguishable prediction error responses. Neuron *111*, 2918-2928.e8. https://doi.org/10.1016/j.neuron.2023.08.015.

25. Vormstein-Schneider, D., Lin, J.D., Pelkey, K.A., Chittajallu, R., Guo, B., Arias-Garcia, M.A., Allaway, K., Sakopoulos, S., Schneider, G., Stevenson, O., et al. (2020). Viral manipulation of functionally distinct interneurons in mice, non-human primates and humans. Nat. Neurosci. *23*, 1629–1636.

26. Friedman, M.J., Wagner, T., Lee, H., Rosenfeld, M.G., and Oh, S. (2024). Enhancer–promoter specificity in gene transcription: molecular mechanisms and disease associations. Exp Mol Med *56*, 772–787. https://doi.org/10.1038/s12276-024-01233-y.

27. Karnuta, J.M., and Scacheri, P.C. (2018). Enhancers: bridging the gap between gene control and human disease. Hum Mol Genet *27*, R219–R227. https://doi.org/10.1093/hmg/ddy167.

28. Lu, T., Ang, C.E., and Zhuang, X. (2022). Spatially resolved epigenomic profiling of single cells in complex tissues. Cell *185*, 4448-4464.e17. https://doi.org/10.1016/j.cell.2022.09.035.

29. Yao, Z., van Velthoven, C.T.J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., Abdelhak, A., Aitken, M., et al. (2023). A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. Nature *624*, 317–332. https://doi.org/10.1038/s41586-023-06812-z.

30. Jorstad, N.L., Song, J.H.T., Exposito-Alonso, D., Suresh, H., Castro-Pacheco, N., Krienen, F.M., Yanny, A.M., Close, J., Gelfand, E., Long, B., et al. (2023). Comparative transcriptomics reveals human-specific cortical features. Science *382*, eade9516. https://doi.org/10.1126/science.ade9516.

31. Xu, C., Lopez, R., Mehlman, E., Regier, J., Jordan, M.I., and Yosef, N. (2021). Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. Molecular Systems Biology *17*, e9620. https://doi.org/10.15252/msb.20209620.

32. Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. Nature. https://doi.org/10.1038/s41586-019-1506-7.

33. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet *53*, 403–411. https://doi.org/10.1038/s41588-021-00790-6.

34. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biology *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

35. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Research *34*, D590–D598. https://doi.org/10.1093/nar/gkj144.

36. Mich, J.K., Sunil, S., Johansen, N., Martinez, R.A., Leytze, M., Gore, B.B., Mahoney, J.T., Ben-Simon, Y., Bishaw, Y., Brouner, K., et al. (2023). Enhancer-AAVs allow genetic access to oligodendrocytes and diverse populations of astrocytes across species. Preprint at bioRxiv, https://doi.org/10.1101/2023.09.20.558718 https://doi.org/10.1101/2023.09.20.558718.

37. Kremers, G.-J., Goedhart, J., van Munster, E.B., and Gadella, T.W.J. (2006). Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius. Biochemistry *45*, 6570–6580. https://doi.org/10.1021/bi0516273.

38. Yardeni, T., Eckhaus, M., Morris, H.D., Huizing, M., and Hoogstraten-Miller, S. (2011). Retro-orbital injections in mice. Lab Anim *40*, 155–160. https://doi.org/10.1038/laban0511-155.

39. Ragan, T., Kadiri, L.R., Venkataraju, K.U., Bahlmann, K., Sutin, J., Taranda, J., Arganda-Carreras, I., Kim, Y., Seung, H.S., and Osten, P. (2012). Serial two-photon tomography for automated ex vivo mouse brain imaging. Nat. Methods *9*, 255–258.

40. Harris, J.A., Mihalas, S., Hirokawa, K.E., Whitesell, J.D., Choi, H., Bernard, A., Bohn, P., Caldejon, S., Casal, L., Cho, A., et al. (2019). Hierarchical organization of cortical and thalamic connectivity. Nature *575*, 195–202.

41. Baliga, N.S. (2001). Promoter analysis by saturation mutagenesis. Biol Proced Online *3*, 64–69. https://doi.org/10.1251/bpo24.

42. Catarino, R.R., and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. Genes Dev *32*, 202–223. https://doi.org/10.1101/gad.310367.117.

43. Hörnblad, A., Bastide, S., Langenfeld, K., Langa, F., and Spitz, F. (2021). Dissection of the Fgf8 regulatory landscape by in vivo CRISPR-editing reveals extensive intra- and inter-enhancer redundancy. Nat Commun *12*, 439. https://doi.org/10.1038/s41467-020-20714-y.

44. Rickels, R., and Shilatifard, A. (2018). Enhancer Logic and Mechanics in Development and Disease. Trends Cell Biol *28*, 608–630. https://doi.org/10.1016/j.tcb.2018.04.003.

45. Singh, G., Mullany, S., Moorthy, S.D., Zhang, R., Mehdi, T., Tian, R., Duncan, A.G., Moses, A.M., and Mitchell, J.A. (2021). A flexible repertoire of transcription factor binding sites and a diversity threshold determines enhancer activity in embryonic stem cells. Genome Res *31*, 564–575. https://doi.org/10.1101/gr.272468.120.

46. Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat Rev Mol Cell Biol *16*, 144–154. https://doi.org/10.1038/nrm3949.

47. Taskiran, I.I., Spanier, K.I., Dickmänken, H., Kempynck, N., Pančíková, A., Ekşi, E.C., Hulselmans, G., Ismail, J.N., Theunis, K., Vandepoel, R., et al. (2024). Cell-type-directed design of synthetic enhancers. Nature *626*, 212–220. https://doi.org/10.1038/s41586-023-06936-2.

48. Hartung, M., and Kisters-Woike, B. (1998). Cre Mutants with Altered DNA Binding Properties. Journal of Biological Chemistry *273*, 22884–22891. https://doi.org/10.1074/jbc.273.36.22884.

49. Grødem, S., Nymoen, I., Vatne, G.H., Rogge, F.S., Björnsdóttir, V., Lensjø, K.K., and Fyhn, M. (2023). An updated suite of viral vectors for in vivo calcium imaging using intracerebral and retro-orbital injections in male mice. Nat Commun *14*, 608. https://doi.org/10.1038/s41467-023-36324-3.

50. de Leeuw, C.N., Korecki, A.J., Berry, G.E., Hickmott, J.W., Lam, S.L., Lengyell, T.C., Bonaguro, R.J., Borretta, L.J., Chopra, V., Chou, A.Y., et al. (2016). rAAV-compatible MiniPromoters for restricted expression in the brain and eye. Mol Brain *9*, 52. https://doi.org/10.1186/s13041-016-0232-4.

51. Hudry, E., and Vandenberghe, L.H. (2019). Therapeutic AAV Gene Transfer to the Nervous System: A Clinical Reality. Neuron *101*, 839–862. https://doi.org/10.1016/j.neuron.2019.02.017.

52. Yao, Z., Liu, H., Xie, F., Fischer, S., Adkins, R.S., Aldridge, A.I., Ament, S.A., Bartlett, A., Behrens, M.M., Van den Berge, K., et al. (2021). A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. Nature *598*, 103–110. https://doi.org/10.1038/s41586-021-03500-8.

53. Madisen, L., Garner, A.R., Shimaoka, D., Chuong, A.S., Klapoetke, N.C., Li, L., van der Bourg, A., Niino, Y., Egolf, L., Monetti, C., et al. (2015). Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. Neuron *85*, 942–958. https://doi.org/10.1016/j.neuron.2015.02.022.

54. Daigle, T.L., Madisen, L., Hage, T.A., Valley, M.T., Knoblich, U., Larsen, R.S., Takeno, M.M., Huang, L., Gu, H., Larsen, R., et al. (2018). A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. Cell *174*, 465-480.e22. https://doi.org/10.1016/j.cell.2018.06.035.

55. Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat Neurosci *13*, 133–140. https://doi.org/10.1038/nn.2467.

56. He, M., Tucciarone, J., Lee, S., Nigro, M.J., Kim, Y., Levine, J.M., Kelly, S.M., Krugikov, I., Wu, P., Chen, Y., et al. (2016). Strategies and Tools for Combinatorial Targeting of GABAergic Neurons in Mouse Cerebral Cortex. Neuron *91*, 1228–1243. https://doi.org/10.1016/j.neuron.2016.08.021.

57. Gong, S., Doughty, M., Harbaugh, C.R., Cummins, A., Hatten, M.E., Heintz, N., and Gerfen, C.R. (2007). Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. J. Neurosci. *27*, 9817–9823. https://doi.org/10.1523/JNEUROSCI.2707-07.2007.

58. Bella, D.J.D., Domínguez-Iturza, N., Brown, J.R., and Arlotta, P. (2024). Making Ramón y Cajal proud: Development of cell identity and diversity in the cerebral cortex. Neuron *0*. https://doi.org/10.1016/j.neuron.2024.04.021.

59. Zeng, H. (2022). What is a cell type and how to define it? Cell *185*, 2739–2755. https://doi.org/10.1016/j.cell.2022.06.031.

60. Doyle, A., McGarry, M.P., Lee, N.A., and Lee, J.J. (2012). The Construction of Transgenic and Gene Knockout/Knockin Mouse Models of Human Disease. Transgenic Res *21*, 327–349. https://doi.org/10.1007/s11248-011-9537-3.

61. Qiu, L., Rivera-Pérez, J.A., and Xu, Z. (2011). A Non-Specific Effect Associated with Conditional Transgene Expression Based on Cre-loxP Strategy in Mice. PLoS One *6*, e18778. https://doi.org/10.1371/journal.pone.0018778.

62. Institute, A. (2023). RAISINs (RNA-seq for Profiling Intact Nuclei with Ribosome-bound mRNA) Nuclei Isolation from Mouse CNS Tissue Protocol.

63. Choi, J.-H., Yu, N.-K., Baek, G.-C., Bakes, J., Seo, D., Nam, H.J., Baek, S.H., Lim, C.-S., Lee, Y.-S., and Kaang, B.-K. (2014). Optimization of AAV expression cassettes to improve packaging capacity and transgene expression in neurons. Mol Brain *7*, 17. https://doi.org/10.1186/1756-6606-7-17.

64. Kim, J.-Y., Grunke, S.D., Levites, Y., Golde, T.E., and Jankowsky, J.L. (2014). Intracerebroventricular viral injection of the neonatal mouse brain for persistent and widespread neuronal transduction. J. Vis. Exp. *2014/09/15*, 51863.

65. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

66. Gray, L.T., Yao, Z., Nguyen, T.N., Kim, T.K., Zeng, H., and Tasic, B. (2017). Layer-specific chromatin accessibility landscapes reveal regulatory networks in adult mouse visual cortex. Elife *6*.

67. Zeng, H., Horie, K., Madisen, L., Pavlova, M.N., Gragerova, G., Rohde, A.D., Schimpf, B.A., Liang, Y., Ojala, E., Kramer, F., et al. (2008). An inducible and reversible mouse genetic rescue system. PLoS Genet. *4*, e1000069.

68. Zhu, F., Gamboa, M., Farruggio, A.P., Hippenmeyer, S., Tasic, B., Schüle, B., Chen-Tsai, Y., and Calos, M.P. (2014). DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. Nucleic Acids Res. *42*, e34.

69. George, S.H.L., Gertsenstein, M., Vintersten, K., Korets-Smith, E., Murphy, J., Stevens, M.E., Haigh, J.J., and Nagy, A. (2007). Developmental and adult phenotyping directly from mutant embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. *104*, 4455–4460.

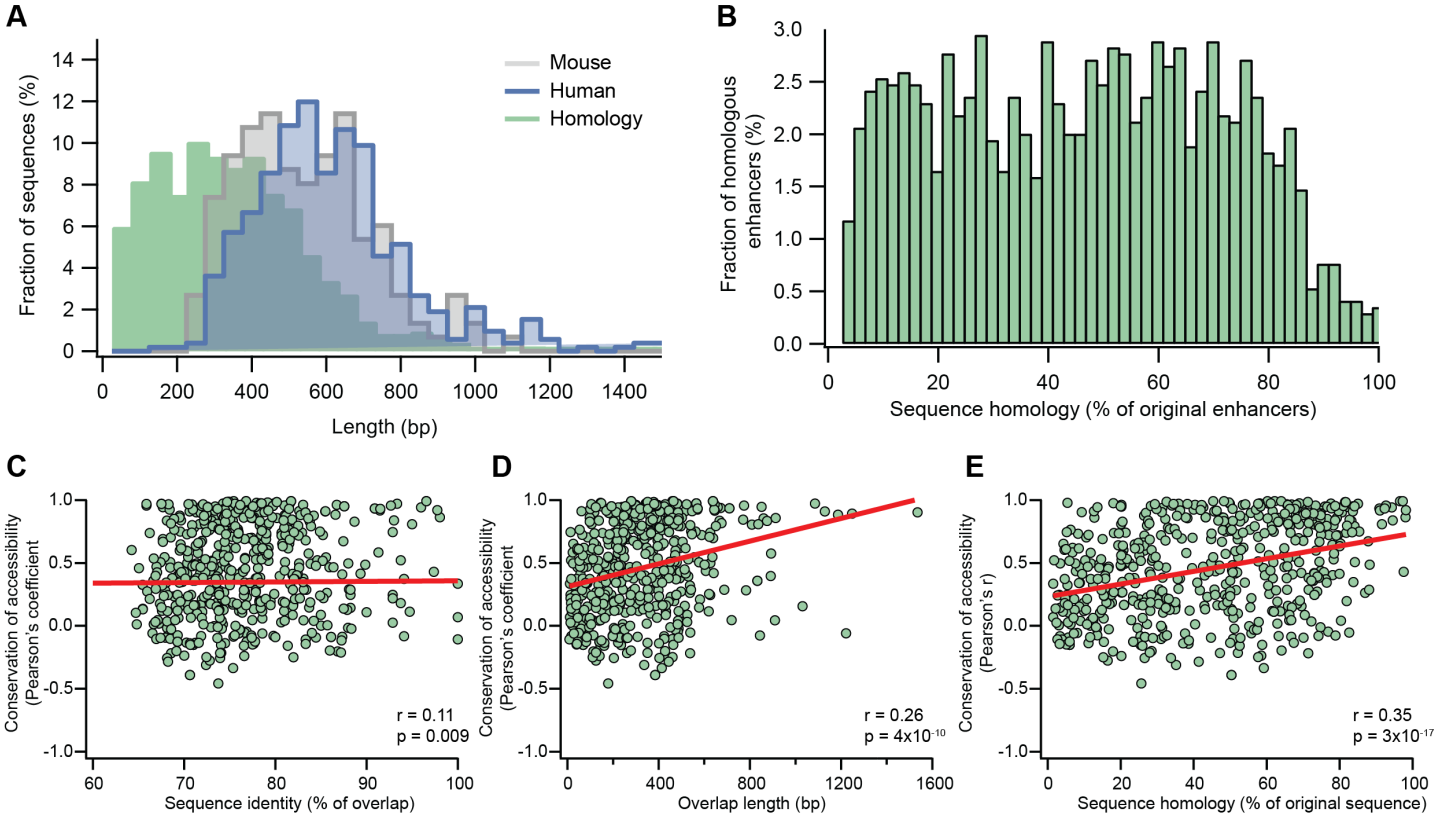


Figure S1: BLAST alignment of mouse and human liftover sequences. A. Histograms showing the proportion of mouse (grey) and human (blue) sequences, according to their length along with a histogram showing the length distribution of the liftover sequences (full green). Bin size = 50 bp. **B.** Histogram of the sequence homology, calculated as the total number of identical base pairs (liftover length x percent identity) of the total length of the original enhancer sequence. Bin size = 2%. **C-E.** Correlation analysis between the conservation of accessibility and the percent identity between the overlapping sequence only (C), overlap length (D) and the sequence homology relative to the length of the original sequence (E). Linear fit lines are shown in red, and the Pearson's coefficient with its p-value are shown at the bottom right corner of each plot.

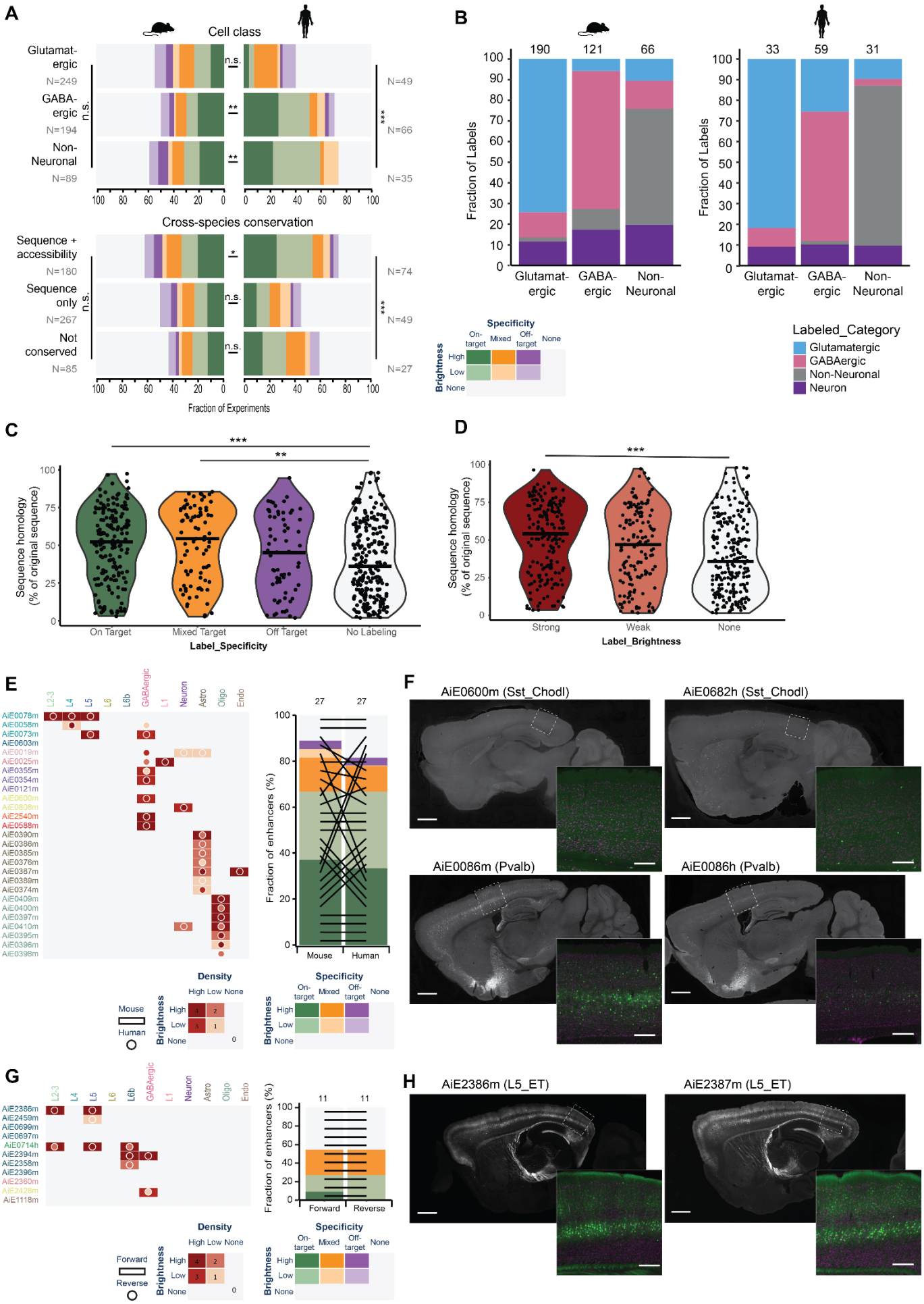

A
Cell class
Glutamat-ergic
N=249
GABA-ergic
N=194
Non-Neuronal
N=89
n.s.
**
**
***
N=49
N=66
N=35
Cross-species conservation
Sequence + accessibility
N=180
Sequence only
N=267
Not conserved
N=85
*
n.s.
n.s.
***
N=74
N=49
N=27
Fraction of Experiments
B
190
121
66
33
59
31
Fraction of Labels
Glutamat-ergic
GABA-ergic
Non-Neuronal
Specificity
On-target
Mixed
Off-target
None
Brightness
High
Low
None
Labeled_Category
Glutamatergic
GABAergic
Non-Neuronal
Neuron
C
***
**
Sequence homology (% of original sequence)
On Target
Mixed Target
Off Target
No Labeling
Label_Specificity
D
***
Sequence homology (% of original sequence)
Strong
Weak
None
Label_Brightness
E
L2-3
L4
L5
L6
L6b
GABAergic
L1
Neuron
Astro
Oligo
Endo
AiE0078m
AiE0058m
AiE0073m
AiE0603m
AiE0019m
AiE0025m
AiE0355m
AiE0354m
AiE0121m
AiE0600m
AiE0808m
AiE2540m
AiE0588m
AiE0390m
AiE0386m
AiE0385m
AiE0376m
AiE0387m
AiE0389m
AiE0374m
AiE0409m
AiE0400m
AiE0397m
AiE0410m
AiE0395m
AiE0396m
AiE0398m
27
27
Fraction of enhancers (%)
Mouse
Human
Density
High Low None
Brightness
High
Low
None
Specificity
On-target
Mixed
Off-target
None
F
AiE0600m (Sst_ChodI)
AiE0682h (Sst_ChodI)
AiE0086m (Pvalb)
AiE0086h (Pvalb)
G
AiE2386m
AiE2459m
AiE0699m
AiE0697m
AiE0714h
AiE2394m
AiE2358m
AiE2396m
AiE2360m
AiE2428m
AiE1118m
11
11
Fraction of enhancers (%)
Forward
Reverse
Density
High Low None
Brightness
High
Low
None
Specificity
On-target
Mixed
Off-target
None
H
AiE2386m (L5_ET)
AiE2387m (L5_ET)

Figure S2: Effects of cross-species conservation and sequence orientation on enhancer scoring. A. Summary plot of enhancer scoring data according to their genome of origin (top) and cross-species conservation of sequence and accessibility (bottom). **B.** Summary plot of labeled cell population class for mouse (left) and human (right) enhancers. **C-D.** Violin plots showing the degree of mouse-human sequence homology according to the different specificity (C) and brightness (D) categories. **E.** Summary plot of the scoring data for mouse (rectangles) alongside its human (circles) orthologous sequence (Left) and a summary plot of the scoring data according to the brightness and specificity, with black lines connecting each pair (Right). **F.** Representative epifluorescence images of parasagittal sections, showing labeling pattern for 2 individual mouse (left) and their human ortholog (right). An expanded view of VISp is shown alongside each section. **G** Summary plot of the scoring data for the forward (rectangles) alongside its reverse compliment (circles) sequence (Left) and a summary plot of the scoring data according to the brightness and specificity, with black lines connecting each pair (Right). **H.** Representative epifluorescence images of parasagittal sections, showing labeling pattern of a mouse enhancer in a forward (left) and reverse orientation (right). An expanded view of VISp is shown alongside each section. Scale bars for full section and expanded view = 1 and 0.2 mm, respectively. P-values lower than 0.05, 0.01 and 0.001 are denoted by *, **, ***, respectively, n.s. = not significant.

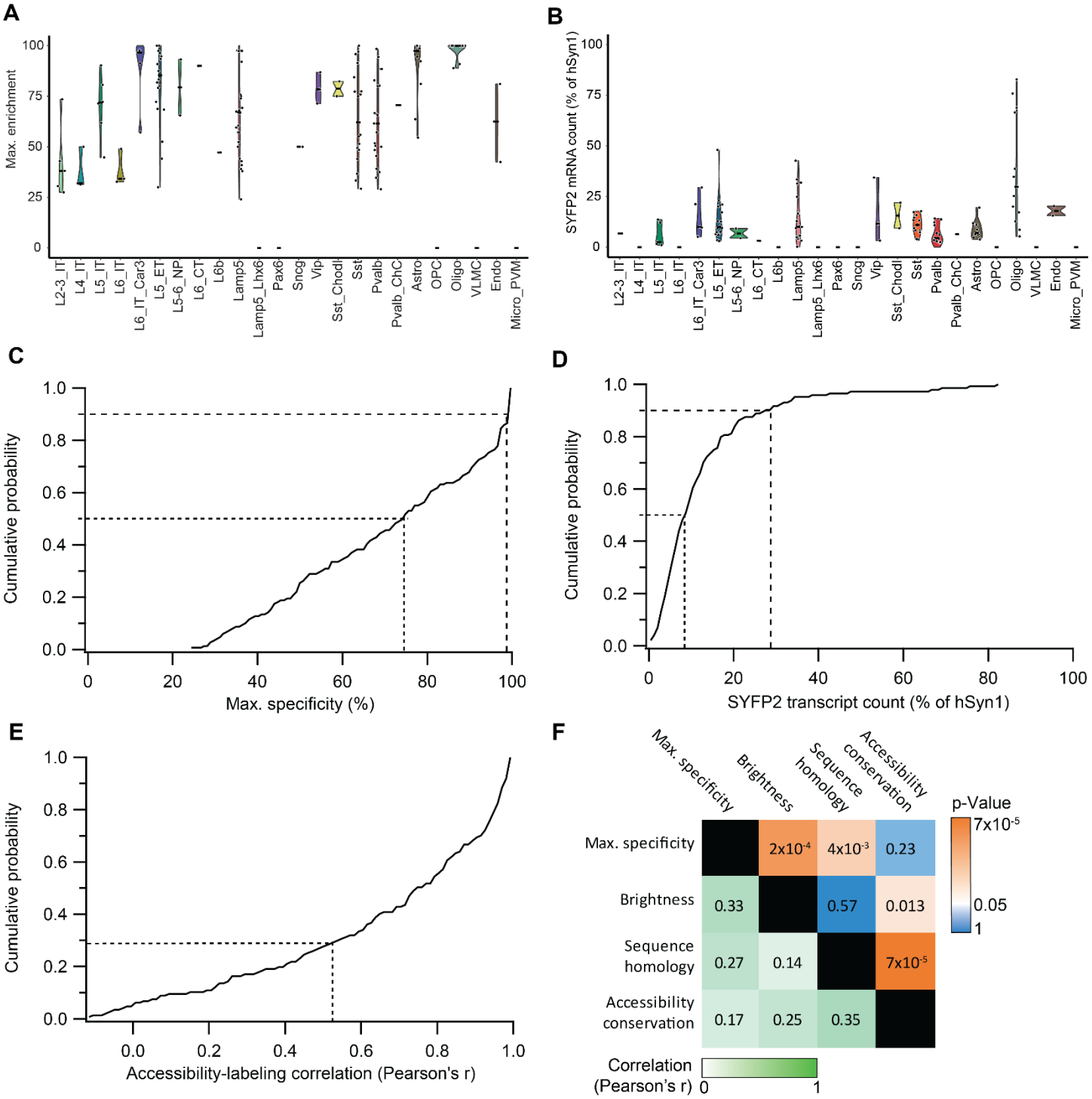


Figure S3: Distribution of specificity and brightness across enhancers. A,B. Violin plots showing the distribution of enhancer max. specificity (A) and SYFP2 mRNA counts normalized to hSyn1 (B), in each of the cortical subclasses. Experiments where the max. specificity was below 50% were excluded from the analysis in (B). **C.** A cumulative distribution plot showing the fraction of enhancers as a function of their specificity, estimated by the maximal fraction of labeled cells. **D.** A cumulative distribution plot showing the fraction of enhancers as a function of their brightness, relative to hSyn1. **E.** A cumulative distribution plot showing the fraction of enhancers as a function of correlation coefficient, between the distribution of labeled cells and distribution of chromatin accessibility, across the cortical subclasses. **F.** Cross-correlation plot showing correlation values (white-green scale, bottom left corner) and their respective p-values (blue-orange scale, top right corner). Dashed lines in plots (A-C) show the median and top 10th percentile of enhancers. P-values in D were corrected for multiple comparisons.

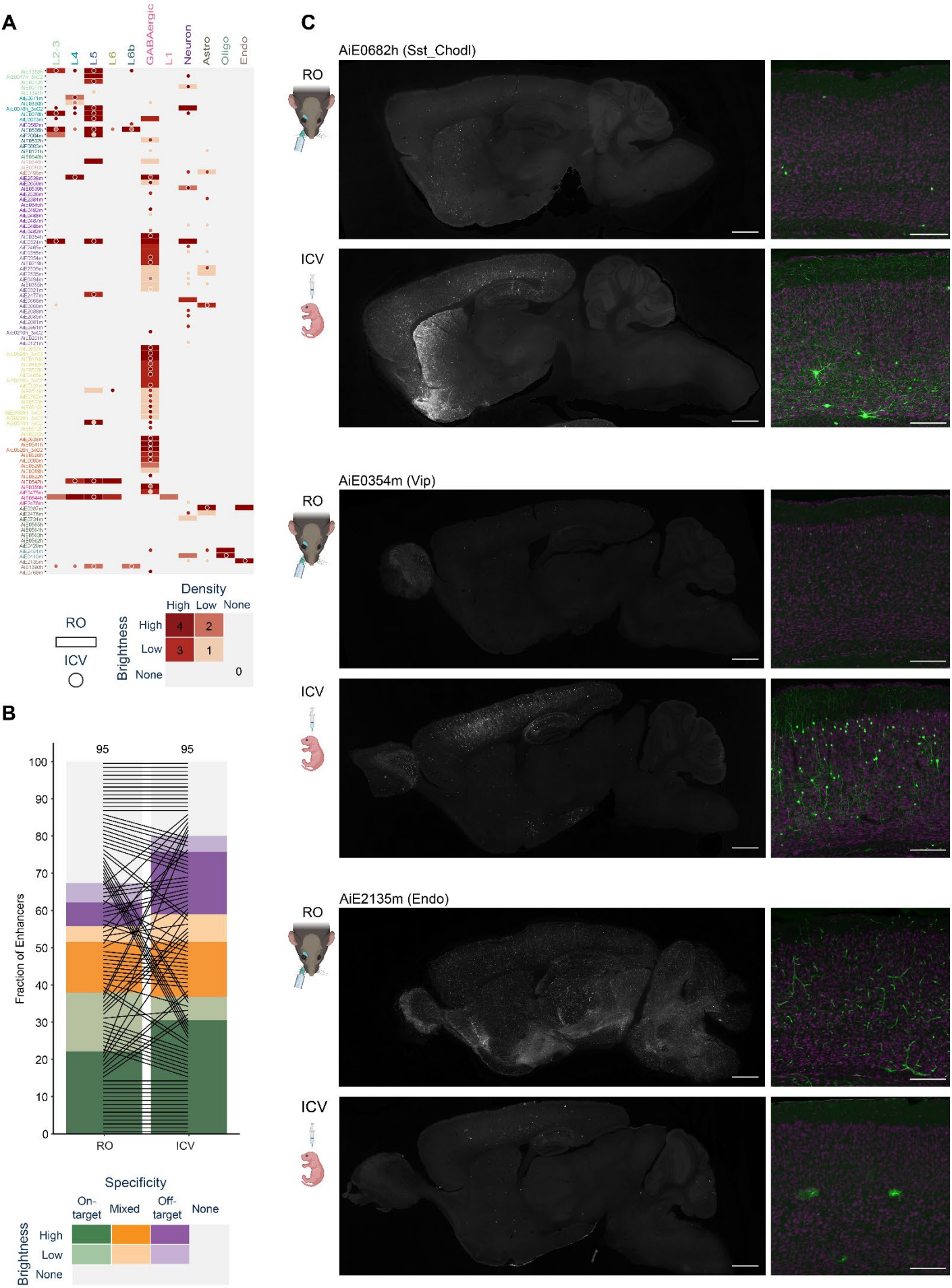

A
L2-3
L4
L5
L6
L6b
GABAergic
L1
Neuron
Astro
Oligo
Endo
Density
High
Low
None
RO
ICV
Brightness
High
Low
None
4
2
3
1
0
B
95
95
100
90
80
70
60
50
40
30
20
10
0
Fraction of Enhancers
RO
ICV
Specificity
On-target
Mixed
Off-target
None
Brightness
High
Low
None
C
AiE0682h (Sst_Chodl)
RO
ICV
AiE0354m (Vip)
RO
ICV
AiE2135m (Endo)
RO
ICV

Figure S4: Comparison of scoring data between RO and ICV viral delivery routes. A. Heatmap of scoring data for the same vectors, delivered RO (rectangles) or ICV (circles). **B.** Summary plot of the scoring data according to the brightness and specificity, with black lines connecting each pair. **C.** Representative epifluorescence images of parasagittal sections of 3 individual enhancers, comparing labeling pattern when the virus was delivered via the RO (top) or ICV (bottom) route. An expanded view of the visual cortex is displayed to the right of the full-sized image. Scale bars for full section and expanded view = 1 and 0.2 mm, respectively.

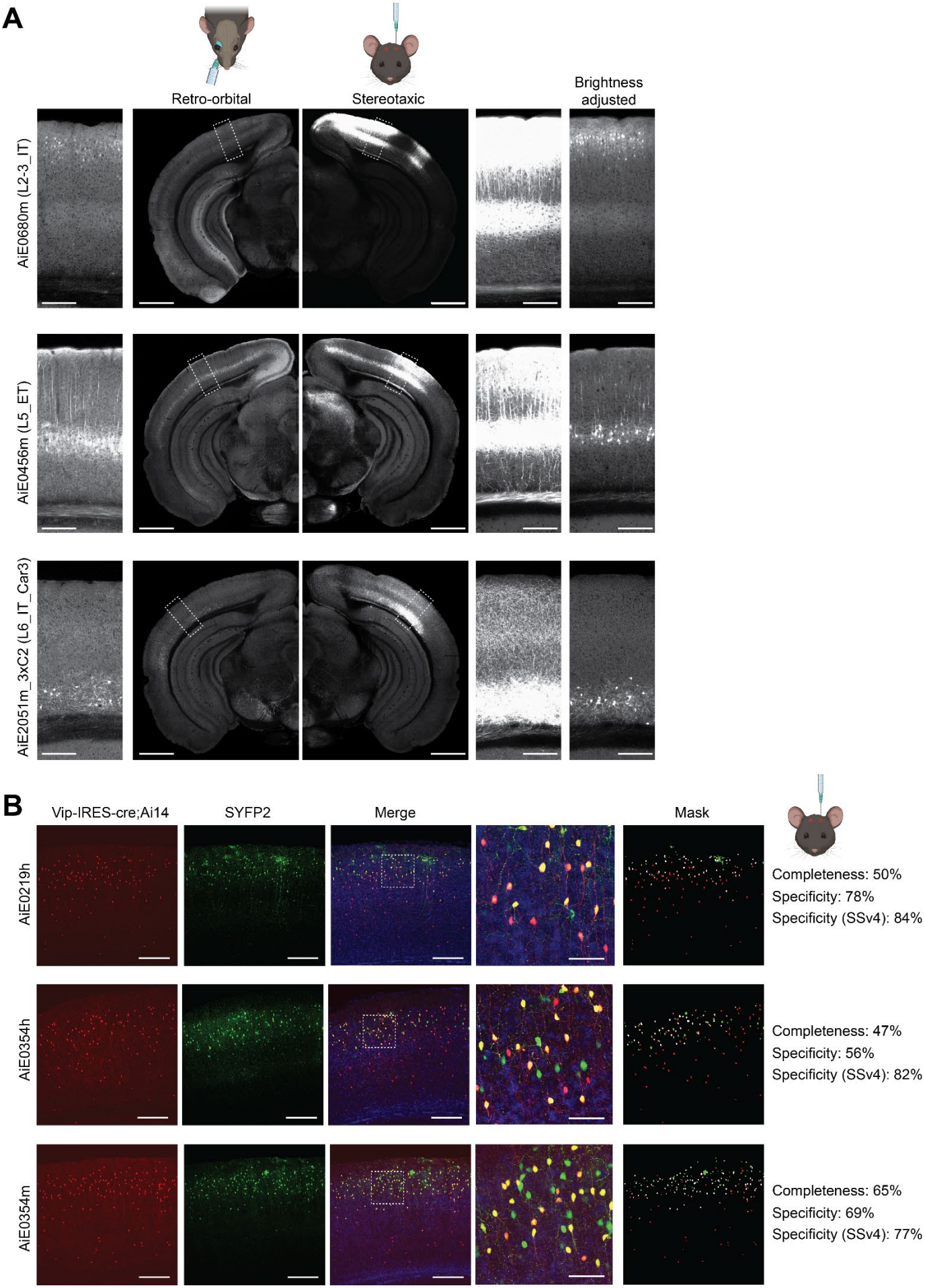

A
Retro-orbital
Stereotaxic
Brightness adjusted
AiE0680m (L2-3_IT)
AiE0456m (L5_ET)
AiE2051m_3xC2 (L6_IT_Car3)
B
Vip-IRES-cre;Ai14
SYFP2
Merge
Mask
AiE0219h
Completeness: 50%
Specificity: 78%
Specificity (SSv4): 84%
AiE0354h
Completeness: 47%
Specificity: 56%
Specificity (SSv4): 82%
AiE0354m
Completeness: 65%
Specificity: 69%
Specificity (SSv4): 77%

Figure S5: Stereotaxic delivery of enhancer AAVs. A. Stereotaxic delivery of three enhancers targeting different subclasses of glutamatergic neurons into VISp results in strong, layer restricted SYFP2 expression. Scale bars for full section and expanded view = 1 and 0.2 mm, respectively. **B.** Stereotaxic delivery of three enhancers targeting Vip interneurons, delivered to the VISp of *Vip-IRES-Cre;Ai14* double transgenic line. For each injection, completeness was calculated as the fraction of SYFP2$^+$/tdTomato$^+$ cells, of all tdTomato$^+$ cells at the injection site, and specificity was calculated as the fraction of SYFP2$^+$/tdTomato$^+$ cells, of all SYFP2$^+$ cells. Scale bars for full VISp view and expanded view = 0.2 and 0.05 mm, respectively.

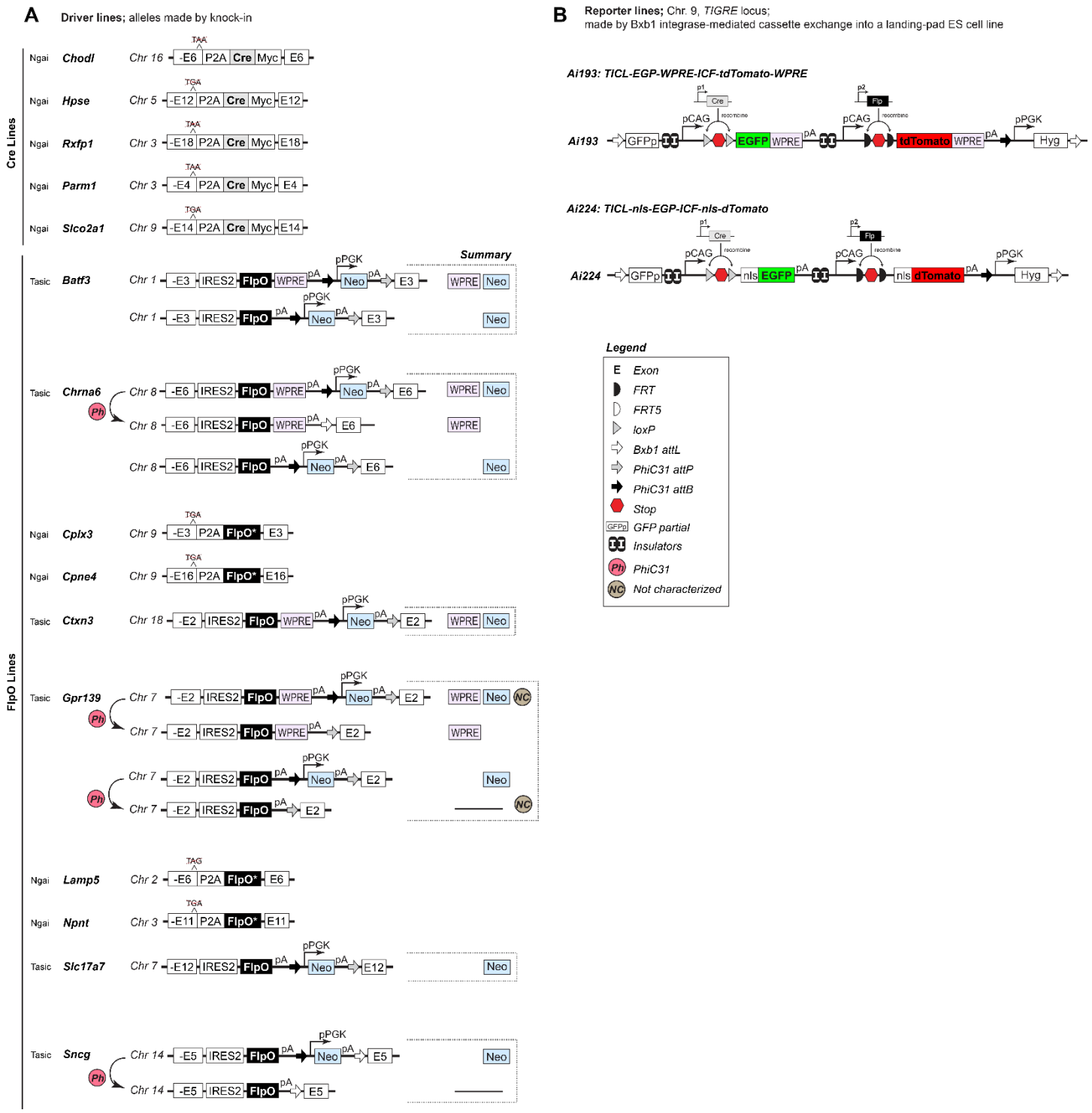


Figure S6: Transgenic line designs. A. Schematics depicting design of the 15 driver lines. Of these, 5 express Cre recombinase whereas 10 express FlpO. For some lines, such as *Chrna6-IRES2-FlpO*, we have versions with WPRE, with Neo present as well as with Neo removed allowing us to compare expression patterns in all three (**Figure S7**). In some instances, the driver lines were used as is and in others, they were crossed with *Rosa26-PhiC31* mice to delete the pPGK-neo selection cassette. **B.** Schematic depicting the design of the two new reporter mice *Ai193* and *Ai224*.

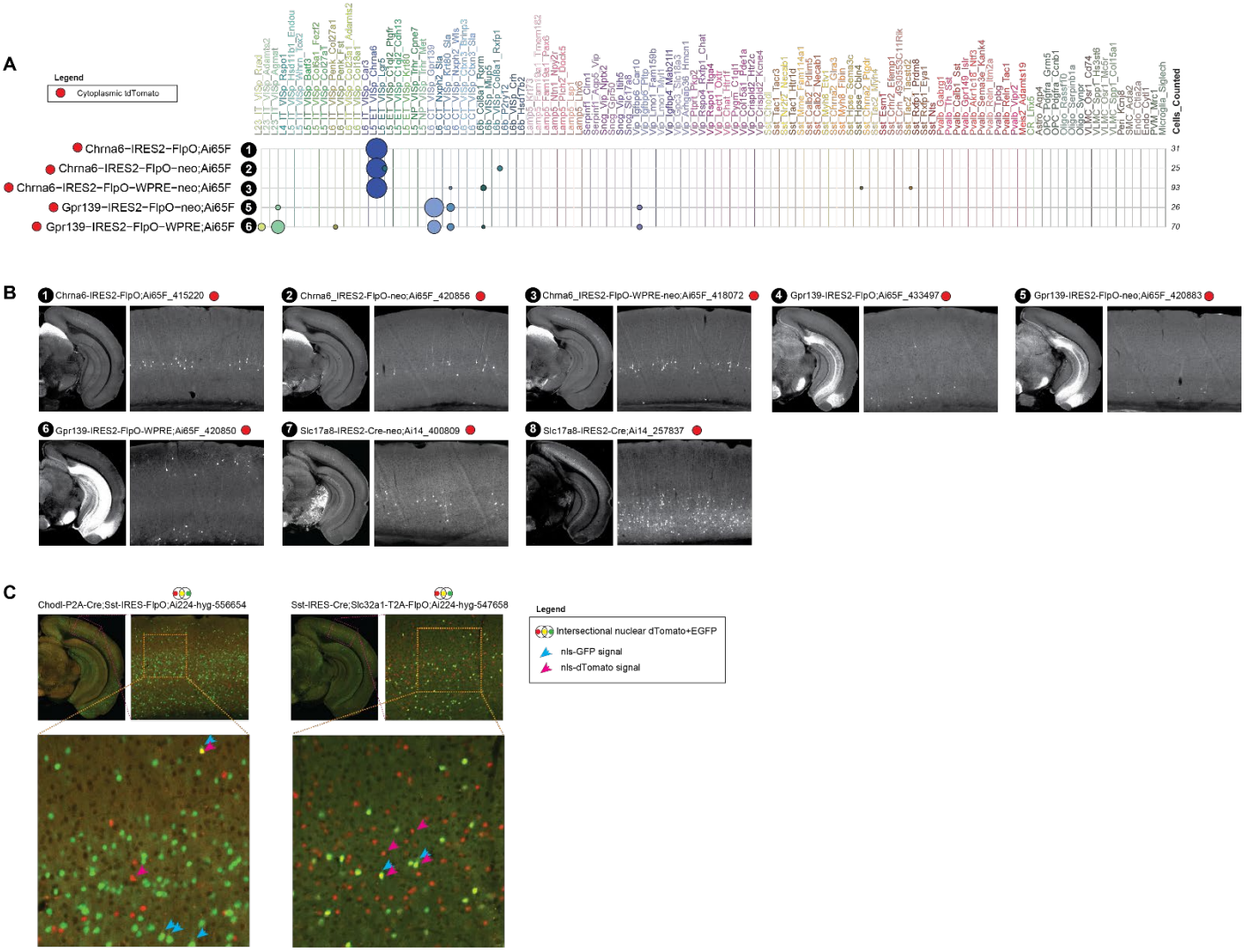


Figure S7: (A-B) The effect of WPRE and the Neo cassette on marker expression. A. scRNA-seq (SSv4) data showing distribution of labeled cells from the different lines with and without Neo or WPRE, mapped to mouse VISp taxonomy at the cluster level. **B.** Images showing differences in expression in each line. **C. Imperfect nuclear localization of nls-EGFP in *Ai224* (TICL-nls-EGFP-ICF-nls-dTomato).** Expression of nls-EGFP (Cre-dependent) and nls-dTomato (Flp-dependent) is faithful in this line; however, the localization is not perfect. The GFP appears mostly nuclear, but some weak cytoplasmic signal can also be observed (blue arrow). In comparison, nls-dTomato appears nuclear (magenta arrow).

Driver	Reporter	Labeling Selectivity (STPT)	Most represented cell type taxon SSv4 (%total cells)	Double driver	Reporter	Most represented cell type taxon SSv4 (%total cells)
Slco2a1-P2A-Cre	Ai14	Not as expected	100% Endo	Slco2a1-P2A-Cre; Slc17a7-IRES2-FlpO-neo	Ai224	N/A
Rxfp1-P2A-Cre	Ai14	Non-selective	N/A	Rxfp1-P2A-Cre; Slc17a7-IRES2-FlpO-neo	Ai65	43% L6_IT_VISp_Car3
Npnt-P2A-FlpO	Ai193	More selective than expected	95% L2-3_IT	NA	NA	N/A
Parm1-P2A-Cre	Ai14	Non-selective	N/A	Parm1-P2A-Cre; Slc17a7-IRES2-FlpO-neo	Ai65	N
Sncg-IRES2-FlpO-neo	Ai65F	As expected	79% Sncg	Slc32a1-IRES-Cre; Sncg-IRES2-FlpO	Ai65	87% Sncg
				Slc17a8-IRES2-Cre; Sncg-IRES2-FlpO	Ai65	80% L5-6_NP
Slc17a7-IRES2-FlpO-neo	Ai65F	As expected	100% Glutamatergic	N/A	N/A	N/A
Gpr139-IRES2-FlpO-WPRE	Ai65F	Not as expected	45% L2-3_IT; 45% L6_CT	Ntsr1-Cre_GN220; Gpr139-IRES2-FlpO-WPRE	Ai65	94% L6_CT_VISp_Gpr139
Gpr139-IRES2-FlpO-neo	Ai65F	As expected	92% L6_CT	N/A	N/A	N/A
Ctxn3-IRES2-FlpO-neo	Ai65F	As expected	38% L6_IT_Car3; 29% L6_CT	Ntsr1-Cre_GN220;Ctxn3-IRES2-FlpO-neo	Ai65	77% L6_CT_VISp_Ctxn3_Brinp3; 23% L6_CT_VISp_Ctxn3_Sla
Chrna6-IRES2-FlpO-WPRE-neo	Ai65F	As expected	91% L5_ET	N/A	N/A	Y
Chrna6-IRES2-FlpO-neo	Ai65F	As expected	96% L5_ET	N/A	N/A	N/A
Chrna6-IRES2-FlpO	Ai65F	As expected	100% L5_ET	N/A	N/A	N/A
Lamp5-P2A-FlpO	RCF-EGFP	Less selective than expected	52% L2-3_IT; 3% Lamp5	Slc32a1-IRES-Cre; Lamp5-P2A-FlpO	Ai65	94% Lamp5
Cpne4-P2A-FlpO	Ai65F	Non selective	51% L4-IT	N/A	N/A	N/A
Cplx3-P2A-FlpO	Ai65F	Less selective than expected	45% L6b	N/A	N/A	N/A
Batf3-IRES2-FlpO-WPRE-neo	Ai65F	Less selective than expected	59% L5-IT	N/A	N/A	N/A
Batf3-IRES2-FlpO-neo	Ai65F	Not as expected	67% in L5-IT	N/A	N/A	N/A
Hpse-P2A-Cre	Ai65L	Less selective than expected	16% Sst_Hpse_Cbln4 + Sst_Hpse_Sema3c	Hpse-P2A-Cre; Slc32a1-T2A-FlpO	Ai65	57% Sst_Hpse_Cbln4 + Sst_Hpse_Sema3c
				Hpse-P2A-Cre; Sst-IRES-FlpO	Ai65	N/A
Chodl-P2A-Cre	Ai14	Less selective than expected	NA	Chodl-P2A-Cre; Sst-IRES-FlpO	Ai65	90% Sst_Chodl

Table S1: Table summarizing the evaluation of the 15 new driver lines. Each line was crossed with a reporter and evaluated by STPT. It was evaluated further by SSV4 only if the expression was as expected or, broader than expected. A second driver line was selected based on the expression pattern and the triple transgenic was similarly evaluated by STPT and SSV4. The most represented cell type is shown as a percentage of total cells labelled.