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] Jazmin Garcia,[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] Emily Kussick,[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] Eric Szelenyi,[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[1,5,*]

[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
[5]Lead contact

[*]Correspondence: bosiljkat@alleninstitute.org

Summary

The mammalian cortex is comprised of cells classified into types according to shared properties. Defining the contribution of each cell type to the processes guided by the cortex is essential for understanding its function in health and disease. We used transcriptomic and epigenomic cortical cell type taxonomies from mouse and human to define marker genes and putative enhancers and created a large toolkit of transgenic lines and enhancer AAVs for selective targeting of cortical cell populations. We report evaluation of fifteen new transgenic driver lines, two new reporter lines, and >800 different enhancer AAVs covering most subclasses of cortical cells. The tools reported here as well as the scaled process of tool creation and modification enable diverse experimental strategies towards understanding mammalian cortex and brain function.

Keywords

Epigenetics, Transcriptomics, Enhancer, Cell type, Neocortex, AAV, Transgenic, Recombinase, Mouse, Human

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[1]. One approach is to gain experimental access to each building block (e.g., a cell type) and manipulate it to observe effects on an animal's physiology and behavior[2–4]. 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. Numerous valuable tools have been developed and utilized to understand cell type functions in the brain, though they may not always achieve the desired level of cell-type resolution[2,3,5,6].

Advances in next generation nucleic acid sequencing and machine learning have transformed cell classification based on single-cell genomics (transcriptomics, epigenomics, etc.)[7–16]. These new methods enable measurements of tens of thousands of molecular properties in individual cells and their subsequent classification into molecularly defined cell types. Cell type taxonomies have hierarchical organization, with cell type taxons of increased granularity frequently referred to as classes, subclasses, supertypes, and types/clusters[9,16]. Single-cell genomics also enables cell type-congruent discovery of marker genes and putative enhancers at all levels of the taxonomy[8,9,11,13,17–21]. These marker genes and enhancers can be used to make tools for precision cell type access, that can be employed to test the correspondence of molecularly defined cell types to 'functional' cell types[3,22–24] (**Figure 1A**).

Here, we relied on historical genomics data[17,18,25–27], and newly-generated single-nucleus multiome data (snMultiome, 10x Genomics, consisting of joint snRNA-seq and snATAC-seq from each nucleus) to create a transcriptomic and epigenomic cross-species (mouse/human) taxonomy consistent with previously published datasets[17,18,25–27]. The newly generated snMultiome data were generated from the adult mouse cortex (somatosensory, motor, and visual areas), whereas the human cortical dataset was derived from the adult middle temporal gyrus[28]. 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 examined a total of 682 putative enhancers sequences: 599 novel and 83 that have been previously reported[17–19,29–31]. We tested these in mice by systemic introduction (retro-orbital injection) of individual enhancer AAVs expressing a fluorescent protein, employing the blood-brain-barrier penetrating PHP.eB capsid[32–34]. For initial evaluation of the enhancer AAVs, we visually examined and scored epifluorescence images of brain sections. 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 attempted to optimize expression from these vectors by enhancer core bashing and concatenation (142 vectors), or expression of recombinase versions that can be combined with reporter mice (39 vectors)[17–19,35,36]. In total, we examined 863 enhancer AAVs in this study. Lastly, 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**).

In summary, we created and characterized a suite of genetic tools and compared them to select existing ones using a common battery of imaging and molecular techniques to provide detailed examination of their specificity and recommendations for their utilization (**Figure 1**). We are in the process of providing the tools through public repositories (Addgene and Jackson Labs) as well as adding data and metadata for all reagents (**Tables S1-S5**) to a new Allen Institute public web resource, the Genetic Tools Atlas (RRID:SCR_025643; https://portal.brain-map.org/genetic-tools/genetic-tools-atlas). These freely available public resources will enable scientists to access the information on these tools and select the best tool for their use.

Results

Selection of putative enhancer sequences from mouse and human genomics data

We selected putative enhancers based on a variety of data and diverse criteria over a span of several years. To identify putative enhancers, we used various chromatin accessibility genomics data for mouse and human cortex, some published[17,18,25–27], some newly created and analyzed (see below). Most putative enhancer selection was based on 1) high and differential chromatin accessibility across the cortical cell subclasses and 2) proximity to marker genes. Putative enhancer sequence conservation across species was not a generally used criterion.

To summarize all the chromatin accessibility data across cell types and represent them in a unified way, as well as select more putative enhancers, we generated a new cortical snMultiome (10x Genomics, bimodal single-nucleus RNA-seq and ATAC-seq) dataset of single nuclei isolated from adult mouse visual, somatosensory, and motor cortices (82,654 nuclei post-QC; **Figure 2A**). We also employed a recently-published snMultiome dataset of nuclei isolated from adult human middle temporal gyrus (MTG, 83,977 nuclei post-QC; **Figure 2A**, right)[28]. For mouse, we mapped the single-nucleus transcriptomes to the recently-published whole mouse brain taxonomy (AIT21) using hierarchical approximate nearest neighbor (HANN) mapping[16]. For the human MTG data, single-nucleus transcriptomes were mapped to the great apes taxonomy with scANVI[37,38]. We compared the great ape taxonomy and a previously published mouse cortical taxonomy[9] 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 (cell subclass). This annotation was performed on the transcriptomic part of the multiome datasets and it resulted in comparable cell populations across mouse and human datasets; similar to what has been described before[37,39].

The transcriptome-based annotation for each nucleus was used to label its respective ATAC-seq profile for downstream analysis. Pseudobulk ATAC-seq data were generated with ArchR[40] and used to identify putative enhancers as differentially-accessible peaks across subclasses by MACS2 and ArchR[40,41]. As an example, we show gene expression data for the *Nos1* gene and the accessibility of two orthologous enhancers targeting Sst_Chodl cell type near this gene in the mouse and human datasets (**Figures 2A**, **2B**).

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**, **S1A**). Each putative enhancer sequence was assigned a unique ID, consisting of the initials AiE for enhancers discovered at the Allen Institute, or ExE for enhancers previously reported by other groups[29,30], followed by a four-digit number and a species indicator ("m" or "h" for mouse or human, respectively). To provide consistency to the enhancer collection, we applied the same naming structure to the previously published enhancers from Allen Institute[17–19] and provide the original names as aliases (**Table S4**). Additionally, each enhancer was associated with one or two target cell populations (TCP), which allow us to predict which population/s they are most likely to be expressed in. 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 also evaluated the degree of cross-species sequence conservation using the UCSC genome browser LiftOver tool[42], followed by BLAST alignment, and correlated the chromatin accessibility pattern of the orthologous pairs, across the cortical subclasses (**Figure 2C**). Most of the selected candidate enhancer sequences were conserved between the two species (n = 554, 81.2%, **Figures 2C, S1A-B**). Among this conserved subset, 59.2% (n =178) of mouse and 39.0% (n = 72) of human enhancers also displayed conserved chromatin accessibility across species (**Figure 2C**). This dataset also included 27 tested 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[17,18], upstream of a minimal promoter (beta-globin, rho or CMV), driving expression of the yellow fluorescent protein, SYFP2[43]. Most plasmids contained the beta-globin minimal promoter (MinBG; 94.6%). The plasmids were packaged into blood-brain-barrier-penetrating, PHP.eB-pseudotyped, AAV vectors[33,34] and delivered retro-orbitally (RO) to ~30 day-old wild-type mice of C57BL/6J background (average age ± SD = 30.8 ± 5.0 days, range = 27-70 days; **Methods**)[44]. 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 by an epifluorescence microscope (**Figure 3A**). We evaluated each experiment by visual inspection and categorized labeled cell populations (LCPs) in the cortex into 11 visually distinguishable categories (**Figure 3B**). 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; **Figures 3B, S2A**, **S2B**). Although we scored each LCP for both parameters, we considered brightness to be a more relevant feature of the enhancer than the density, as this factor is more easily comparable across different populations of labeled cells, and is more routinely used to evaluate enhancer strength[45–47]. In addition to annotating the neocortical populations, we also scored all labeled regions, brain-wide, using Allen Mouse Brain Reference Atlas naming scheme[48] as these enhancer AAVs could be useful tools for other brain regions. For brevity, we are reporting only the neocortical labeling patterns here. The Allen Genetic Tools Atlas web portal (in preparation) will contain labeling scores for all tools and all brain regions.

Since cell type distribution and cell shapes across neocortical layers for all TCPs are known[9,49,50], we visually matched each observed LCP with the expected pattern of the TCP to derive initial evaluation on target specificity. Following this TCP-LCP matching, 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 3C, 3D**). Multiple TCPs can belong to a single LCP, for example, in the case of different subclasses of GABAergic neurons. Therefore, this level of analysis can only provide putative on-target specificity for groups clearly differentiable by visual inspection. However, as presented below, this scoring scheme identified many highly specific enhancers and provided a nomination path for their additional characterization. Of the 682 unique enhancers screened, ~55% produced labeling in the cortex, ~43% were putatively on-target or had mixed (on- and off-target) labeling, and ~30% were putatively on-target (**Figure 3E**).

We also used the scoring data to examine the distributions of scores across cell subclasses and species, as well as the degree of sequence/accessibility conservation (**Figure 3F, S3A – S3D**). For human enhancers, we detect a statistically significant lower-on target rate for glutamatergic cell class (**Figure S3A**, top). We observe a similar trend for mouse enhancers, but it did not reach statistical significance (**Figure S3A**, top). We also observe a higher on-target rate for enhancers with conserved sequence and accessibility (**Figure S3A**, bottom). Specifically, we note that the precise degree of conservation, measured by the fraction of overlapping base pairs (**Figure S1E**) strongly affects both the specificity and brightness of the enhancer (**Figure S3B**). When comparing orthologous pairs, no significant differences in score distribution were observed overall (**Figures S3C**). However, high variability within the categories was noted, showing that the differences in orthologous sequences can affect enhancer specificity and strength (**Figure 3D**). We also found that orientation of enhancers in the context of AAV appears to have no effect on their performance, (**Figures S3E, S3F**), as has previously been observed for the SV40 enhancer[51].

In summary, we describe 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 at the subclass cell type level, when evaluated in the mouse brain, with roughly 30% 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 (scRNA-seq) with Smart-seq v4 (SSv4) and/or serial two-photon tomography (STPT)[52] (**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 isolated SYFP2(+)/DAPI(-) cells with FACS (**Figure S4A**). These cells were subjected to scRNA-seq and their transcriptomes were mapped to the mouse VISp cortical taxonomy (**Methods**). These data enabled us to quantify the fraction of cells belonging to any cell type for each experiment (**Figures 4B**, **4C**, **S4B**, **S4C**). In addition, we quantified the expression level of the SYFP mRNA (median SYFP2 mRNA count per million transcripts) and normalized it for each cell type to the median expression of *hSyn1*-promoter/enhancer-driven SYFP2 mRNA (**Figure 4B, S4D**).

We find that our initial screen produced many highly specific enhancer AAVs: more than 50% of the enhancers examined with scRNA-seq exhibited target specificity of >70% at the subclass cell-type level and 28% were >90% specific (**Figure S4C**). However, we note that SYFP2 mRNA expression driven by the enhancer AAVs was relatively low for most of the enhancers, when compared with the pan-neuronal *hSyn1*-promoter/enhancer (average ratio ± SD = 13.3 ± 15.3% of hSyn1; **Figure S4D**). We also observed that ~29% of enhancers exhibited labeling distribution patterns that were not significantly correlated with the distribution of accessibility across subclasses (**Figure S4E**). This unexpected result could be partially explained by the way cells were collected and sorted for sequencing: To avoid false negatives in FACS, we set the sorting gates relatively stringently, to collect the brightest cells (**Figure S4A**)[19]. However, this can lead to biases in the results, in cases where two or more populations are labeled, with the more abundant population being dimmer than the smaller one, as in the case of AiE2543m (**Figure 4D**). Note that evaluation of specificity by scRNA-seq of FACS-isolated cells can also lead to biases in cases where labeled population includes cell types particularly sensitive to the dissociation and sorting process (see below)[8,9].

For enhancers with putative on-target labeling patterns, we generated whole-brain STPT image sets to further interrogate the enhancer AAV labeling patterns. Compared to five epifluorescence images, STPT enables more complete evaluation of the entire cortex. STPT also examines all other parts of the brain and shows if the enhancer is cortex-specific or active in other brain regions. Finally, brain-wide visualization reveals axonal projection patterns and can help distinguish between different types of cortical projection neurons. We examined ~34% (180/532) of mouse enhancers and ~41% (62/150) human enhancers for a total of 242 STPT whole-brain image datasets (representative examples in **Figure 4D**). These data complement the SSv4 data (**Figure 4B**), as well as confirm and extend primary screen data, as they cover the whole brain with a series of coronal images. In all cases, we find that the STPT data are in full agreement with the labeling patterns observed in the primary screen.

We cross-correlated the specificity of each enhancer, measured as the maximal fraction of labeled cells, with the SYFP2 transcript count and cross-species conservation of sequence and accessibility (**Figure S4F**). We find 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 S4F**). It should be noted that this analysis was performed only on the SSv4 data, which comprise only enhancers with demonstrated putative on-target specificity in our primary screen, and therefore is not representative of the entire collection.

These analyses highlighted enhancers which delivered high specificity as potentially useful tools for selective targeting of cortical cell types. The relevant experimental data for SYFP2-expressing vectors in this study were organized in supplementary material (**Documents S1-S5**) and will be available at the Genetic Tools Atlas web portal (RRID:SCR_025643; https://portal.brain-map.org/genetic-tools/genetic-tools-atlas).

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[53–56]. We proceeded to fragment a subset of enhancers into three putative “cores” (C1, C2, and C3; average length ± SD = 212 ± 75 bp) that tile the originally-examined sequence with 50 bp overlap[17]. 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 have previously shown that enhancer concatenation can increase reporter expression[17–19]. 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[57,58].

We core-bashed a total of 82 enhancers, creating 3xC 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 occasionally labeled additional cell populations, which were not labeled by the original enhancer (**Figures 5C, 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. To further compare the select bashed versions with the original enhancers, we collected SSv4 scRNA-seq data from VISp. We compared the labeling specificity and SYFP2 mRNA expression levels of the concatenated bashed enhancer with the original enhancer. The behavior of individual enhancers varied, but on average, we observed ~3-fold increase in SYFP2 mRNA count for the bashed 3x concatenated enhancer (average ratio ± SD = 274 ± 217% of full-length enhancer) with a slight reduction in the specificity (average ratio ± SD = 91 ± 30% of full-length enhancer; **Figures 5E, 5F**). This is in agreement with previous studies that show that enhancers can support target gene expression in additive fashion, as well as sub- and supra-additive fashion depending on the enhancers themselves[59,60].

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[61] and removed the WPRE sequence in a subset of enhancer vectors. The latter two choices were implemented to counter loss of specificity that was more frequently observed in constructs with iCre and WPRE compared to the original constructs with fluorescent proteins, as reported previously[22]. We evaluated the recombination-mediated reporter labeling and compared it to the original pattern defined by constructs with SYFP2 (**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**).

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[62] (**Figures S5A – S5C**). 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[62]. 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 S5C**), which is consistent with previous observations where alternative approaches were tested to target this population[21,33,63]. It has also been reported that certain cell subclasses such as VLMCs, OPCs and microglia have virtually absent transduction with PhP.eB-serotyped AAVs and hence are difficult to target via our current screening strategy, which exclusively uses the PhP.eB capsid for all vectors[64]. All the vectors designed for these cell subclasses led to off-target/mixed-target labeling (**Figure 3F**) and were not further investigated by SSv4.

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 (**Figures S6A, S6B**). We also note that enhancers that exhibited weak fluorescence when delivered RO displayed increased signal intensity when delivered stereotaxically, likely due to higher multiplicity of infection (MOI) at the target region (**Figure S6A**). Finally, we provide a summary of 'hall-of-fame' enhancer tools with highest available strength and specificity of labeling for cell subclasses or clusters (**Table S5**).

New transgenic lines targeting cortical cell types

Based on the transcriptomic taxonomies of cell types in the mouse cortex[8,9,13], we selected marker genes that could label specific cell type taxons (subclasses, supertypes or clusters). 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 S7A, 7, 8**). We also made two new reporter lines that address the unmet need for Cre "AND/OR" Flp reporters: *Ai193*[65] and *Ai224*[66] (**Figure S7B**).

Initial characterization for the new driver lines was performed by crossing to a fluorescent reporter line (*Ai14* for Cre, *Ai65F* for Flp transgenes, respectively)[67,68] and subsequent STPT on the whole mouse brain. We examined labeling of cortical cells for expected cell distribution (**Table S1**). Most driver lines produced expected labeling based on cell types expressing the targeted marker gene and were further analyzed by single-cell transcriptomic profiling (SSv4) of fluorescently labeled FACS-isolated cells. Others that produced broad labeling were not analyzed by SSv4. The single-cell transcriptomes were then mapped to the VISp taxonomy as above[9] (**Figures 7, 8**).

Many lines were included into intersectional crosses to examine if additional specificity can be achieved up to the level of single transcriptomic cell-type cluster. The intersectional crosses were performed by first making double recombinase transgenics (Cre and Flp; **Figure 1A**), followed by crossing to a dual "AND" recombinase reporter (e.g., *Ai65* or *RCFL-H2B-GFP*)[69,70]. This crossing scheme prevents unintended recombination and permanent reporter modification in the germline with certain recombinase lines, if they are crossed to a reporter first, and then to the second recombinase[71]. The triple transgenic mice were examined by STPT and SSv4 (**Figures 7, 8**). In most cases, the use of triple transgenics refined reporter expression and resulted in greater specificity (**Table S1).** 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-type labeling (**Figures 8B$_{29+31}$**, **8B$_{30+31}$**, **8C$_{39+40}$**) and had observed expected overlap between the two labeling patterns.

Some cell type taxons, such as the IT glutamatergic subclasses and clusters, are difficult to target with high specificity (**Figure 7A**), perhaps because many markers for IT types tend to be continuously expressed across related IT clusters and subclasses[9,72]. This appears to be true as well for the enhancer viruses that target these populations (**Figure 4B**). A notable exception is the L6_IT_Car3 cluster, which is labeled by several unique markers in addition to being specifically targeted by several enhancer viruses (**Figures 4B**, **7A$_{11}$**, **7B$_{11}$**).

We successfully targeted L5_ET types by viruses at both subclass and cluster level (**Figures 4B**, **S4B**, **7A$_{16,17}$**, **7C$_{16,17}$**). We have also generated a specific transgenic line targeting the *Chrna6* gene that labels the unique L5_ET_Chrna6 cell type (**Figures 7A$_{15}$**, **7C$_{15}$**). For the L5_NP subclass, we have previously profiled transgenic lines based on the *Slc17a8* gene (*Slc17a8-iCre* and *Slc17a8-IRES-Cre*)[9,73]. Here, we report a highly specific enhancer virus for the L5_NP subclass (**Figures 4B**, **7A$_{19}$**, **7C$_{19}$**).

For L6_CT and L6b types, we show that a judicious use of crosses can result in nearly single cell-type cluster labeling (**Table S1).** For example, *Ctxn3-IRES2-FlpO* labels an assortment of glutamatergic IT cell types, as well as select L6_CT clusters (**Figures 7A$_{13}$**, **7B$_{13}$**). Crossing this line with a L6_CT-specific line, *Ntsr1-Cre_GN220*[74] (**Figures 7A$_{26}$**, **7D$_{26}$**), 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$_{24}$**, **7B$_{24}$**; **Table S7**). Similarly, we generated a tool that can exhibit a remarkably specific pattern of expression at the level of single L6_CT cluster in a triple transgenic cross. *Gpr139-IRES2-FlpO*, a line based on *Gpr139*, is expressed in several glutamatergic clusters in different layers, including a specific L6_CT cluster. When this line is crossed with *Ntsr1-Cre_GN220* and *Ai65*, we find that it almost exclusively labels only the L6_CT_VISp_Gpr139 cluster (**Figures 7A$_{22\text{-}23}$**, **7D$_{22\text{-}23}$**).

The two new reporters we made, *Ai193*[65] and *Ai224*[66] (**Figure S7B**), report Cre "AND/OR" Flp labeling, unlike *Ai65*[70], which requires both Cre "AND" Flp to express a fluorescent protein (FP, tdTomato in this case). These two lines have separate transcription units that independently report the presence of Cre or Flp, resulting in GFP expression for Cre, tdTomato (or dTomato) expression for Flp, and both FPs when Cre and Flp are present. *Ai193* expresses the FPs in the cytoplasm (**Figure 7E**), whereas *Ai224* expresses them in the nucleus

(**Figure 7F**). When compared with previously widely used reporters (*Ai14*, *Ai65* or *Ai65F*; **Figure S8D**), both lines appear to faithfully report Cre and Flp recombinase expression by expressing either one or both marker genes. Note that the localization of GFP in *Ai224*, while predominantly nuclear, is imperfect (**Figure S8F**).

For GABAergic cell types, we selected the *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 7A$_4$**, **7B$_4$**, **8A$_4$**, **8B$_4$**). To exclude expression in glutamatergic cells, we generated a triple transgenic containing the pan-GABAergic line *Slc32a1-IRES-Cre*[75]. We show that *Lamp5-2A-FlpO;Slc32a1-IRES-Cre*;*Ai65* exclusively labels the Lamp5 subclass of GABAergic interneurons in the cortex (**Figures 8A$_{31}$**, **8B$_{31}$**). We also find that this subclass can be very specifically labeled by many enhancer viruses (**Figures 8A$_{29\text{-}30}$**, **8B$_{29\text{-}30}$**). 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$_{33}$**, **8B$_{33}$**). The generation of the triple transgenic *Sncg-IRES-FlpO;Slc32a1-IRES-Cre*;*Ai65* experimental animals excluded the endothelial cells, labeling only the Sncg GABAergic cells (**Figures 8A$_{34}$**, **8B$_{34}$**). In addition, we were able to use the Sncg line to label an individual cell type cluster, Sncg_Slc17a8, a unique Sncg type enriched in the frontal cortex[9]. To achieve this, we crossed this line with the *Slc17a8-IRES-Cre* driver line[73] followed by *Ai65*. We observed exquisitely specific labeling for the Sncg_Slc17a8 cluster, as shown by SSv4 of cells isolated from the Anterior Lateral Motor (ALM) cortex of the triple transgenic *Sncg-IRES-FlpO;Slc17a8-IRES-Cre*;*Ai65* mice (**Figures 8A$_{35\text{-}36}$**, **8B$_{36}$**). Note that the new Sncg line is useful beyond cortex: when combined with reporter viruses, we used it to characterize the Sncg GABAergic neurons in the hippocampus[76].

To gain access to the Sst_Chodl cell type cluster, which corresponds to sleep-active *Nos1* and *Tacr1*-expressing cells[77], we made a Cre line that targets the gene *Chodl* (*Chodl-P2A-Cre*). In initial characterization, we found that this line also labels other cell types (**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 8C$_{39}$**). We also discovered two enhancer viruses that quite specifically label this cell type (**Figures 4B,C, 8C$_{40}$**). We note that one of the enhancers (AiE0600m) is located ~250 kbp from the *Nos1* gene (**Figure 2A**). The previously described *Nos1-CreERT2*[78] also labels Sst_Chodl, but not cleanly (**Figure 8C$_{37\text{-}38}$**). However, *Nos1-CreERT2;Sst-IRES-FlpO;Ai65* quite specifically labels the Sst_Chodl cell type (**Figure 8C$_{37\text{-}38}$**).

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 cluster in the Sst 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 cluster labeling (**Figure 8C$_{45\text{-}46}$**; **Table S1**). Therefore, each cross needs to be tested before assuming that the tools based on the marker genes will provide expected labeling.

We note that the exact components used to create a transgenic line influence the labeling pattern that the line will produce. 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 (thalamocortical projection neurons), whereas the *Pvalb-IRES-Cre* line does not[68]. 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, compared to both *Gpr139-IRES2-FlpO and Gpr139-IRES2-FlpO-neo* (**Figures S7A$_{5\text{-}6}$**, **S7B$_{4\text{-}6}$**).

As observed before[9,79,80], presence of the Neomycin selection cassette (Neo) can affect transgene expression. Some of the lines we include in this paper were generated by gene targeting without drug selection (the Ngai lines; **Figure S7**), whereas others were generated using antibiotic selection and contained Neo. It is a common practice to remove Neo after the line is established[68]. 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[79]. In most cases, the removal of Neo results in an expansion of expression as in the case of *Slc17a8-IRES2-Cre* (**Figure S8B$_{6-7}$**). It does not appear to make a difference for the *Gpr139-IRES2-FlpO* line (**Figure S8B$_{31-32}$**) or for the *Chrna6-IRES2-FlpO line* (**Figure S8B$_{1-2}$**). Note that in the latter case, the addition of WPRE in the *Chrna6-IRES2-FlpO-WPRE-neo* did not appear to change cell type labeling: L5_ET_Chrna6 was still the dominant labeled type, with the same rare L6b cluster (L6b_Col8a1_Rprm) expressing the *Chrna6* mRNA also labeled (**Figure S8B$_{1-3}$**)[8,9].

In this study, we used SSv4 to evaluate the expression pattern of transgenic lines. However, it is important to compare, whenever possible, the SSv4 data with other modalities to confirm labeling patterns. For example, the SSv4 data for *Npnt-2A-FlpO;Ai65* show cells from both L2-3 and L5 whereas that for *Npnt-2A-FlpO;Ai193* only includes L2-3 cells implying *Ai193* only labels a subset of cells expressing Npnt. However, consistent with *in situ* hybridization results (https://mouse.brain-map.org/experiment/show/71670677; **Figure S8C**)[81], *Npnt-2A-FlpO;Ai65F* and *Npnt-2A-FlpO;Ai193* show identical expression patterns by STPT with labeling of both L2-3 and L5 cells (**Figure S8C$_{7-8}$**). This discrepancy can be explained by our previous report that some cell types (specifically cortical Pvalb and L5 ET cells) are sensitive to FACS isolation, and survive the process with variable success experiment-to-experiment[8,9]. Therefore, the absence of L5_ET cells in some SSv4 experiments (**Figure S8A$_{9-10}$**) is likely an effect of FACS, and not differential reporter sensitivity.

We sought to further explore the effect of reporter line choice by comparing labeling patterns in crosses where the same driver lines were crossed with previously characterized reporters (*Ai14* and *Ai65/Ai65F*) and the new AND/OR reporters (*Ai193* and *Ai224*). In every case examined, we found that there was no discernable difference between the patterns of labelled cells. For example, triple-transgenic animals obtained by crossing *Chodl-P2A-Cre;Sst-IRES-FlpO* with *Ai65* or *Ai224* show a similar number and distribution of labelled cells (red cells in the *Ai65* cross and yellow-double labelled cells in the *Ai224* cross; **Figure S8D**). Likewise, the labelling patterns in double transgenic crosses of *Sst-IRES-FlpO* or *Chodl-P2A-Cre* with *Ai224, Ai14*, *RCL-H2B-EGFP*[69] or *Ai193,* show that the labeling pattern, at least for these reporters, depends on the recombinase lines employed (**Figure S8D**).

Finally, we expected that the knock-in recombinase-based labeling may be more inclusive of all cells that express a certain marker gene compared to enhancer AAVs. However, we show at least one instance where the enhancer virus faithfully captures the expression pattern of a marker gene while the knock-in recombinase line-based labeling is incomplete. *Cplx3* is a marker gene for L6b neurons as shown by scRNA-seq[9,13] and RNA *in situ* hybridization[81] (https://mouse.brain-map.org/experiment/show/70928340; **Figure S8E**). The expression pattern for AiE2359m, as well other L6b enhancers we discovered, mirrors *Cplx3* expression whereas the transgenic cross, *Cplx3-P2A-FlpO;Ai193* excludes *Cplx3*+ cells in the lateral entorhinal area, and L1 interneurons, which also express this gene[82]. This effect is independent of the reporter line as we also observe it with a different reporter in *Cplx3-P2A-FlpO;Ai65* (**Figure S8E**).

In summary, we provide a wealth of characterization data for the tools we created and underline the importance of careful evaluation of each tool or a combination of tools for the intended purpose. We strongly suggest evaluation of each tool/tool combination with more than one data modality, as some modalities have modality-specific pitfalls that can be highlighted with a different modality.

Discussion

The definition of cell types and their taxonomy in the central nervous system and genetic access to individual types are essential for our understanding of how they contribute to nervous system function and dysfunction[1,3,11,83]. We utilized single-cell transcriptomic and epigenomic data to identify marker genes for the generation of new transgenic lines, as well as putative enhancers for the creation of enhancer AAVs.

Currently, the use of transgenic mouse lines is the main approach for gaining access to molecularly identified cellular populations[8,67,84,85]. Transgenic mouse lines are created by inserting exogenous DNA (e.g., fluorophore, recombinase or transcription factor) into the mouse genome. A subset of transgenic mouse lines, the knock-in lines, are created by inserting a single copy of exogenous DNA, at a specific position, frequently within a marker gene in the mouse genome[67,86]. All the transgenic driver lines described in this paper were generated by knocking-in a recombinase into the endogenous gene locus. This contrasts with randomly integrated transgenes where the endogenous elements have been taken out of context and depending on the size/regulatory elements included, copy number and insertion site, dramatic variations in labeling patterns may occur[74,87] or may even behave as enhancer traps[88].

The knock-in approach takes advantage of the endogenous genomic regulatory elements to enable selective expression of various transgenes for cell labeling, monitoring and/or perturbation, and frequently produces expected cell-type labeling. However the efficiency, specificity, and strength of cell type labeling by transgenes generated in the same locus can vary depending on the exact components inserted[68,89], and we report several examples of this phenomenon in this study. In a number of cases, the driver lines labeled unintended cell types in which the marker gene expression was weak or not observed in adult mice, which is important to note especially if functional reporters are to be driven by them[90]. In order to restrict recombination to the target population, we generated triple transgenic mice using two separate driver lines, whose recombination patterns intersect only in the population of choice, with a dual reporter line requiring the presence of both recombinases for its expression[24,69,70,91].

Generation, validation, and maintenance of transgenic lines is expensive and laborious. Transgenes are integrated into the genome and although methods for their modification once established have become recently available[92], these are not widely utilized. Moreover, site-specific transgenesis in other mammalian species is difficult and costly[93,94] Therefore, it is not practical to rely on transgenic lines for all experiments. The use of enhancer AAVs[17,18,22,30,95] and other viral[32,96] or non-viral technologies[97] that do not require germline modification can potentially circumvent all these obstacles. However, there are caveats to be considered when applying enhancer AAVs: 1) even for the same vector, the degree of specificity and expression strength heavily depend, among other factors, on the concentration, delivery method, and viral serotype, which can lead to larger variability within and across experiments than with mouse transgenes; 2) access to some populations can be more difficult, due to their decreased susceptibility to viral transduction (e.g., microglia)[64]. For any untested crosses or combinations of transgene and virus, we advise the user to characterize them before assuming specificity based on the marker genes or enhancers used.

In this study, we provide a detailed and comprehensive report on the development, screening and use of enhancer AAVs in the context of the mouse cortex. We demonstrate that, at the subclass level of cell-type resolution, the mammalian species of origin appears to have little effect on enhancer performance, suggesting that many of the enhancer sequences described here could be useful in other mammalian species, even if their orthologs are not present in the species of interest. To evaluate enhancer specificity at a transcriptomic cell type (cluster) 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, Figure S4A**). We also found that the SYFP mRNA expression was significantly lower when driven by enhancer AAVs compared to the pan-neuronal hSyn1 promoter/enhancer, which could make enhancers insufficiently strong to drive functional cargo such as effectors and indicators. However, we show that in many cases, this limitation can 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 S5, S6**). We demonstrate that all these approaches can lead to a marked increase in cargo expression, often with tolerable specificity loss.

The tools reported here, as well as the scaled and standardized process used to create and evaluate them, provide an unprecedented resource that should enable diverse experimental strategies for understanding mammalian cortex function, including access to many previously inaccessible cortical types. We are in the process of making all materials, data and metadata associated with the study publicly available. Moreover, in an associated study[98], our standardized experimental process for enhancer evaluation described here has advanced our understanding of the basic features underlining enhancer performance.

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) protocols 1508, 1802, 1806, 2105, and 2406, 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. 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. See **Table S2** for a full list of all transgenic mouse lines included in this study.

10x Genomics snMultiome data generation

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. Mice were anesthetized with 2.5-3.5% isoflurane. Following perfusion, brains were isolated quickly, and frozen for 2 minutes in liquid nitrogen vapor. Frozen brain samples were sectioned on a cryostat to obtain 300 µm sections. Micro knives were used to microdissect the area of interest according to the Allen reference atlas. Images were collected pre- and post-microdissection to document which brain regions were profiled. Dissectates were placed in 12-well plates in the cryostat after collection.

Following dissection, nuclei were isolated with the RAISINs protocol (RNA-seq for Profiling Intact Nuclei with Ribosome-bound mRNA) # PF0334[99] developed based on a previously published protocol[100]. Briefly, tissue samples were placed in CST buffer 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[28]. 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[40]. Pseudobulk coverage data were used for identification of peaks in VISp subclasses using Macs2 as implemented by ArchR (version 1.0.2) [40,41]. 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) [101]. 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. See **Table S3** for a full list of all plasmids included in this study.

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. [102–104].

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 $2.5\text{x}10^{-2}$ to $2.5\text{x}10^{-8}$ were used for the measurement. Serial dilutions of $2.5\text{x}10^{-5}$ to $2.5\text{x}10^{-8}$ were used for fitting the linear dynamic range. Viral titer was calculated by averaging virus concentration of two dilutions within the linear dynamic 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[44]: Male and female C57BL/6J mice, aged P27–P33, were anesthetized using isoflurane. Each mouse received an injection of $5\text{x}10^{11}$ 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[105]: 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, male and female C57BL/6J mice, aged P45–P90, were anesthetized with isoflurane before injecting with filtered PHP.eB-pseudotyped AAV 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 $1.5\text{x}10^{9}$ 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 sagittally along the midline, 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 v4 sample preparation and analysis

Sample preparation for SMART-Seq was performed using the SMART-Seq v4 kit (Takara Cat#634894) as described previously[9]. 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[106] 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[72] (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, individual cells were FACS-isolated 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, two or more collections were made using an "upper" (layers 1-4) and a "lower" (layers 5-6) dissection strategy, and the data were pooled. In summary, for transgenic lines, cells were summed across all layer collections, which may include single layers or combinations. This may contribute to biases in cell-type compositions reported.

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[67,68,70]. 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. The 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 we employed a recombinase-mediated cassette exchange (RMCE) strategy similar to that previously described[70], but instead of using Flp recombinase for targeting, we used Bxb1 integrase[107] 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[108], 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. Genomic DNA 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 analyses

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.

Data Availability

Mouse 10x Multiome data and single-cell RNA-seq data are available at the Neuroscience Multi-omic Archive (NeMO):
https://data.nemoarchive.org/biccn/grant/u19_zeng/zeng/multimodal/sncell/10xMultiome_RNAseq/mouse/raw
https://data.nemoarchive.org/biccn/grant/u19_zeng/zeng/multimodal/sncell/10xMultiome_ATACseq/mouse/raw

SSv4 scRNA-seq data are available at NeMO:
https://data.nemoarchive.org/biccn/grant/rf1_tasic/tasic/transcriptome/scell/SSv4/mouse/raw/

Primary screen epifluorescence and serial-two-photon tomography (STPT) data are available at the Brain Image Library (BIL): https://doi.org/10.35077/g.1162

Primary screen epifluorescence data are available at BIL:
https://download.brainimagelibrary.org/4e/fa/4efaa61008dfb900/

Serial-two-photon tomography (STPT) data are available at BIL:
https://download.brainimagelibrary.org/bc/59/bc59278fe0669df7/

Code Availability

Analysis methods used in this manuscript include SCVI (https://github.com/scverse/scvi-tools) and Scanpy (https://github.com/scverse/scanpy) for data processing integration and visualization, ArchR for ATAC-seq analyses (https://github.com/GreenleafLab/ArchR), as well as scrattch.hicat and scrattch.bigcat for cell type mapping and identification (https://github.com/AllenInstitute/scrattch.hicat). Plots were generated using R packages ggplot2, dendextend, which are available from both CRAN and github. Additional code used for data processing and analysis in this manuscript will be made available upon request.

Author contributions

Conceptualization, T.L.D., H.Z., B.P.L., J.N., J.T.T., B.Tasic; ***Methodology - Transgenic lines***, S.N., T.L.D., D.A.S., S.A., T.B., J.Bendrick, J. Bowlus, G.B., B.C., R.K.C., M.D., C.Halterman, O.H., C.Huang, R.L., A.Y.L., M.L., G.H.L., E.L., D.Rette, J.P.S., L.Shulga, L.Siverts, E.S., M.W., T.W., B.W., M.R., K.R., H.Z., B.Tasic; ***Methodology - Viral vectors***, M.H., T.L.D., J.K.M., R.A.M., X.O., J.R.R., N.Donadio, L.G., N.G., F.H., S.H., S.K., E.K., R.K., G.H.L., D.N., K.S., C.T., N.W., T.Z., S.Y., B.P.L., J.T.T., B.Tasic; ***Investigation***, Y.B., M.H., S.N., T.L.D., D.D., J.K.M., M.J.T., J.R.R., A.A., S.B., J.Bendrick, D.B., K.B., T.C., A.B.C., M.C., N.Dotson, T.E., A.G., J.Gloe, L.G., F.H., W.H., S.H., Z.J., M.K., G.H.L., N.L., J.M., R.M., E.M., K.N., V.O., A.Oyama, T.P., E.P., C.A.P., L.P., S.R., C.R., A.R., N.S., A.R.S., M.T., C.T., A.T., N.W., J.A., N.Dee, K.A.S., H.Z., B.P.L., J.T.T., B.Tasic; ***Resources***, S.M.; ***Formal Analysis***, Y.B., M.H., S.N., T.L.D., D.D., A.Oster, R.C., S.C., M.Gabitto, J.G., J.Goldy, B.B.G., L.G., A.H., N.J., B.K., C.L., N.L., E.M., D.Rocha, A.S., S.Somasundaram, C.T.J.V., M.E.W., Z.Y., B.P.L., J.T.T., B.Tasic; ***Data curation***, Y.B., M.H., S.N., D.D., S.W.W., A.Oster, J.R.R., N.G., R.E.A.S., B.P.L., J.T.T.; ***Project Administration***, A.Oster, T.M., R.E.A.S., S.M.S., B.Thyagarajan; ***Visualization***, Y.B., M.H., S.N., D.D., B.Tasic; ***Supervision***, Y.B., S.N., T.L.D., M.J.T., T.E.B., T.C., A.B.C., M.Gasperini, D.H., R.L., T.M., L.N., L.P., C.T.J.V., Z.Y., N.Dee, M.R., K.R., S.M., S.M.S., K.A.S., L.E., J.W., S.Y., E.S.L., H.Z., B.P.L., J.N., J.T.T., B.Tasic; ***Writing – Original Draft***, Y.B., M.H., S.N., D.D., B.Tasic; ***Writing – Review & Editing***, Y.B., M.H., S.N., D.D., S.W.W., B.P.L., J.T.T., B.Tasic; ***Funding Acquisition***, E.S.L., H.Z., B.P.L., J.N., J.T.T., B.Tasic.

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. We thank Bing Ren and colleagues for open access sharing of the CATlas resource for mouse brain snATAC-seq data. This work was funded by the Allen Institute for Brain Science in addition to the NIH grants U19MH114830 to H.Z., RF1MH121274 to B.T., and RF1MH114126 and UG3MH120095 to E.S.L., J.T., and B.P.L. 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. Allen Institute for Brain Science has filed patent applications for many enhancer AAVs described in this manuscript with multiple authors listed as inventors. B.P.L., B.B.G., J.K.M, and E.S.L. are founders of EpiCure Therapeutics, Inc. The other authors declare no competing interests.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Microsoft Copilot in order to improve readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. 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.
2. DeNardo, L., and Luo, L. (2017). Genetic strategies to access activated neurons. Curr. Opin. Neurobiol. *45*, 121–129.
3. 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.
4. 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.
5. 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.
6. Qiu, L., Zhang, B., and Gao, Z. (2022). Lighting Up Neural Circuits by Viral Tracing. Neurosci. Bull. *38*, 1383–1396. https://doi.org/10.1007/s12264-022-00860-7.
7. Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138–1142.
8. 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.
9. 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.
10. 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.
11. Tasic, B. (2018). Single cell transcriptomics in neuroscience: cell classification and beyond. Current Opinion in Neurobiology *50*, 242–249. https://doi.org/10.1016/j.conb.2018.04.021.
12. 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.
13. 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.
14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. Furlanis, E., Dai, M., Garcia, B.L., Vergara, J., Pereira, A., Pelkey, K., Tran, T., Gorissen, B.L., Vlacho, A., Hairston, A., et al. (2024). An enhancer-AAV toolbox to target and manipulate distinct interneuron subtypes. Preprint at bioRxiv, https://doi.org/10.1101/2024.07.17.603924 https://doi.org/10.1101/2024.07.17.603924.

22. 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.

23. Gong, L., Liu, X., Wu, J., and He, M. (2022). Emerging strategies for the genetic dissection of gene functions, cell types, and neural circuits in the mammalian brain. Mol Psychiatry *27*, 422–435. https://doi.org/10.1038/s41380-021-01292-x.

24. Zhang, Q., Liu, X., Gong, L., and He, M. (2023). Combinatorial genetic strategies for dissecting cell lineages, cell types, and gene function in the mouse brain. Dev Growth Differ *65*, 546–553. https://doi.org/10.1111/dgd.12902.

25. Bakken, T.E., Jorstad, N.L., Hu, Q., Lake, B.B., Tian, W., Kalmbach, B.E., Crow, M., Hodge, R.D., Krienen, F.M., Sorensen, S.A., et al. (2021). Comparative cellular analysis of motor cortex in human, marmoset and mouse. Nature *598*, 111–119. https://doi.org/10.1038/s41586-021-03465-8.

26. Li, Y.E., Preissl, S., Hou, X., Zhang, Z., Zhang, K., Qiu, Y., Poirion, O.B., Li, B., Chiou, J., Liu, H., et al. (2021). An atlas of gene regulatory elements in adult mouse cerebrum. Nature *598*, 129–136. https://doi.org/10.1038/s41586-021-03604-1.

27. 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*.

28. 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.

29. 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.

30. 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.

31. Lee, B.R., Dalley, R., Miller, J.A., Chartrand, T., Close, J., Mann, R., Mukora, A., Ng, L., Alfiler, L., Baker, K., et al. (2023). Signature morphoelectric properties of diverse GABAergic interneurons in the human neocortex. Science *382*, eadf6484. https://doi.org/10.1126/science.adf6484.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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 Res *34*, D590-598. https://doi.org/10.1093/nar/gkj144.

43. 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.

44. 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.

45. Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., et al. (2022). Nonlinear control of transcription through enhancer–promoter interactions. Nature *604*, 571–577. https://doi.org/10.1038/s41586-022-04570-y.

46. Kwasnieski, J.C., Fiore, C., Chaudhari, H.G., and Cohen, B.A. (2014). High-throughput functional testing of ENCODE segmentation predictions. Genome Res *24*, 1595–1602. https://doi.org/10.1101/gr.173518.114.

47. Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A genome-wide framework for mapping gene regulation via cellular genetic screens. Cell *176*, 377-390.e19. https://doi.org/10.1016/j.cell.2018.11.029.

48. Wang, Q., Ding, S.-L., Li, Y., Royall, J., Feng, D., Lesnar, P., Graddis, N., Naeemi, M., Facer, B., Ho, A., et al. (2020). The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. Cell *181*, 936-953.e20. https://doi.org/10.1016/j.cell.2020.04.007.

49. Gouwens, N.W., Sorensen, S.A., Baftizadeh, F., Budzillo, A., Lee, B.R., Jarsky, T., Alfiler, L., Baker, K., Barkan, E., Berry, K., et al. (2020). Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. Cell *183*, 935-953.e19.

50. Sorensen, S.A., Gouwens, N.W., Wang, Y., Mallory, M., Budzillo, A., Dalley, R., Lee, B., Gliko, O., Kuo, H.-C., Kuang, X., et al. (2023). Connecting single-cell transcriptomes to projectomes in mouse visual cortex. bioRxiv. https://doi.org/10.1101/2023.11.25.568393.

51. Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308. https://doi.org/10.1016/0092-8674(81)90413-X.

52. 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.

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

54. 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.

55. 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.

56. 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.

57. 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.

58. 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.

59. Bothma, J.P., Garcia, H.G., Ng, S., Perry, M.W., Gregor, T., and Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. eLife *4*, e07956. https://doi.org/10.7554/eLife.07956.

60. Hay, D., Hughes, J.R., Babbs, C., Davies, J.O.J., Graham, B.J., Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Sharpe, J.A., Suciu, M.C., et al. (2016). Genetic dissection of the α-globin super-enhancer in vivo. Nat Genet *48*, 895–903. https://doi.org/10.1038/ng.3605.

61. 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.

62. 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.

63. 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.

64. Shi, H., He, Y., Zhou, Y., Huang, J., Maher, K., Wang, B., Tang, Z., Luo, S., Tan, P., Wu, M., et al. (2023). Spatial atlas of the mouse central nervous system at molecular resolution. Nature *622*, 552–561. https://doi.org/10.1038/s41586-023-06569-5.

65. McGovern, D.J., Ly, A., Ecton, K.L., Huynh, D.T., Prévost, E.D., Gonzalez, S.C., McNulty, C.J., Rau, A.R., Hentges, S.T., Daigle, T.L., et al. (2022). Ventral tegmental area glutamate neurons mediate nonassociative consequences of stress. Mol Psychiatry. https://doi.org/10.1038/s41380-022-01858-3.

66. Liwang, J.K., Kronman, F.A., Minteer, J.A., Wu, Y.-T., Vanselow, D.J., Ben-Simon, Y., Taormina, M., Parmaksiz, D., Way, S.W., Zeng, H., et al. (2023). epDevAtlas: Mapping GABAergic cells and microglia in postnatal mouse brains. bioRxiv, 2023.11.24.568585. https://doi.org/10.1101/2023.11.24.568585.

67. 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.

68. 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.

69. 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.

70. 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.

71. Steinmetz, N.A., Buetfering, C., Lecoq, J., Lee, C.R., Peters, A.J., Jacobs, E.A.K., Coen, P., Ollerenshaw, D.R., Valley, M.T., de Vries, S.E.J., et al. (2017). Aberrant Cortical Activity in Multiple GCaMP6-Expressing Transgenic Mouse Lines. eNeuro *4*, ENEURO.0207-17.2017. https://doi.org/10.1523/ENEURO.0207-17.2017.

72. 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.

73. Grimes, W.N., Seal, R.P., Oesch, N., Edwards, R.H., and Diamond, J.S. (2011). Genetic targeting and physiological features of VGLUT3+ amacrine cells. Vis. Neurosci. *28*, 381–392.

74. 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.

75. Vong, L., Ye, C., Yang, Z., Choi, B., Chua, S., and Lowell, B.B. (2011). Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. Neuron *71*, 142–154. https://doi.org/10.1016/j.neuron.2011.05.028.

76. Dudok, B., Klein, P.M., Hwaun, E., Lee, B.R., Yao, Z., Fong, O., Bowler, J.C., Terada, S., Sparks, F.T., Szabo, G.G., et al. (2021). Alternating sources of perisomatic inhibition during behavior. Neuron *109*, 997-1012.e9. https://doi.org/10.1016/j.neuron.2021.01.003.

77. Morairty, S.R., Dittrich, L., Pasumarthi, R.K., Valladao, D., Heiss, J.E., Gerashchenko, D., and Kilduff, T.S. (2013). A role for cortical nNOS/NK1 neurons in coupling homeostatic sleep drive to EEG slow wave activity. Proceedings of the National Academy of Sciences *110*, 20272–20277. https://doi.org/10.1073/pnas.1314762110.

78. Taniguchi, H., He, M., Wu, P., Kim, S., Paik, R., Sugino, K., Kvitsiani, D., Fu, Y., Lu, J., Lin, Y., et al. (2011). A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. Neuron *71*, 995–1013. https://doi.org/10.1016/j.neuron.2011.07.026.

79. Zhuang, J., Wang, Y., Ouellette, N.D., Turschak, E.E., Larsen, R.S., Takasaki, K.T., Daigle, T.L., Tasic, B., Waters, J., Zeng, H., et al. (2021). Laminar distribution and arbor density of two functional classes of thalamic inputs to primary visual cortex. Cell Rep *37*, 109826. https://doi.org/10.1016/j.celrep.2021.109826.

80. Tasic, B., Hippenmeyer, S., Wang, C., Gamboa, M., Zong, H., Chen-Tsai, Y., and Luo, L. (2011). Site-specific integrase-mediated transgenesis in mice via pronuclear injection. Proceedings of the National Academy of Sciences *108*, 7902–7907. https://doi.org/10.1073/pnas.1019507108.

81. Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature *445*, 168–176. https://doi.org/10.1038/nature05453.

82. Ben-Simon, Y., Kaefer, K., Velicky, P., Csicsvari, J., Danzl, J.G., and Jonas, P. (2022). A direct excitatory projection from entorhinal layer 6b neurons to the hippocampus contributes to spatial coding and memory. Nat Commun *13*, 4826. https://doi.org/10.1038/s41467-022-32559-8.

83. Di Bella, D.J., 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 *112*, 2091–2111. https://doi.org/10.1016/j.neuron.2024.04.021.

84. Harris, J.A., Hirokawa, K.E., Sorensen, S.A., Gu, H., Mills, M., Ng, L.L., Bohn, P., Mortrud, M., Ouellette, B., Kidney, J., et al. (2014). Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. Front Neural Circuits *8*, 76. https://doi.org/10.3389/fncir.2014.00076.

85. Arias, A., Manubens-Gil, L., and Dierssen, M. (2022). Fluorescent transgenic mouse models for whole-brain imaging in health and disease. Front. Mol. Neurosci. *15*. https://doi.org/10.3389/fnmol.2022.958222.

86. 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.

87. Laboulaye, M.A., Duan, X., Qiao, M., Whitney, I.E., and Sanes, J.R. (2018). Mapping Transgene Insertion Sites Reveals Complex Interactions Between Mouse Transgenes and Neighboring Endogenous Genes. Front. Mol. Neurosci. *11*. https://doi.org/10.3389/fnmol.2018.00385.

88. 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.

89. 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.

90. Müller-Komorowska, D., Opitz, T., Elzoheiry, S., Schweizer, M., Ambrad Giovannetti, E., and Beck, H. (2020). Nonspecific Expression in Limited Excitatory Cell Populations in Interneuron-Targeting Cre-driver Lines Can Have Large Functional Effects. Front. Neural Circuits *14*. https://doi.org/10.3389/fncir.2020.00016.

91. Palicz, R., Pater, B., Truschow, P., Witte, M., and Staiger, J.F. (2024). Intersectional strategy to study cortical inhibitory parvalbumin-expressing interneurons. Sci Rep *14*, 2829. https://doi.org/10.1038/s41598-024-52901-y.

92. Miura, H., Nakamura, A., Kurosaki, A., Kotani, A., Motojima, M., Tanaka, K., Kakuta, S., Ogiwara, S., Ohmi, Y., Komaba, H., et al. (2024). Targeted insertion of conditional expression cassettes into the mouse genome using the modified i-PITT. BMC Genomics *25*, 568. https://doi.org/10.1186/s12864-024-10250-0.

93. Chenouard, V., Remy, S., Tesson, L., Ménoret, S., Ouisse, L.-H., Cherifi, Y., and Anegon, I. (2021). Advances in Genome Editing and Application to the Generation of Genetically Modified Rat Models. Front. Genet. *12*. https://doi.org/10.3389/fgene.2021.615491.

94. Feng, G., Jensen, F.E., Greely, H.T., Okano, H., Treue, S., Roberts, A.C., Fox, J.G., Caddick, S., Poo, M., Newsome, W.T., et al. (2020). Opportunities and limitations of genetically modified nonhuman primate models for neuroscience research. Proc Natl Acad Sci U S A *117*, 24022–24031. https://doi.org/10.1073/pnas.2006515117.

95. Tremblay, S., Acker, L., Afraz, A., Albaugh, D.L., Amita, H., Andrei, A.R., Angelucci, A., Aschner, A., Balan, P.F., Basso, M.A., et al. (2020). An Open Resource for Non-human Primate Optogenetics. Neuron *108*, 1075-1090.e6. https://doi.org/10.1016/j.neuron.2020.09.027.

96. Qian, Y., Li, J., Zhao, S., Matthews, E.A., Adoff, M., Zhong, W., An, X., Yeo, M., Park, C., Yang, X., et al. (2022). Programmable RNA sensing for cell monitoring and manipulation. Nature *610*, 713–721. https://doi.org/10.1038/s41586-022-05280-1.

97. Brown, D.W., Wee, P., Bhandari, P., Bukhari, A., Grin, L., Vega, H., Hejazi, M., Sosnowski, D., Ablack, J., Clancy, E.K., et al. (2024). Safe and effective in vivo delivery of DNA and RNA using proteolipid vehicles. Cell *187*, 5357-5375.e24. https://doi.org/10.1016/j.cell.2024.07.023.

98. Johansen, N.J., Kempynck, N., Zemke, N.R., Somasundaram, S., Winter, S.D., Hooper, M., Dwivedi, D., Lohia, R., Wehbe, F., Li, B., et al. (2024). Evaluating Methods for the Prediction of Cell Type-Specific Enhancers in the Mammalian Cortex. Preprint at bioRxiv, https://doi.org/10.1101/2024.08.21.609075 https://doi.org/10.1101/2024.08.21.609075.

99. Institute, A. (2023). RAISINs (RNA-seq for Profiling Intact Nuclei with Ribosome-bound mRNA) nuclei isolation from mouse CNS tissue protocol.

100. Drokhlyansky, E., Smillie, C.S., Van Wittenberghe, N., Ericsson, M., Griffin, G.K., Eraslan, G., Dionne, D., Cuoco, M.S., Goder-Reiser, M.N., Sharova, T., et al. (2020). The Human and Mouse Enteric Nervous System at Single-Cell Resolution. Cell *182*, 1606-1622.e23. https://doi.org/10.1016/j.cell.2020.08.003.

101. 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.

102. Protocols.io VC0002 Allen Institute for Brain Science (2024). Production of Crude AAV Virus Extract.

103. Protocols.io VC0003 Allen Institute for Brain Science (2024). Production of AAV Virus by Iodixanol Gradient Ultracentrifugation.

104. Protocols.io VC0005 Allen Institute for Brain Science (2024). Viral Titer Quantification by Digital Droplet PCR.

105. 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.

106. 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.

107. 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.

108. 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.

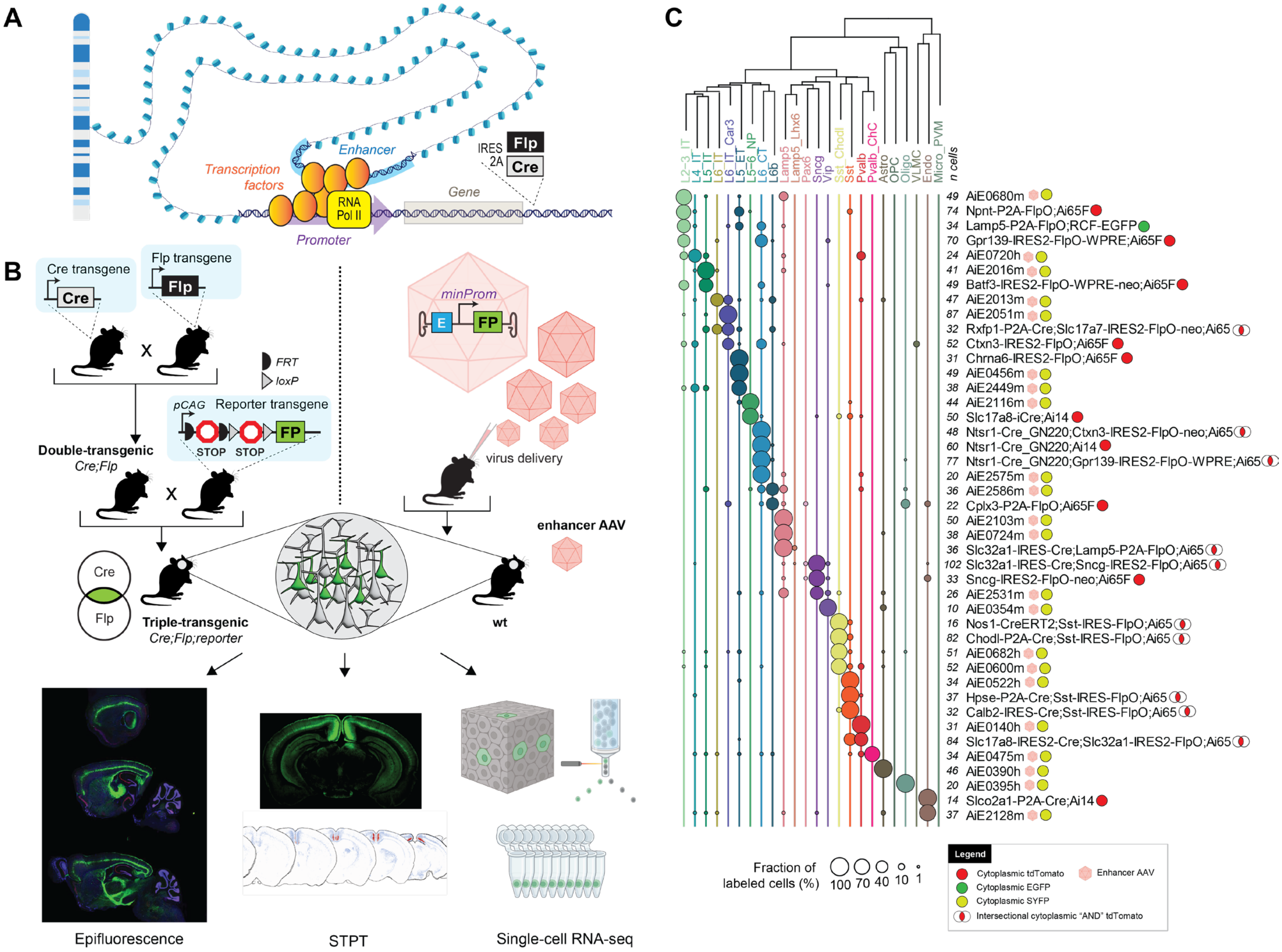


Figure 1: Cell-type-specific genetic tool design and characterization. A. A schematic illustration of gene regulation and location for recombinase insertions for knock-in transgenic mouse generation. **B.** Diagrams describing two strategies for genetic tool generation and characterization. *Left*: knock-in transgenic mouse lines are generated by insertions of recombinases into cell-type-specific differentially expressed genes. Generation of experimental animals requires one or more crosses to other recombinase lines and reporters. *Bottom*: The characterization of tool expression in brains of experimental animals is performed by three modalities: epifluorescence on select brain sections, serial-two-photon tomography (STPT) on whole brain, and single-cell RNA-seq on the visual cortex. *Right*: viral vectors utilize enhancers to achieve tool specificity. Generation of experimental animals requires retroorbital, intracerebroventricular or stereotaxic virus delivery to the animal. **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.

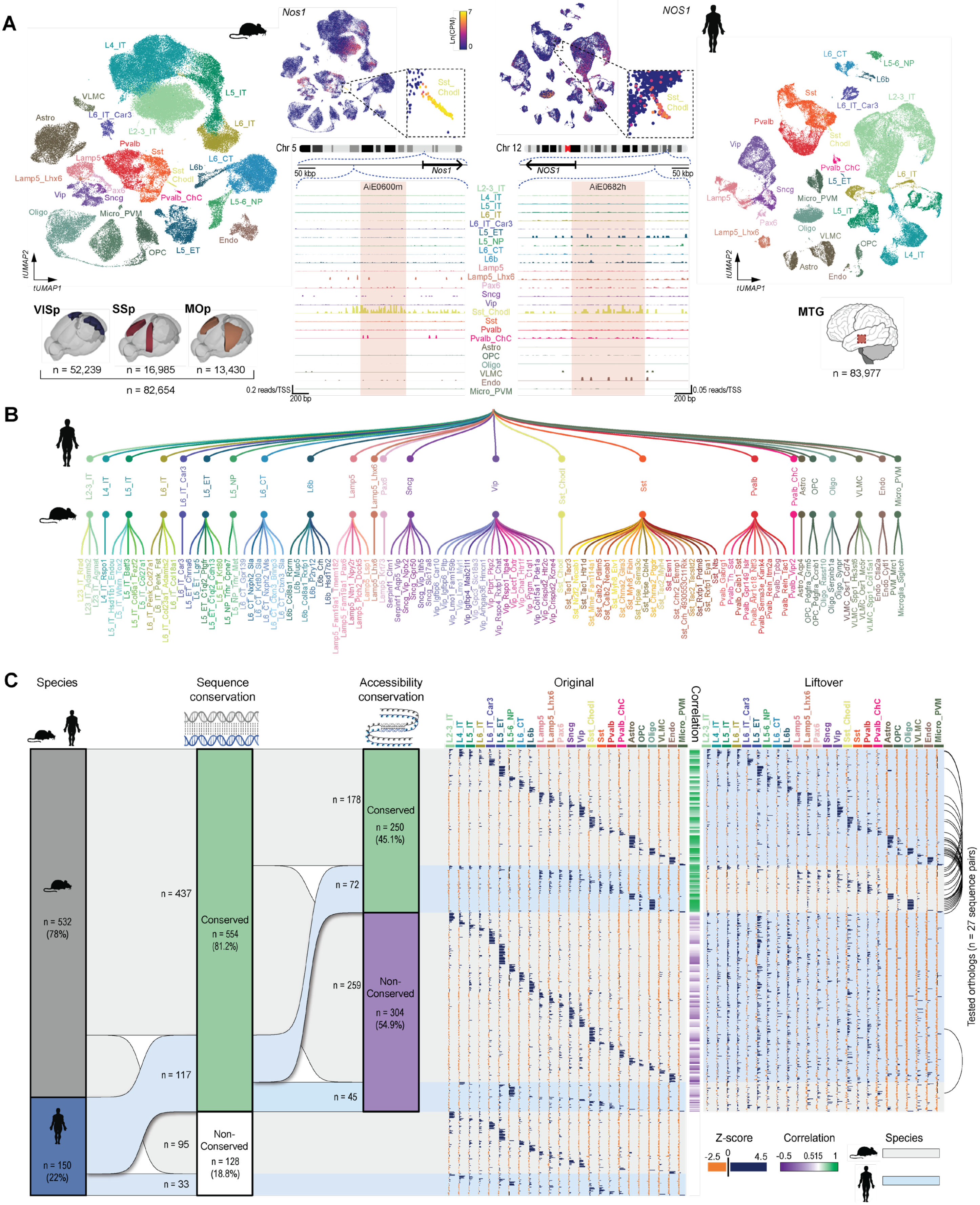

A
Nos1
NOS1
Ln(CPM)
7
0
L4_IT
L5_IT
L6_IT
VLMC
L6_IT_Car3
L2-3_IT
Astro
Lamp5
Pvalb
Sst
Sst_Chodl
L6b
L6_CT
Lamp5_Lhx6
Pax6
Sncg
Vip
Pvalb_ChC
L5-6_NP
Oligo
Micro_PVM
OPC
L5_ET
Endo
tUMAP1
tUMAP2
VISp
SSp
MOp
n = 52,239
n = 16,985
n = 13,430
n = 82,654
Chr 5
Chr 12
50 kbp
AiE0600m
AiE0682h
0.2 reads/TSS
0.05 reads/TSS
200 bp
L5_NP
MTG
n = 83,977
B
C
Species
Sequence conservation
Accessibility conservation
Original
Liftover
Correlation
n = 532 (78%)
n = 150 (22%)
n = 437
n = 117
n = 95
n = 33
Conserved
n = 554 (81.2%)
Non-Conserved
n = 128 (18.8%)
n = 178
n = 72
n = 259
n = 45
Conserved
n = 250 (45.1%)
Non-Conserved
n = 304 (54.9%)
Tested orthologs (n = 27 sequence pairs)
Z-score
-2.5
0
4.5
Correlation
-0.5
0.515
1
Species

Figure 2: Selection of putative enhancer sequences. A. Mouse and human single-nucleus transcriptomes obtained from single-nucleus multiomes represented in a transcriptomic UMAP and labeled according to cortical cell subclasses. Mouse nuclei were collected from the primary visual (VISp), somatosensory (SSp) and motor (MOp) cortices; human nuclei were collected from the middle 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 ‘native’ putative enhancer sequences tested (n = 682), divided by the genome of origin, followed by cross-species conservation of sequence and accessibility (left). The modified sequences produced by concatenation are not included. 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.

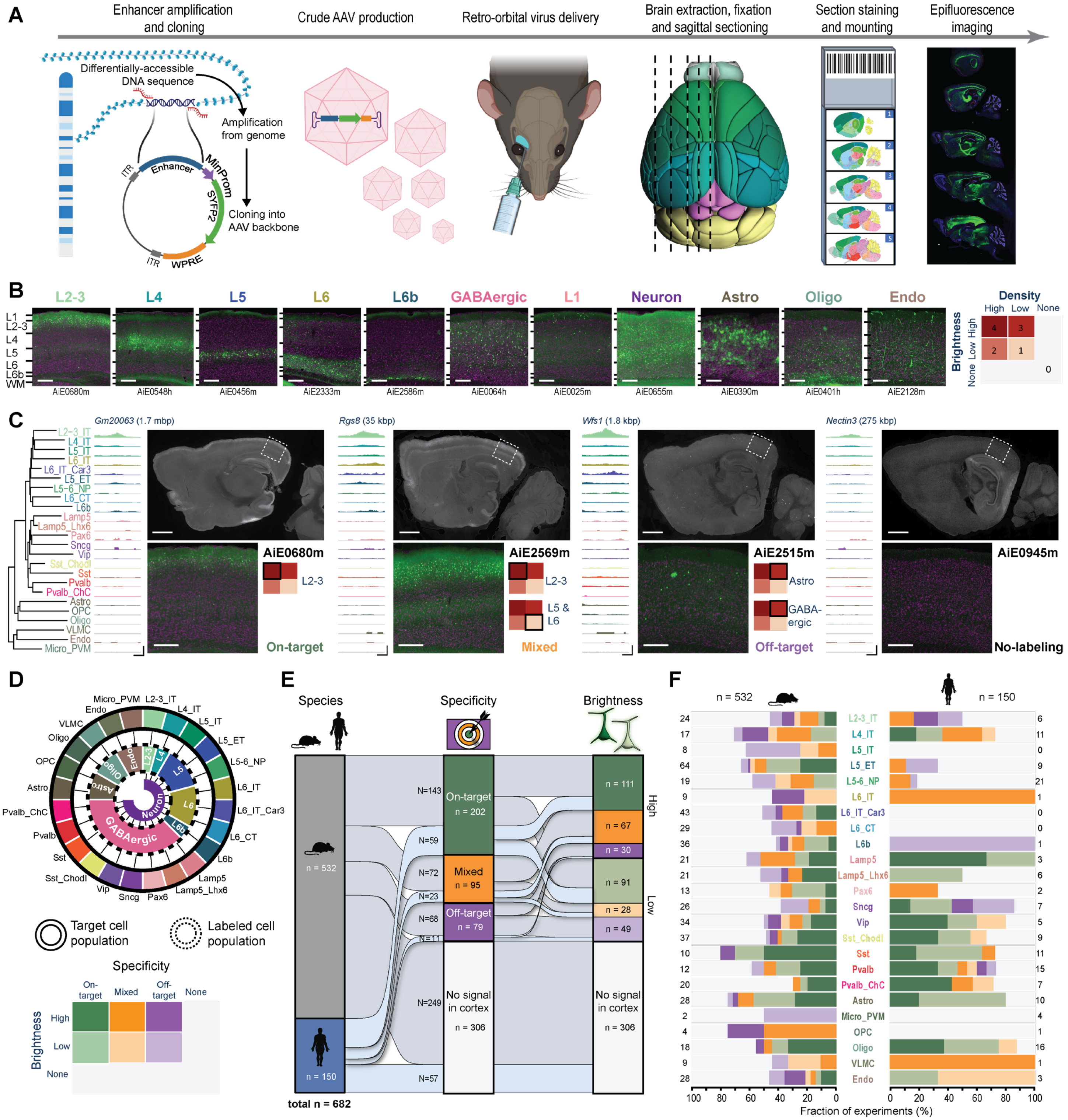

A
Enhancer amplification and cloning
Crude AAV production
Retro-orbital virus delivery
Brain extraction, fixation and sagittal sectioning
Section staining and mounting
Epifluorescence imaging
Differentially-accessible DNA sequence
Amplification from genome
Cloning into AAV backbone
ITR
Enhancer
MinProm
SYFP2
WPRE
ITR
B
L2-3
L4
L5
L6
L6b
GABAergic
L1
Neuron
Astro
Oligo
Endo
L1
L2-3
L4
L5
L6
L6b
WM
AiE0680m
AiE0548h
AiE0456m
AiE2333m
AiE2586m
AiE0064h
AiE0025m
AiE0655m
AiE0390m
AiE0401h
AiE2128m
Density
High
Low
None
Brightness
High
Low
None
4
3
2
1
0
C
Gm20063 (1.7 mbp)
Rgs8 (35 kbp)
Wfs1 (1.8 kbp)
Nectin3 (275 kbp)
L2-3_IT
L4_IT
L5_IT
L6_IT
L6_IT_Car3
L5_ET
L5-6_NP
L6_CT
L6b
Lamp5
Lamp5_Lhx6
Pax6
Sncg
Vip
Sst_Chodl
Sst
Pvalb
Pvalb_ChC
Astro
OPC
Oligo
VLMC
Endo
Micro_PVM
AiE0680m
L2-3
On-target
AiE2569m
L2-3
L5 &
L6
Mixed
AiE2515m
Astro
GABA-ergic
Off-target
AiE0945m
No-labeling
D
Micro_PVM
L2-3_IT
Endo
L4_IT
VLMC
L5_IT
Oligo
L5_ET
OPC
L5-6_NP
Astro
L6_IT
Pvalb_ChC
L6_IT_Car3
Pvalb
L6_CT
Sst
L6b
Sst_Chodl
Lamp5
Vip
Lamp5_Lhx6
Sncg
Pax6
Endo
Oligo
Astro
L2-3
L4
L5
L6
L6b
GABAergic
Neuron
Target cell population
Labeled cell population
Specificity
On-target
Mixed
Off-target
None
Brightness
High
Low
None
E
Species
Specificity
Brightness
n = 532
n = 150
total n = 682
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
High
Low
F
n = 532
n = 150
24
17
8
64
19
9
43
29
36
21
21
13
26
34
37
10
12
20
28
2
4
18
9
28
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
Micro_PVM
OPC
Oligo
VLMC
Endo
6
11
0
9
21
1
0
0
1
3
6
2
7
5
9
11
15
7
10
4
1
16
1
3
100
80
60
40
20
0
0
20
40
60
80
100
Fraction of experiments (%)

Figure 3: A pipeline for enhancer AAV screening in the mouse brain. A. Graphical summary of the enhancer screening pipeline. **B.** Representative images of the visual cortex, showing examples of 11 categories of visually distinguishable cortical populations used to evaluate the labeling pattern of each enhancer AAV. Individual enhancer IDs are shown beneath each image (left), and a scoring matrix for evaluating brightness of the fluorescent signal and the labeling density, compared to the density expected for each category (right). **C.** Representative tracks of chromatin accessibility for four individual enhancers targeting across the cortical subclasses, demonstrating differential accessibility in L2-3_IT, alongside representative epifluorescence image sets, showing the resulting labeling pattern, and the score given to each in VISp. The closest L2-3_IT marker gene is shown above each set of tracks, along with the distance from the enhancer to its TSS. Scale bars below the tracks represent 100 bp (horizontal) and 0.3 RPKM/cell (vertical). **D.** Schematic describing the approach used to determine target specificity, according to the alignment between the TCP and LCP (top left) and a matrix used for classifying all enhancers, based on a combination of their target specificity and signal brightness (bottom left). **E.** Summary plot showing performance of all tested enhancers (n = 682) according to the categories (right). > 50% of enhancers (n = 376) exhibited signal in the cortex, ~43% were putatively on-target or had mixed (on- and off-target) labeling, and ~30% were putatively on-target (202/682). **F.** Proportion of enhancers in each of the categories specified in (D), according to their genome of origin (top, n = 150 for human and 532 for mouse) and TCP. Numerical values represent the number of tested enhancers in each column. For images, scale bars for full section and expanded views are 1.0 and 0.2 mm, respectively.

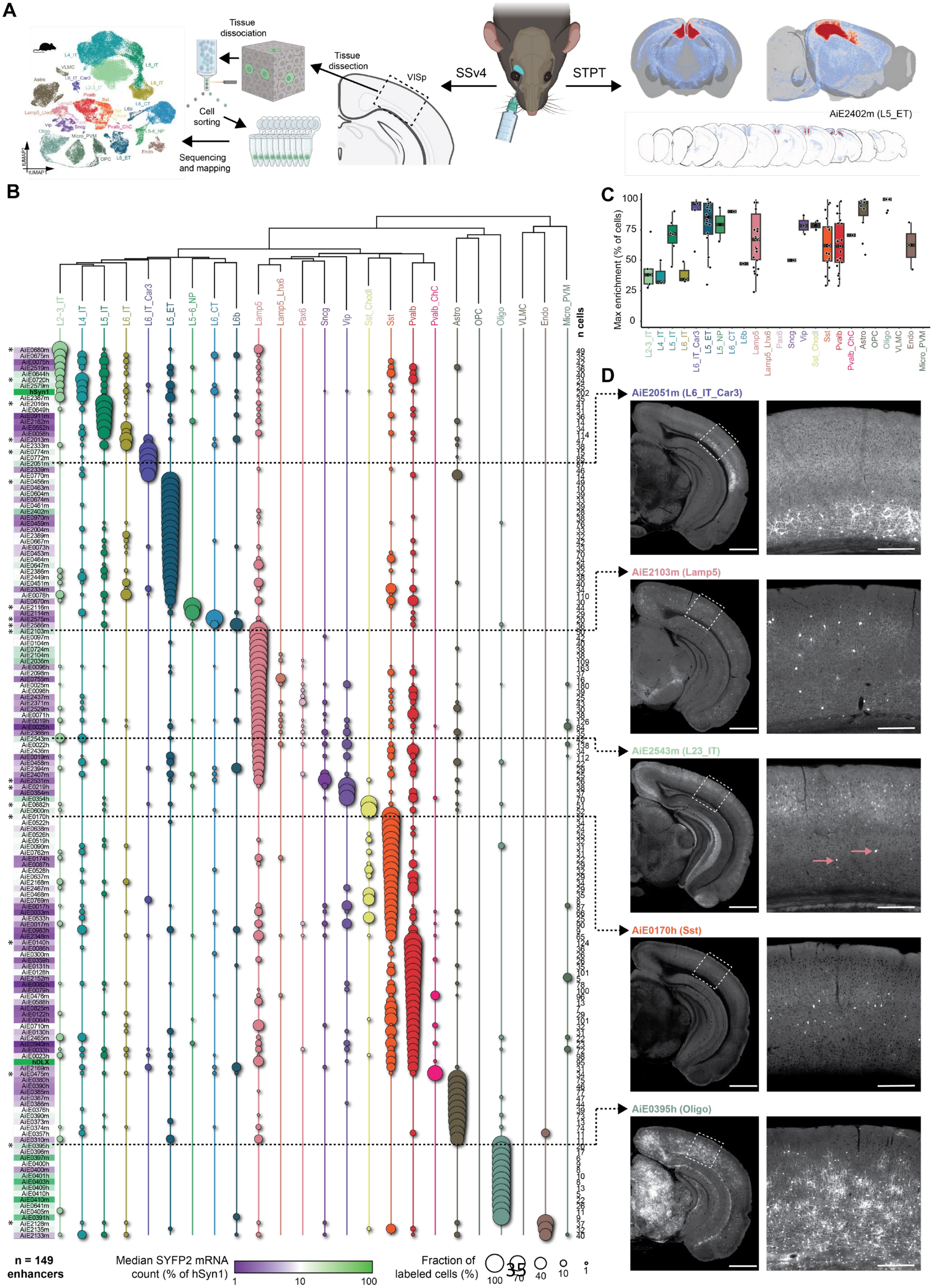

A
Tissue dissociation
Cell sorting
Sequencing and mapping
Tissue dissection
VISp
SSv4
STPT
AiE2402m (L5_ET)
tUMAP1
tUMAP2
L4_IT
L5_IT
VLMC
L6_IT_Car3
L6_IT
L2-3_IT
Astro
Pvalb
Sst
L6_CT
Lamp5_Lhx6
L6b
Vip
Sncg
Pvalb_ChC
L5-6_NP
Oligo
Micro_PVM
Endo
OPC
L5_ET
B
L2-3_IT
L4_IT
L5_IT
L6_IT
L6_IT_Car3
L5_ET
L5-6_NP
L6_CT
L6b
Lamp5
Lamp5_Lhx6
Pax6
Sncg
Vip
Sst_Chodl
Sst
Pvalb
Pvalb_ChC
Astro
OPC
Oligo
VLMC
Endo
Micro_PVM
n cells
hSyn1
hDLX
n = 149 enhancers
Median SYFP2 mRNA count (% of hSyn1)
1
10
100
Fraction of labeled cells (%)
100
70
40
10
1
C
Max enrichment (% of cells)
100
75
50
25
0
L2-3_IT
L4_IT
L5_IT
L6_IT
L6_IT_Car3
L5_ET
L5_NP
L6_CT
L6b
Lamp5
Lamp5_Lhx6
Pax6
Sncg
Vip
Sst_Chodl
Sst
Pvalb
Pvalb_ChC
Astro
OPC
Oligo
VLMC
Endo
Micro_PVM
D
AiE2051m (L6_IT_Car3)
AiE2103m (Lamp5)
AiE2543m (L23_IT)
AiE0170h (Sst)
AiE0395h (Oligo)

Figure 4: Secondary validation of target specificity, with scRNA-seq and whole-brain imaging. A. Schematic of methods for secondary validation. **B.** scRNA-seq analysis using SSv4, of FACS-sorted SYFP2$^+$ cells from the mouse visual cortex, following RO administration of the enhancer AAV. The fraction of cells mapped to each cortical subclass corresponds to 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. Total number of enhancers examined is n = 149. Asterisks denote the top performing enhancers for each subclass, i.e., the ones with highest proportion of cells mapping to the subclass of interest. **C.** Box plot showing for each cortical population, all enhancer AAVs for which that population was the main enriched target population. The thick black bars represent medians, color-coded boxes represent top and bottom 25%, and whiskers represent top and bottom 10%. Data for individual enhancers is shown as superimposed black circles. **D.** Representative STPT images for five enhancers, with an expanded view of VISp displayed to the right. Dashed arrows connect each image set to its corresponding SSv4 data. In the case of AiE2543m, which labels L2-3_IT cells and Lamp5 cells, pink arrows point to sparse, yet brightly labeled non-L2-3 neurons, which are likely the Lamp5 interneurons. These are overrepresented in SSv4 (B), likely due to the stringent gating strategy in FACS focusing on the highly fluorescent cells. For images, scale bars for the hemisphere and the VISp magnified view are 1.0 and 0.2 mm, respectively.

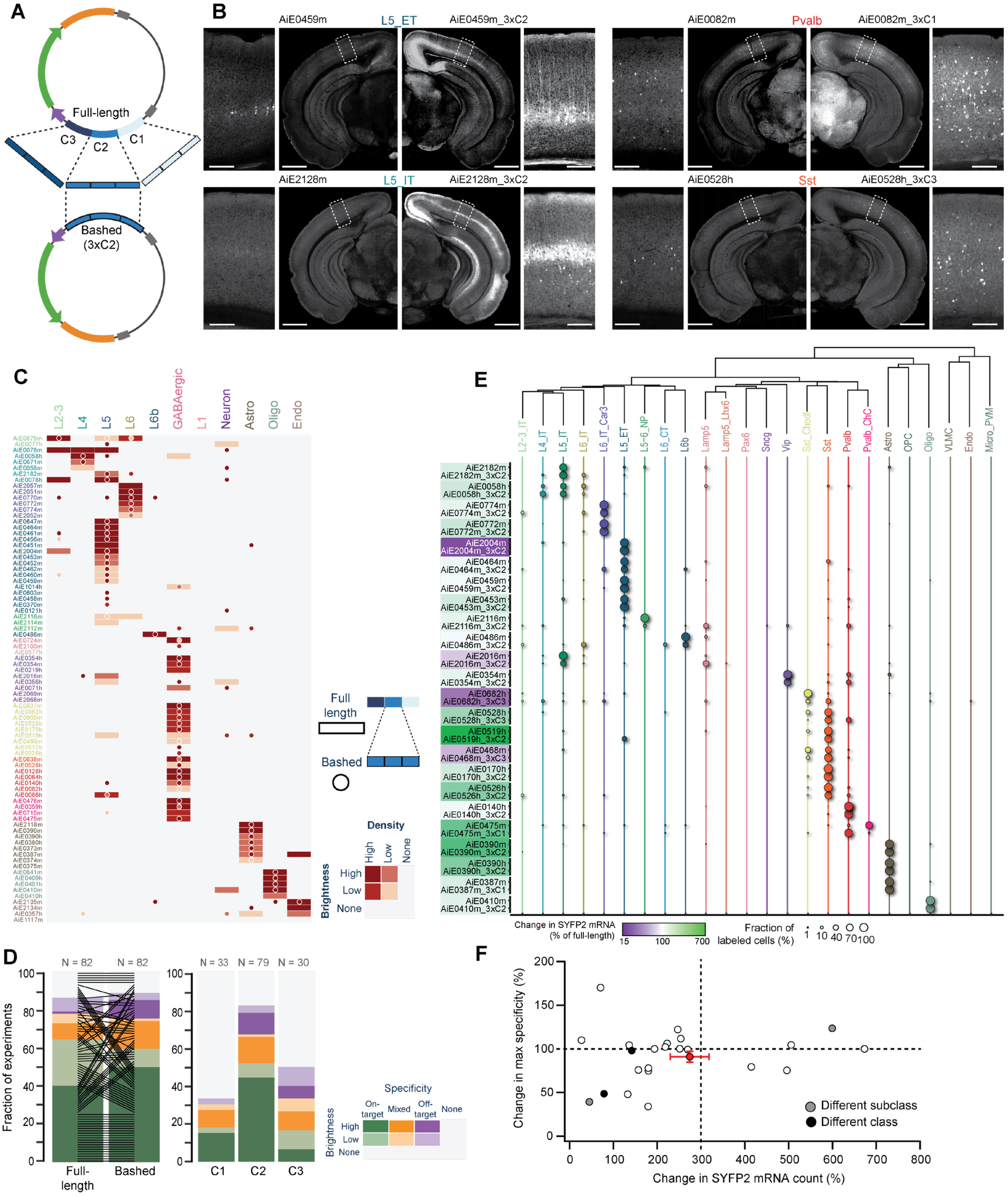

A
Full-length
C3
C2
C1
Bashed
(3xC2)
B
AiE0459m
L5_ET
AiE0459m_3xC2
AiE0082m
Pvalb
AiE0082m_3xC1
AiE2128m
L5_IT
AiE2128m_3xC2
AiE0528h
Sst
AiE0528h_3xC3
C
L2-3
L4
L5
L6
L6b
GABAergic
L1
Neuron
Astro
Oligo
Endo
Full length
Bashed
Density
High
Low
None
Brightness
High
Low
None
E
L2-3_IT
L4_IT
L5_IT
L6_IT
L6_IT_Car3
L5_ET
L5-6_NP
L6_CT
L6b
Lamp5
Lamp5_Lhx6
Pax6
Sncg
Vip
Sst_Chodl
Sst
Pvalb
Pvalb_ChC
Astro
OPC
Oligo
VLMC
Endo
Micro_PVM
AiE2182m
AiE2182m_3xC2
AiE0058h
AiE0058h_3xC2
AiE0774m
AiE0774m_3xC2
AiE0772m
AiE0772m_3xC2
AiE2004m
AiE2004m_3xC2
AiE0464m
AiE0464m_3xC2
AiE0459m
AiE0459m_3xC2
AiE0453m
AiE0453m_3xC2
AiE2116m
AiE2116m_3xC2
AiE0486m
AiE0486m_3xC2
AiE2016m
AiE2016m_3xC2
AiE0354m
AiE0354m_3xC2
AiE0682h
AiE0682h_3xC3
AiE0528h
AiE0528h_3xC3
AiE0519h
AiE0519h_3xC2
AiE0468m
AiE0468m_3xC3
AiE0170h
AiE0170h_3xC2
AiE0526h
AiE0526h_3xC2
AiE0140h
AiE0140h_3xC2
AiE0475m
AiE0475m_3xC1
AiE0390m
AiE0390m_3xC2
AiE0390h
AiE0390h_3xC2
AiE0387m
AiE0387m_3xC1
AiE0410m
AiE0410m_3xC2
Change in SYFP2 mRNA (% of full-length)
15
100
700
Fraction of labeled cells (%)
1
10
40
70
100
D
N = 82
N = 82
N = 33
N = 79
N = 30
Fraction of experiments
100
80
60
40
20
0
Full-length
Bashed
C1
C2
C3
Specificity
On-target
Mixed
Off-target
None
Brightness
High
Low
None
F
Change in max specificity (%)
200
150
100
50
0
Change in SYFP2 mRNA count (%)
0
100
200
300
400
500
600
700
800
Different subclass
Different class

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 four individual full-length enhancers (left) and their best bashed version (right), which was selected according to the combination of brightness and specificity. A magnified view of VISp is shown alongside each hemisection. Scale bars for full section and expanded view are 1.0 and 0.2 mm, respectively. **C.** Heatmap showing the scoring results of epifluorescence image sets of the full-length enhancer (rectangles) alongside its best bashed version (circles); n = 82 pairs. **D.** Summary of the scoring data in (C), sorted according to brightness and specificity of the full-length vs. the bashed enhancer (left) and for the different cores tested (right). **E.** Dot plot of SSv4 data for full-length enhancers and their bashed counterpart shown in pairs, with circle size denoting the fraction of cells mapped to each of the cortical subclasses in each experiment. The color overlaying each pair name corresponds to the relative change in SYFP2 transcript count of the bashed relative to the full-length enhancer; n = 24 pairs. **F.** Change in specificity vs. change in SYFP2 transcript count for all enhancer pairs in (E). Average and SEM for all experiments corresponds to the red dot with error bars. Pairwise comparisons for individual enhancers correspond to white dots if no change in specificity if observed. If the bashed version preferentially labeled a different subclass or class compared to the corresponding full-length enhancer, the dots are grey or black, respectively.

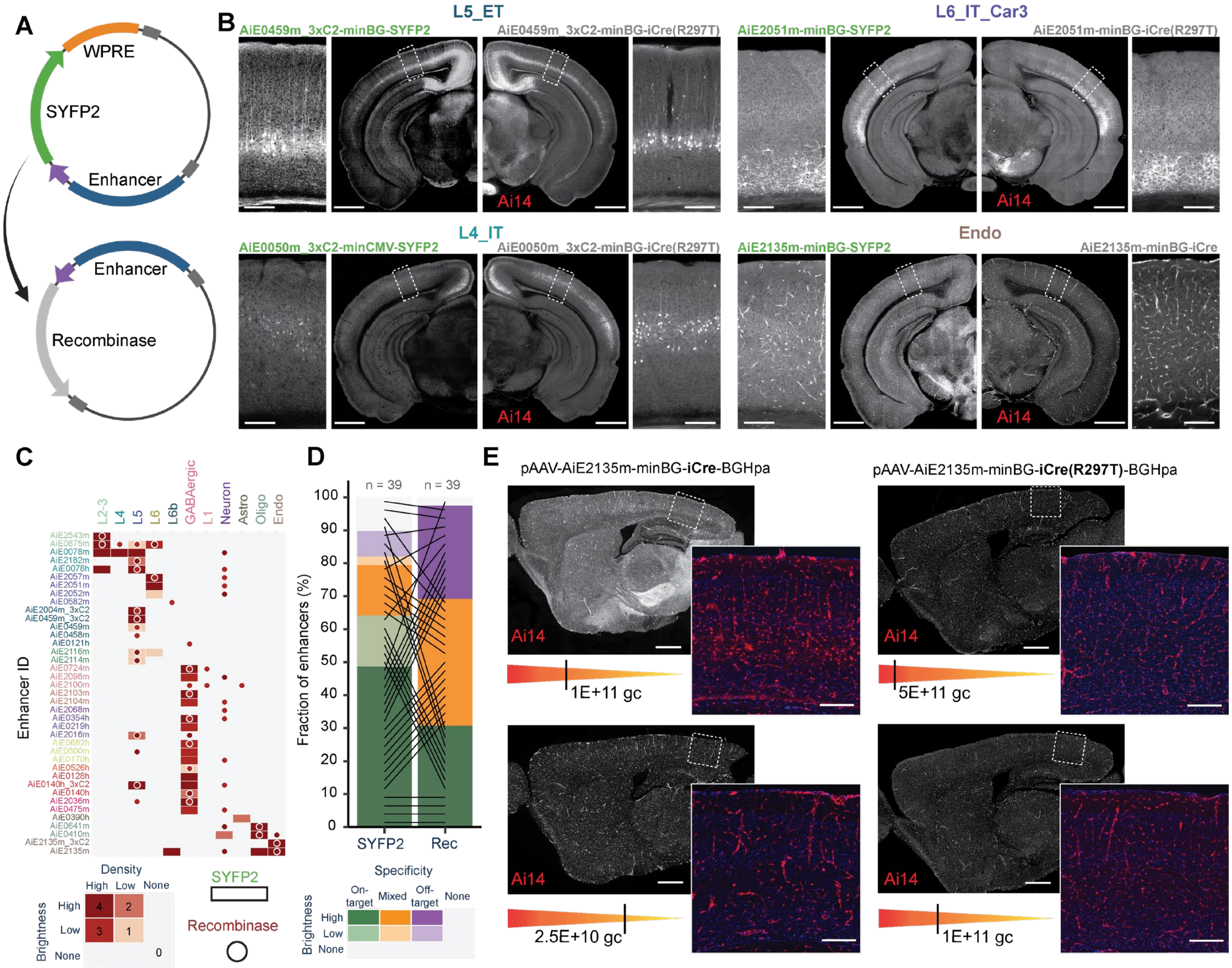


Figure 6: Recombinase-expressing enhancer-AAVs. A. Schematic representation of the vector design. **B.** Representative STPT images of coronal hemisections, showing labeling pattern for four individual enhancers expressing SYFP2 delivered to a wild-type mouse (left) and the same enhancers driving iCre(R297T), delivered to the *Ai14* reporter mouse (right). A magnified view of VISp is shown alongside each hemisection. **C.** Heatmap showing the scoring of epifluorescence imaging data of the enhancers driving SYFP2 (rectangles) or a recombinase (circles); n = 39 pairs. **D.** Summary plot of the scoring data in (C), comparing 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 (gc, genome copies). Scale bars in (B) and (E), 1.0 and 0.2 mm for full section and expanded view, respectively.

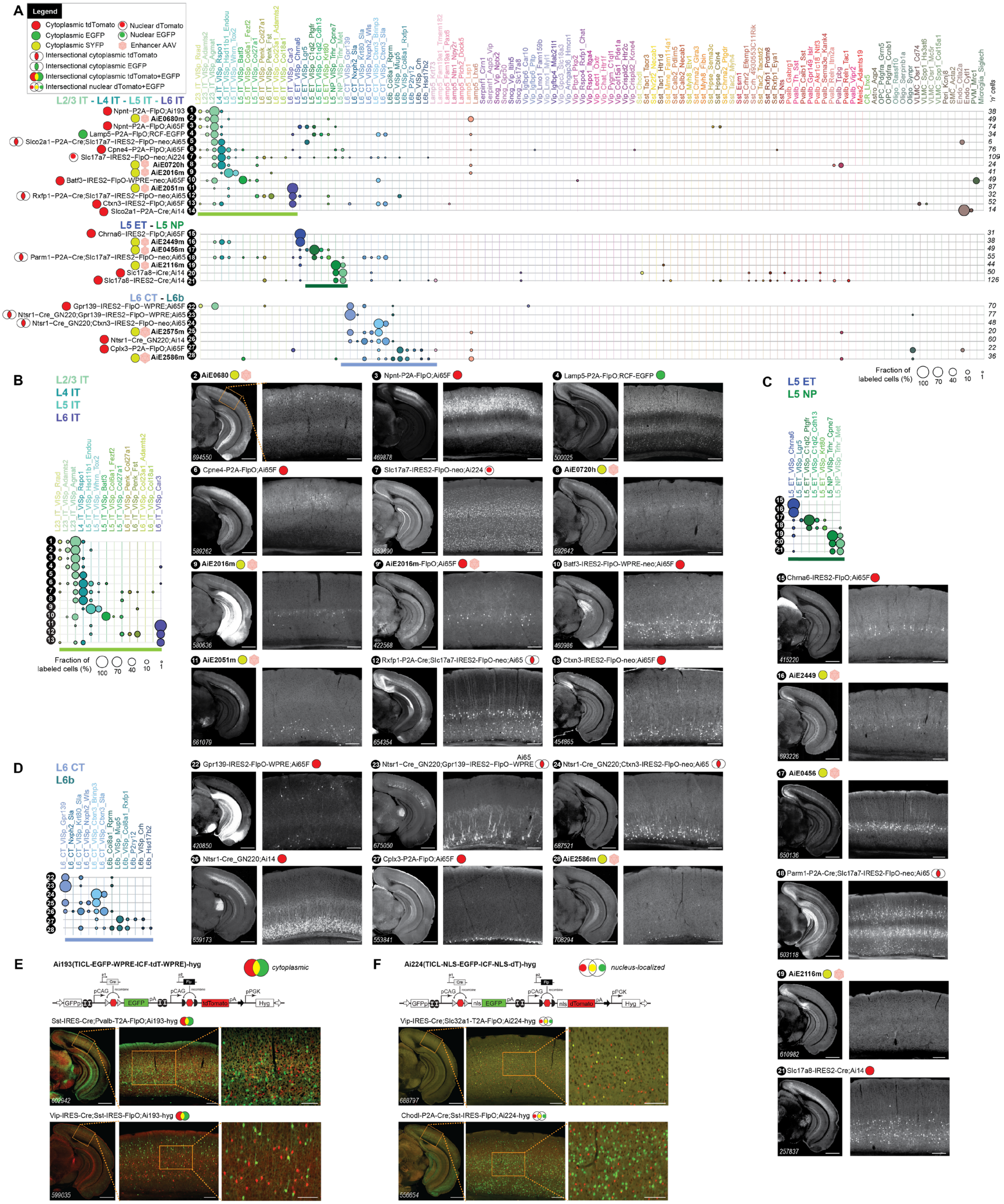

A
Legend
Cytoplasmic tdTomato
Nuclear dTomato
Cytoplasmic EGFP
Nuclear EGFP
Cytoplasmic SYFP
Enhancer AAV
Intersectional cytoplasmic tdTomato
Intersectional cytoplasmic EGFP
Intersectional cytoplasmic tdTomato+EGFP
Intersectional nuclear dTomato+EGFP
L2/3 IT - L4 IT - L5 IT - L6 IT
Npnt-P2A-FlpO;Ai193
AiE0680m
Npnt-P2A-FlpO;Ai65F
Lamp5-P2A-FlpO;RCF-EGFP
Slco2a1-P2A-Cre;Slc17a7-IRES2-FlpO-neo;Ai65
Cpne4-P2A-FlpO;Ai65F
Slc17a7-IRES2-FlpO-neo;Ai224
AiE0720h
AiE2016m
Batf3-IRES2-FlpO-WPRE-neo;Ai65F
AiE2051m
Rxfp1-P2A-Cre;Slc17a7-IRES2-FlpO-neo;Ai65
Ctxn3-IRES2-FlpO-neo;Ai65F
Slco2a1-P2A-Cre;Ai14
L5 ET - L5 NP
Chrna6-IRES2-FlpO;Ai65F
AiE2449m
AiE0456m
Parm1-P2A-Cre;Slc17a7-IRES2-FlpO-neo;Ai65
AiE2116m
Slc17a8-iCre;Ai14
Slc17a8-IRES2-Cre;Ai14
L6 CT - L6b
Gpr139-IRES2-FlpO-WPRE;Ai65F
Ntsr1-Cre_GN220;Gpr139-IRES2-FlpO-WPRE;Ai65
Ntsr1-Cre_GN220;Ctxn3-IRES2-FlpO-neo;Ai65
AiE2575m
Ntsr1-Cre_GN220;Ai14
Cplx3-P2A-FlpO;Ai65F
AiE2586m
'n' cells
38
49
74
34
6
76
109
24
41
49
87
32
52
14
31
38
49
55
44
50
126
70
77
48
20
60
22
36
Fraction of labeled cells (%)
100
70
40
10
1
B
L2/3 IT
L4 IT
L5 IT
L6 IT
2 AiE0680
3 Npnt-P2A-FlpO;Ai65F
4 Lamp5-P2A-FlpO;RCF-EGFP
6 Cpne4-P2A-FlpO;Ai65F
7 Slc17a7-IRES2-FlpO-neo;Ai224
8 AiE0720h
9 AiE2016m
9' AiE2016m-FlpO;Ai65F
10 Batf3-IRES2-FlpO-WPRE-neo;Ai65F
11 AiE2051m
12 Rxfp1-P2A-Cre;Slc17a7-IRES2-FlpO-neo;Ai65
13 Ctxn3-IRES2-FlpO-neo;Ai65F
694550
469878
500025
589262
653690
692642
580636
422568
460986
661079
654354
454865
C
L5 ET
L5 NP
15 Chrna6-IRES2-FlpO;Ai65F
16 AiE2449
17 AiE0456
18 Parm1-P2A-Cre;Slc17a7-IRES2-FlpO-neo;Ai65
19 AiE2116m
21 Slc17a8-IRES2-Cre;Ai14
415220
693226
650136
603118
610982
257837
D
L6 CT
L6b
22 Gpr139-IRES2-FlpO-WPRE;Ai65F
23 Ntsr1-Cre_GN220;Gpr139-IRES2-FlpO-WPRE Ai65
24 Ntsr1-Cre_GN220;Ctxn3-IRES2-FlpO-neo;Ai65
26 Ntsr1-Cre_GN220;Ai14
27 Cplx3-P2A-FlpO;Ai65F
28 AiE2586m
420850
675050
687521
659173
553841
708294
E
Ai193(TICL-EGFP-WPRE-ICF-tdT-WPRE)-hyg
cytoplasmic
GFPp
pCAG
Cre
recombinase
EGFP
pA
Flp
tdTomato
pPGK
Hyg
Sst-IRES-Cre;Pvalb-T2A-FlpO;Ai193-hyg
602942
Vip-IRES-Cre;Sst-IRES-FlpO;Ai193-hyg
599035
F
Ai224(TICL-NLS-EGFP-ICF-NLS-dT)-hyg
nucleus-localized
nls
dTomato
Vip-IRES-Cre;Slc32a1-T2A-FlpO;Ai224-hyg
668797
Chodl-P2A-Cre;Sst-IRES-FlpO;Ai224-hyg
556654

Figure 7: Characterization of new transgenic driver lines, preferentially targeting glutamatergic subclasses and clusters. A. scRNA-seq (SSv4) data showing distribution of labeled cells mapped to the mouse VISp taxonomy at the cluster level. 28 different experimental conditions (tools numbered **1-28**) were grouped into panels according to predominant cell types labeled. They may include transgenic drivers and reporters as indicated or may be wild-type animals that received a systemic delivery of enhancer viruses (marked by pink hexagons). **B.** Focused view of tools **1-13** that label IT neurons from Layer 2-3, L4, L5 and L6, including the previously reported enhancer AiE2016m (originally called mscRE16)[22] expressing SYFP in a wild-type animal (**9**) and driving a FlpO recombinase in Ai65F (**9***). **C.** Same as in (B) for tools **14-21** that label ET and NP neurons in L5. **D.** Same as in (B) for tools **22-28** that label L6_CT and the L6b neurons. **E.** Schematics and representative sections from STPT data for a new Flp-Cre:AND/OR reporter line, *Ai193* (TICL-EGFP-WPRE-ICF-tdT-WPRE)-hyg. The line was tested in triple transgenic crosses with two recombinase lines. Cells that express the Cre recombinase are EGFP-positive (green) and those that express FlpO are tdTomato-positive (red). Cells that express both appear yellow. **F.** Same as in (E) for a new reporter line, *Ai224* (TICL-NLS-EGFP-ICF-NLS-dT)-hyg, where fluorophores are nucleus-localized. Scale bars: 1.0 and 0.2 mm for full section and expanded view; 0.1 mm for further expanded view in (E) and (F).

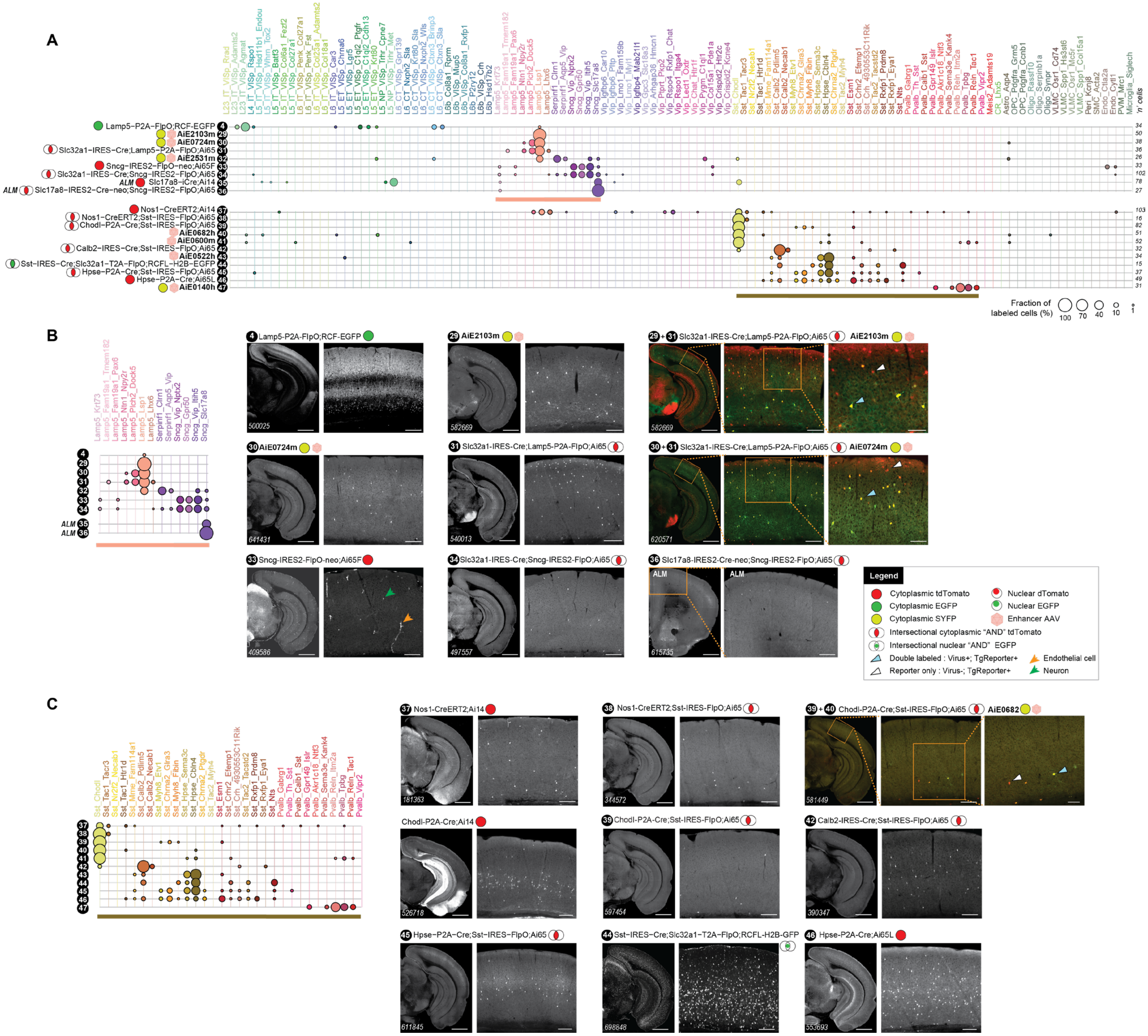


Figure 8: Characterization of new transgenic driver lines, preferentially targeting GABAergic subclasses and clusters. A. scRNA-seq (SSv4) data, same as in Figure 7A but for 20 lines (tools numbered **4, 29-34**, **37-47**) targeting GABAergic types grouped into panels according to predominant cell types labeled. **B.** Focused representation of the same data as in (A) for tools labeling clusters within Lamp5 and Sncg GABAergic cortical subclasses. All data (tools numbered **4**, **29-34**) are from VISp, except for tools **35** and **36** that were characterized in cortical area ALM. The tools **29+31** and **30+31** show expression of two different Lamp5-expressing viruses in the triple transgenic mouse that expresses highly specifically only in Lamp5 interneurons (*Slc32a1-IRES-Cre;Lamp5-P2A-FlpO;Ai65* – see tool **31**) .**C.** Focused representation of the same data as in (A) for tools **37-47** labeling clusters in Sst and Pvalb GABAergic cortical subclasses. Scale bars: 1.0 and 0.2 mm for full section and expanded view; 0.1 mm for further expanded view for tools **29+31**, **30+31**, **39+40**.

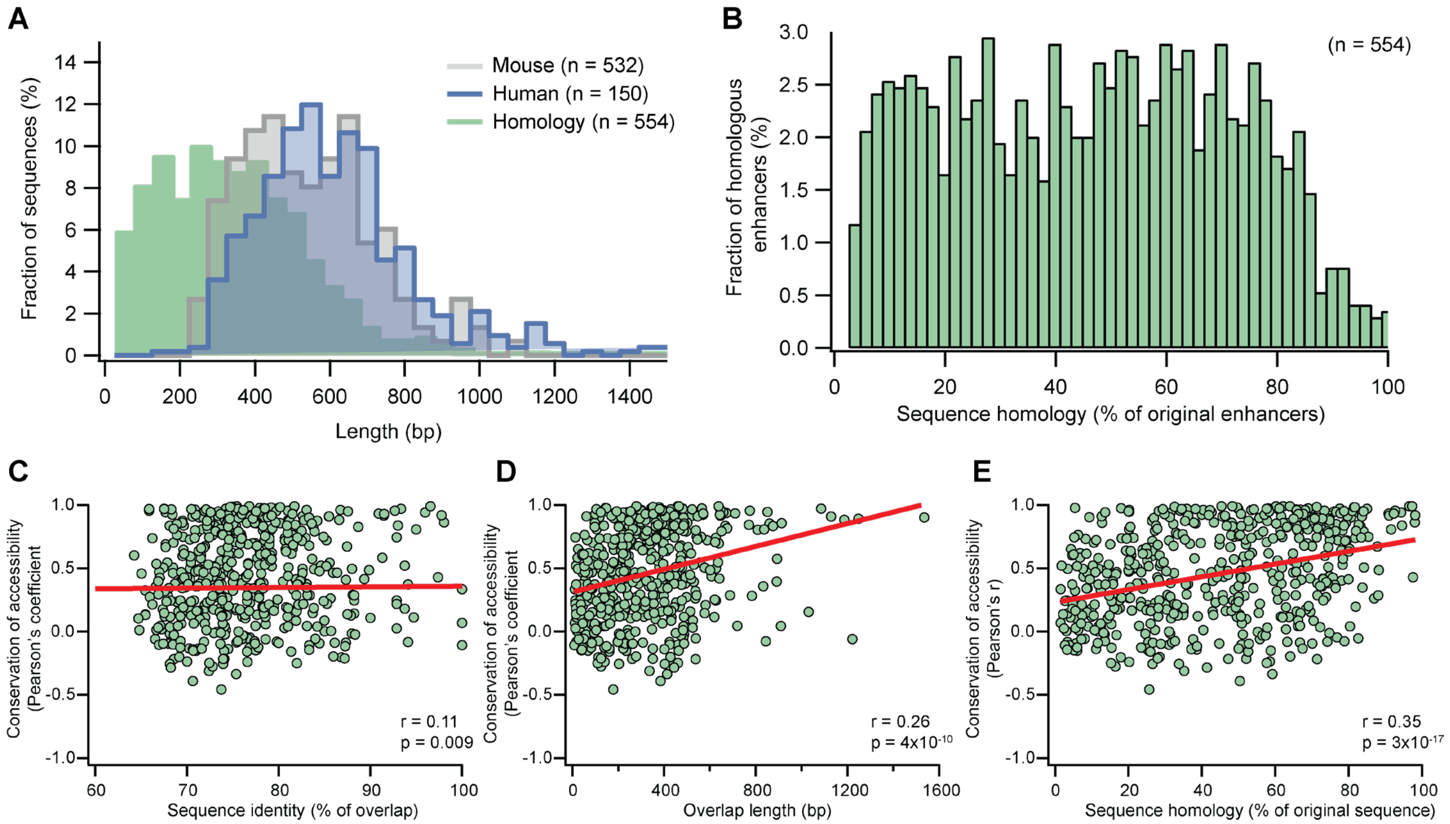


Figure S1: Analysis of sequence alignment across species. A. Histograms showing the proportion of mouse (grey) and human (blue) sequences, according to their length along with a histogram showing the length distribution of the liftover sequences (full green). Bin size = 50 bp. **B.** Histogram of the 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.

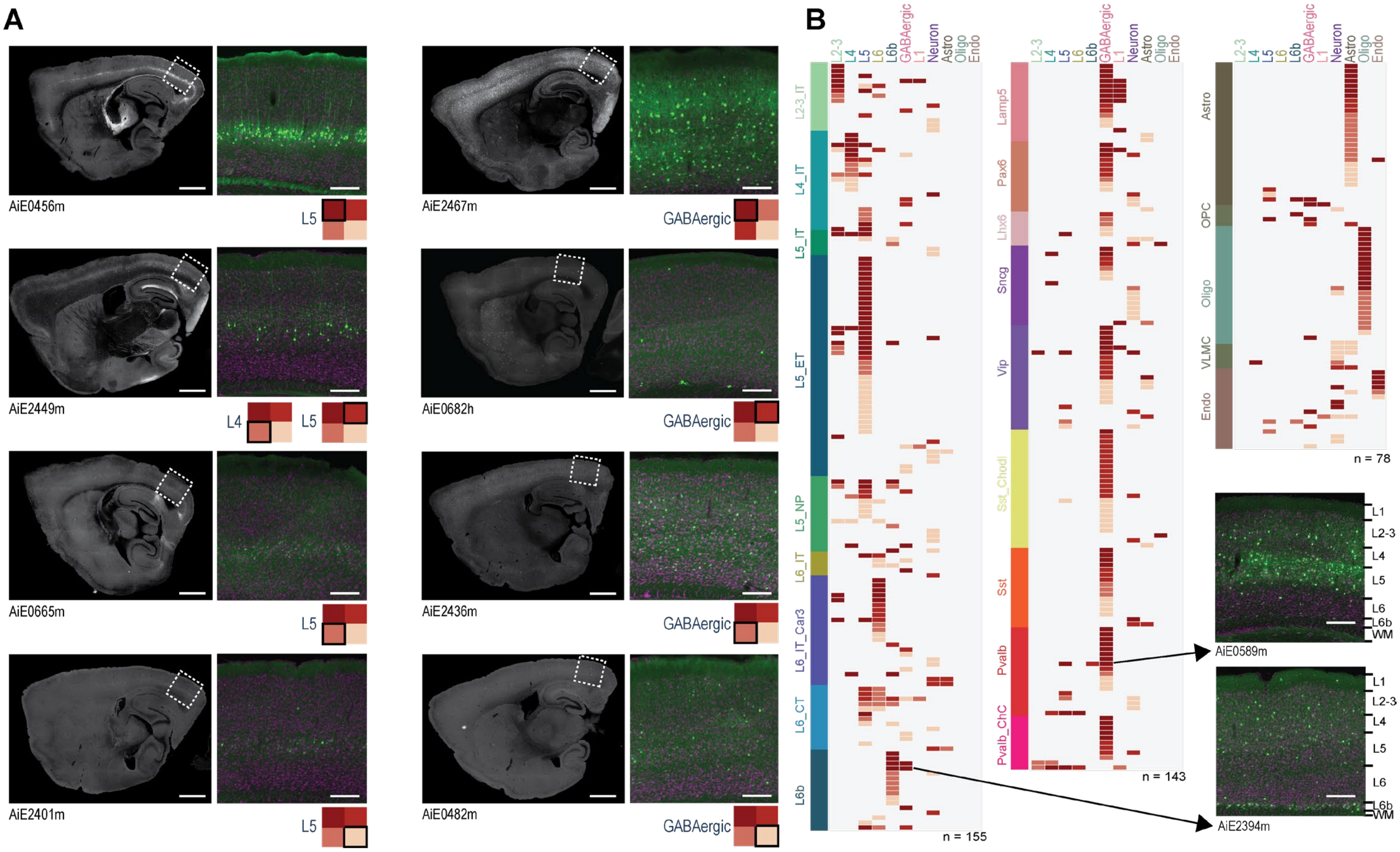


Figure S2: Enhancer scoring based on epifluorescence image sets. A. Representative sagittal images from eight experiments with focus on two cell categories: L5 (left column) or GABAergic (right). Scoring according to the scheme in Figure 3B is shown below each set of images. **B.** A heat map of all enhancers evaluated, which produced any labeling in the neocortex, arranged according to the cell population where their accessibility was highest, showing the identity of the labeled cell population and the labeling quality, according to the scheme in Figure 3B. The number of individual experiments in each category is shown below each plot (total n = 376). For images, scale bars for full section and expanded views are 1.0 and 0.2 mm, respectively.

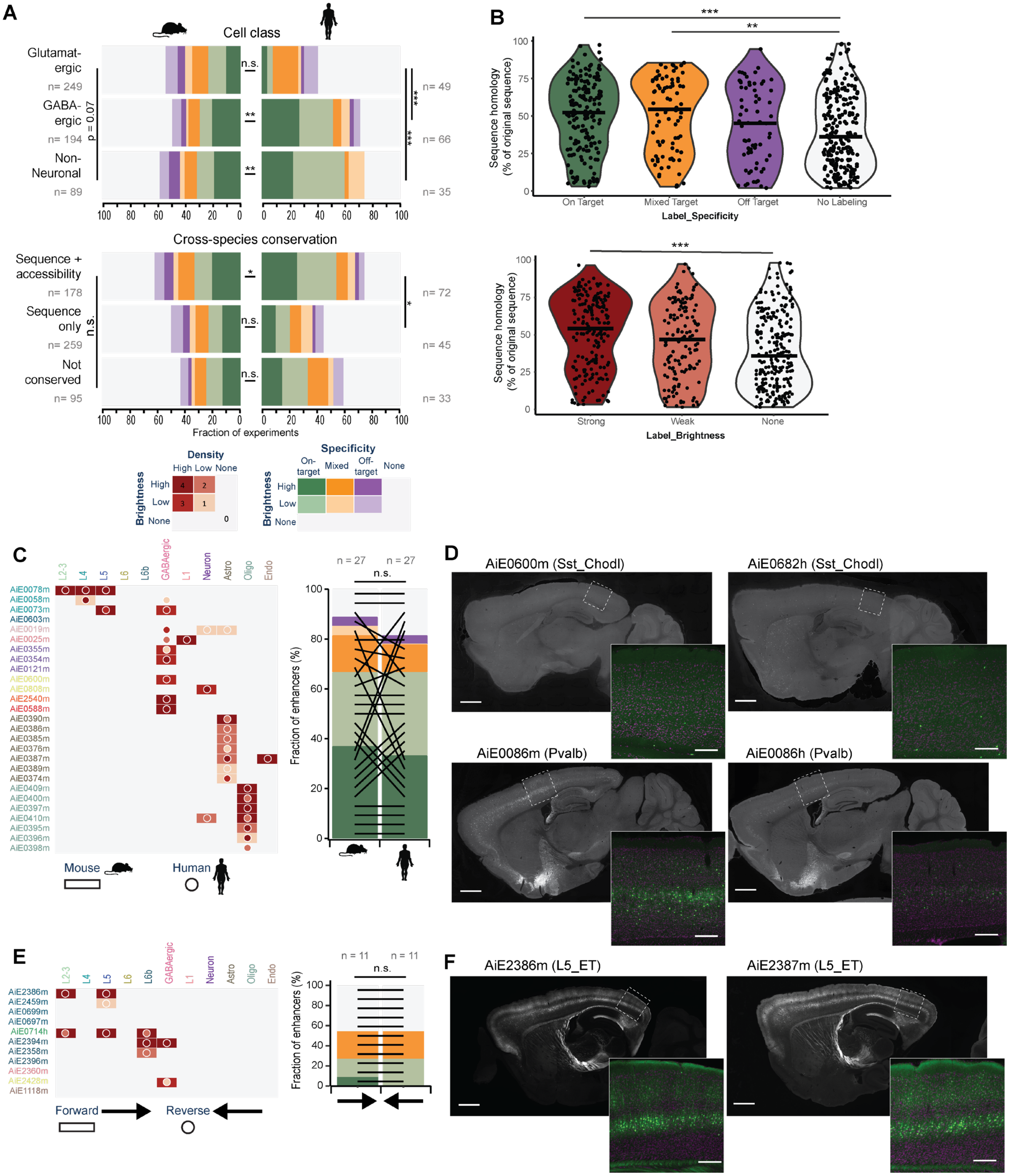
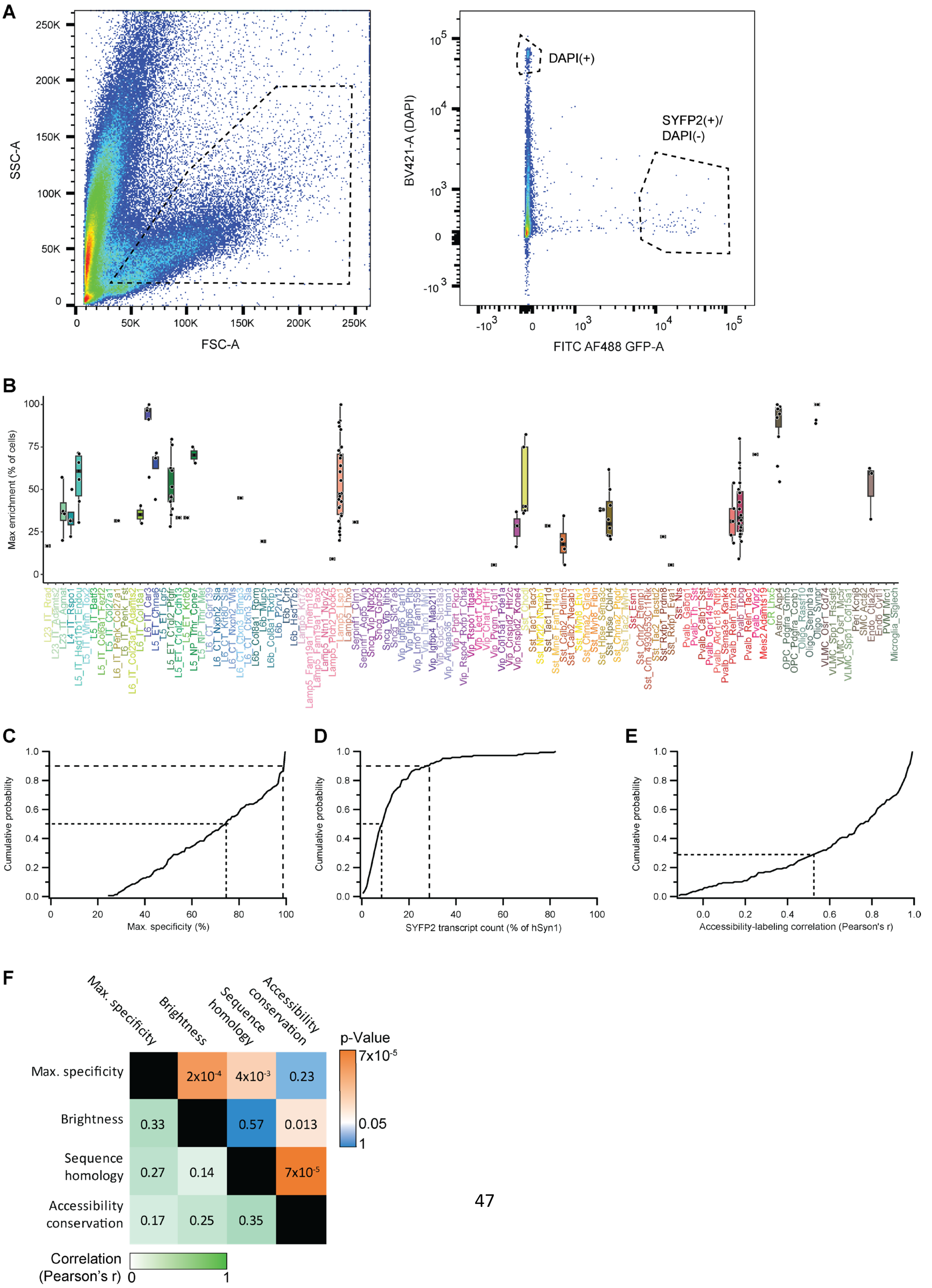

A
Cell class
Glutamat-ergic
n= 249
GABA-ergic
n= 194
Non-Neuronal
n= 89
p = 0.07
n.s.
**
**
***
***
n= 49
n= 66
n= 35
100 80 60 40 20 0 0 20 40 60 80 100
Cross-species conservation
Sequence + accessibility
n= 178
Sequence only
n= 259
Not conserved
n= 95
n.s.
*
n.s.
n.s.
*
n= 72
n= 45
n= 33
Fraction of experiments
Density
High Low None
Brightness
High
Low
None
4 2 3 1 0
Specificity
On-target
Mixed
Off-target
None
B
***
**
Sequence homology (% of original sequence)
100 75 50 25 0
On Target
Mixed Target
Off Target
No Labeling
Label_Specificity
***
Strong
Weak
None
Label_Brightness
C
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
Mouse
Human
n = 27
n = 27
n.s.
Fraction of enhancers (%)
100 80 60 40 20 0
D
AiE0600m (Sst_Chodl)
AiE0682h (Sst_Chodl)
AiE0086m (Pvalb)
AiE0086h (Pvalb)
E
L2-3 L4 L5 L6 L6b GABAergic L1 Neuron Astro Oligo Endo
AiE2386m
AiE2459m
AiE0699m
AiE0697m
AiE0714h
AiE2394m
AiE2358m
AiE2396m
AiE2360m
AiE2428m
AiE1118m
Forward
Reverse
n = 11
n = 11
n.s.
Fraction of enhancers (%)
100 80 60 40 20 0
F
AiE2386m (L5_ET)
AiE2387m (L5_ET)

Figure S3: Effects of cross-species conservation and sequence orientation on enhancer performance. **A.** Summary plot of enhancer scoring data according to their genome of origin (top) and cross-species conservation of sequence and accessibility (bottom). **B.** Violin plots showing the degree of mouse-human sequence homology according to the different specificity (top) and brightness (bottom) categories. P-values were calculated with one-way ANOVA, followed by a pairwise post-hoc analysis (Tukey) corrected for multiple comparisons (Bonferroni). Significance levels are the same as in (A). **C.** 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). **D.** Representative epifluorescence images of sagittal sections, showing labeling pattern for two individual mouse enhancers (left) and their human orthologs (right). **E.** Summary plot of the scoring data for complementary oriented sequences (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 sagittal sections, showing identical labeling pattern for a complementary pair. P-values were calculated using a chi-squared test. P-values < 0.05, < 0.01 and < 0.001 are denoted by *, **, ***, respectively, n.s. = not significant. Scale bars for full section and higher-magnification view = 1.0 and 0.2 mm, respectively.



	Max. specificity	Brightness	Sequence homology	Accessibility conservation
Max. specificity		$2x10^{-4}$	$4x10^{-3}$	0.23
Brightness	0.33		0.57	0.013
Sequence homology	0.27	0.14		$7x10^{-5}$
Accessibility conservation	0.17	0.25	0.35	

Figure S4: Distribution of specificity and brightness across enhancer AAVs. A. Representative plots of the FACS gating strategy for selective collection of cells labeled by the Lamp5 enhancer AiE2103m: Forward (FSC-A) and side scatter (SSC-A) were used to select objects matching size and granularity of cortical cells (left) and this fraction was further separated according to signal detected in the DAPI and FITC channels (right), in order to avoid collection of DAPI+ cells, whose membrane is likely compromised. Dashed boxes indicate the gates applied for sample collection. **B.** Box plots showing the distribution of enhancer maximum specificity in each of the cortical clusters. **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.

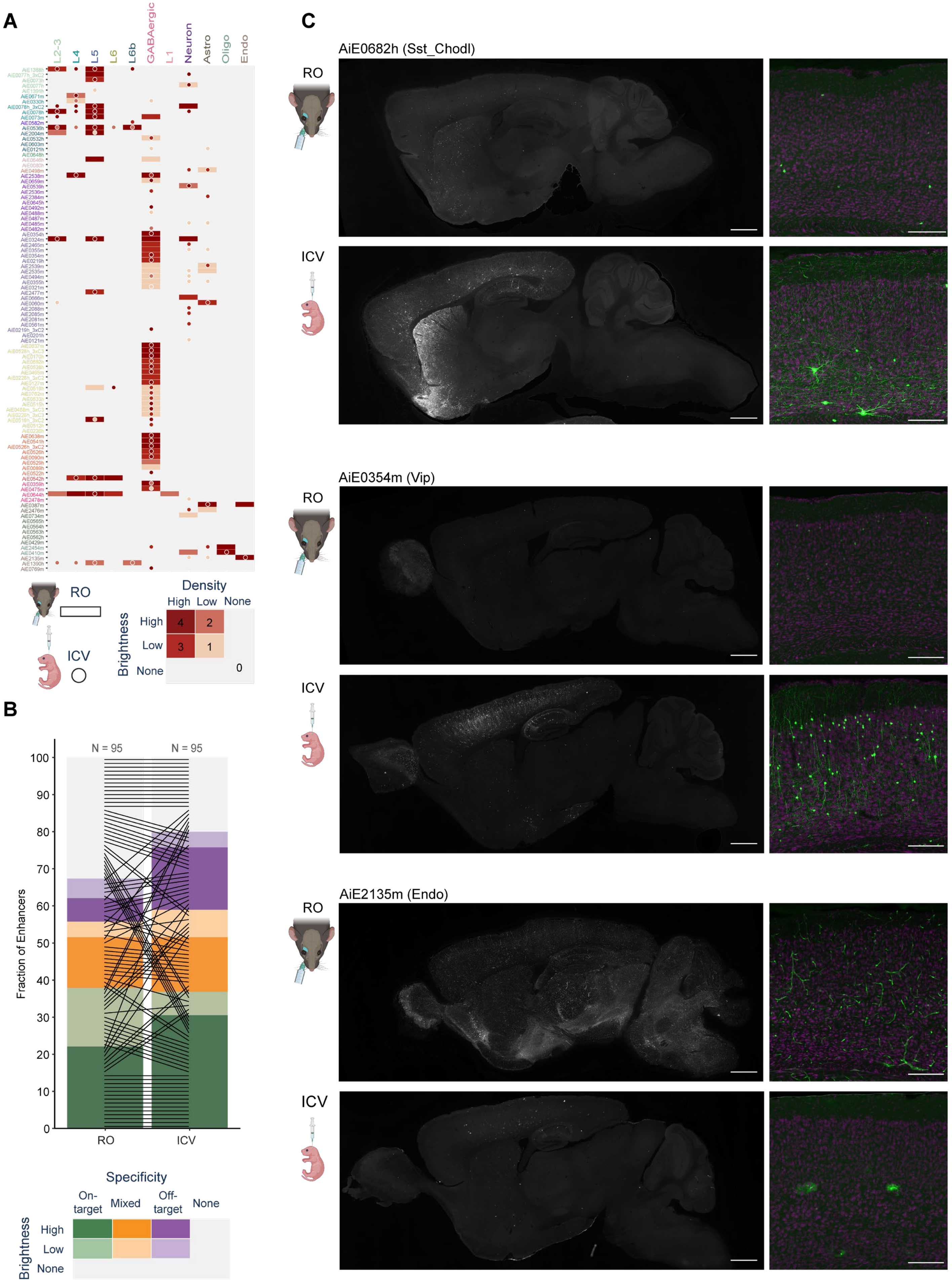

A
L2-3
L4
L5
L6
L6b
GABAergic
L1
Neuron
Astro
Oligo
Endo
RO
ICV
Density
High
Low
None
Brightness
High
Low
None
4
2
3
1
0
B
N = 95
N = 95
Fraction of Enhancers
0
10
20
30
40
50
60
70
80
90
100
RO
ICV
Specificity
On-target
Mixed
Off-target
None
Brightness
High
Low
None
C
AiE0682h (Sst_Chodl)
RO
ICV
AiE0354m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

Figure S5: Comparison RO and ICV viral delivery routes by image scoring data. 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 sagittal sections of three 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 = 1.0 and 0.2 mm, respectively.

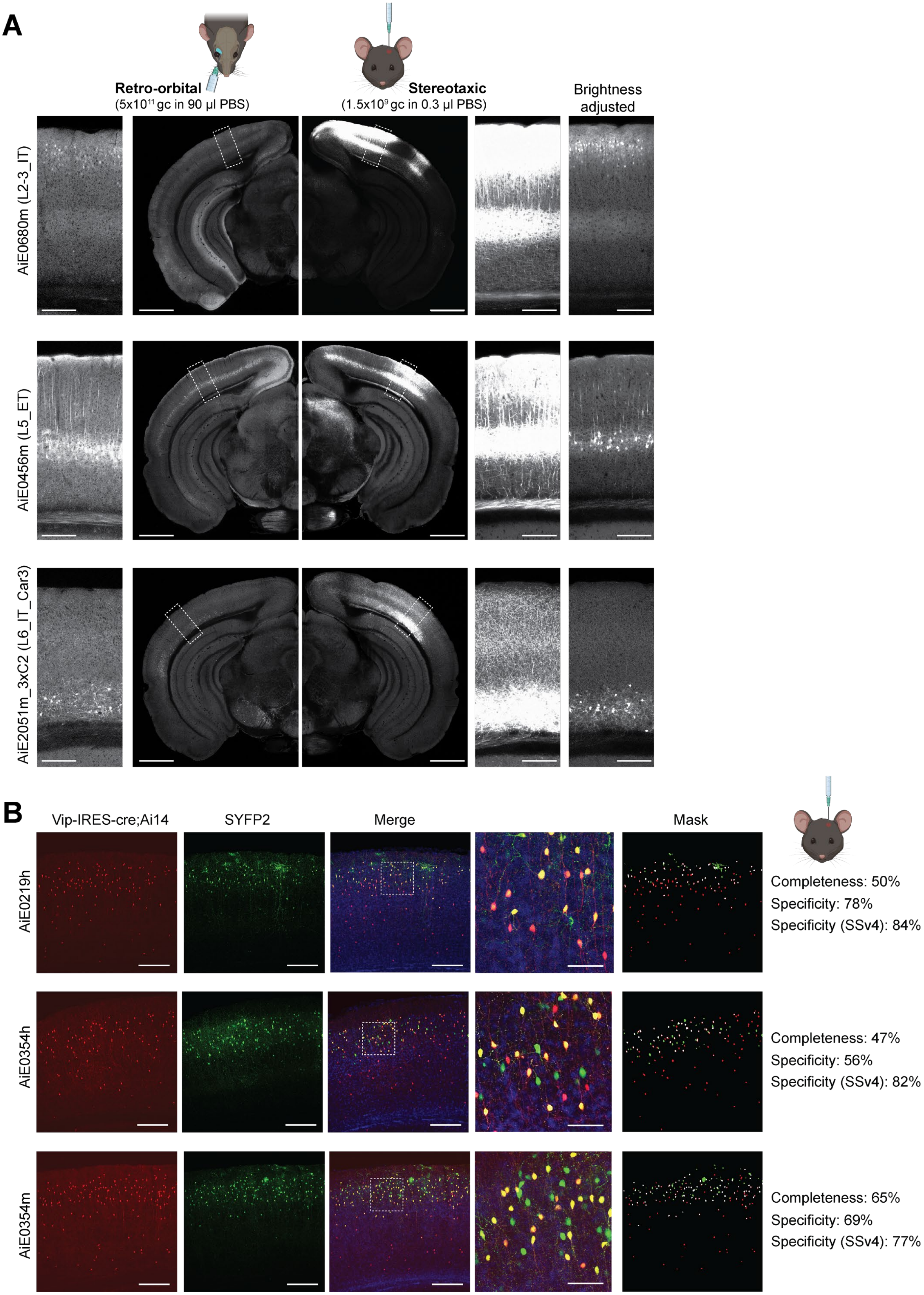

A
Retro-orbital
(5x10¹¹ gc in 90 µl PBS)
Stereotaxic
(1.5x10⁹ gc in 0.3 µl PBS)
Brightness adjusted
AiE0680m (L2-3_IT)
AiE0456m (L5_ET)
AiE2051m_3xC2 (L6_IT_Car3)
B
Vip-IRES-cre;Ai14
SYFP2
Merge
Mask
AiE0219h
Completeness: 50%
Specificity: 78%
Specificity (SSv4): 84%
AiE0354h
Completeness: 47%
Specificity: 56%
Specificity (SSv4): 82%
AiE0354m
Completeness: 65%
Specificity: 69%
Specificity (SSv4): 77%

Figure S6: Stereotaxic delivery of enhancer AAVs. A. Stereotaxic delivery into VISp of three enhancers AAVs targeting different subclasses of glutamatergic neurons, resulted in strong, layer restricted SYFP2 expression. Scale bars for full section and expanded view = 1.0 and 0.2 mm, respectively. **B.** Stereotaxic delivery of three enhancers targeting Vip interneurons (green), delivered to the VISp of *Vip-IRES-Cre;Ai14* double transgenic line (red). 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. Specificity results were compared with SSv4 measurements for each vector, following RO delivery of 5x10^{11} genome copies (gc). n = 1 experimental animal for all experiments shown in this figure. Scale bars for full VISp view and expanded view = 0.2 and 0.05 mm, respectively.

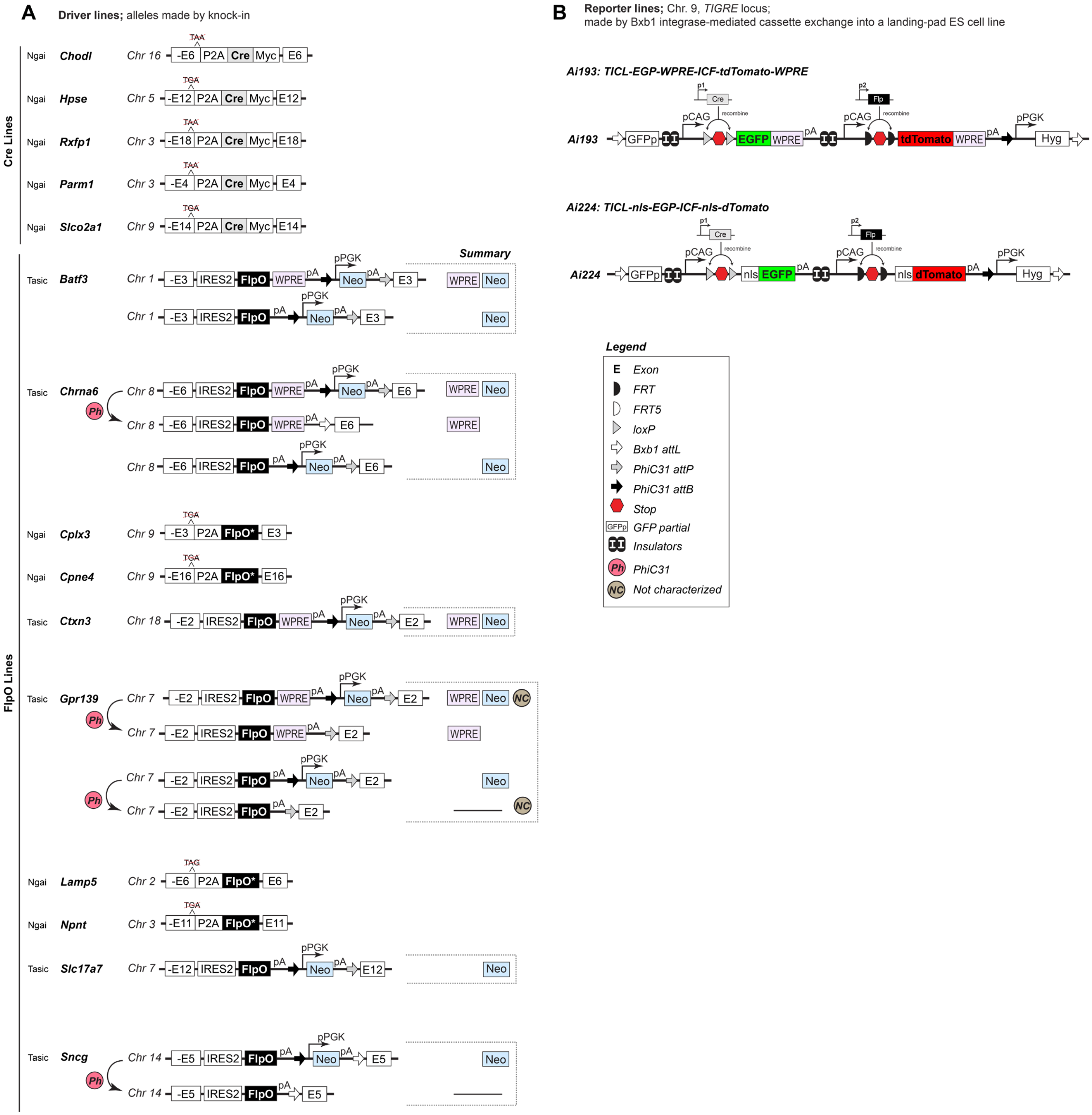


Figure S7: Transgenic line designs. A. Schematics depicting design of the 15 driver lines. Of these, five express Cre recombinase whereas 10 express FlpO. For some lines, such as *Chrna6-IRES2-FlpO*, we have versions with WPRE, with Neo present as well as with Neo removed allowing us to compare expression patterns in all three. 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*.

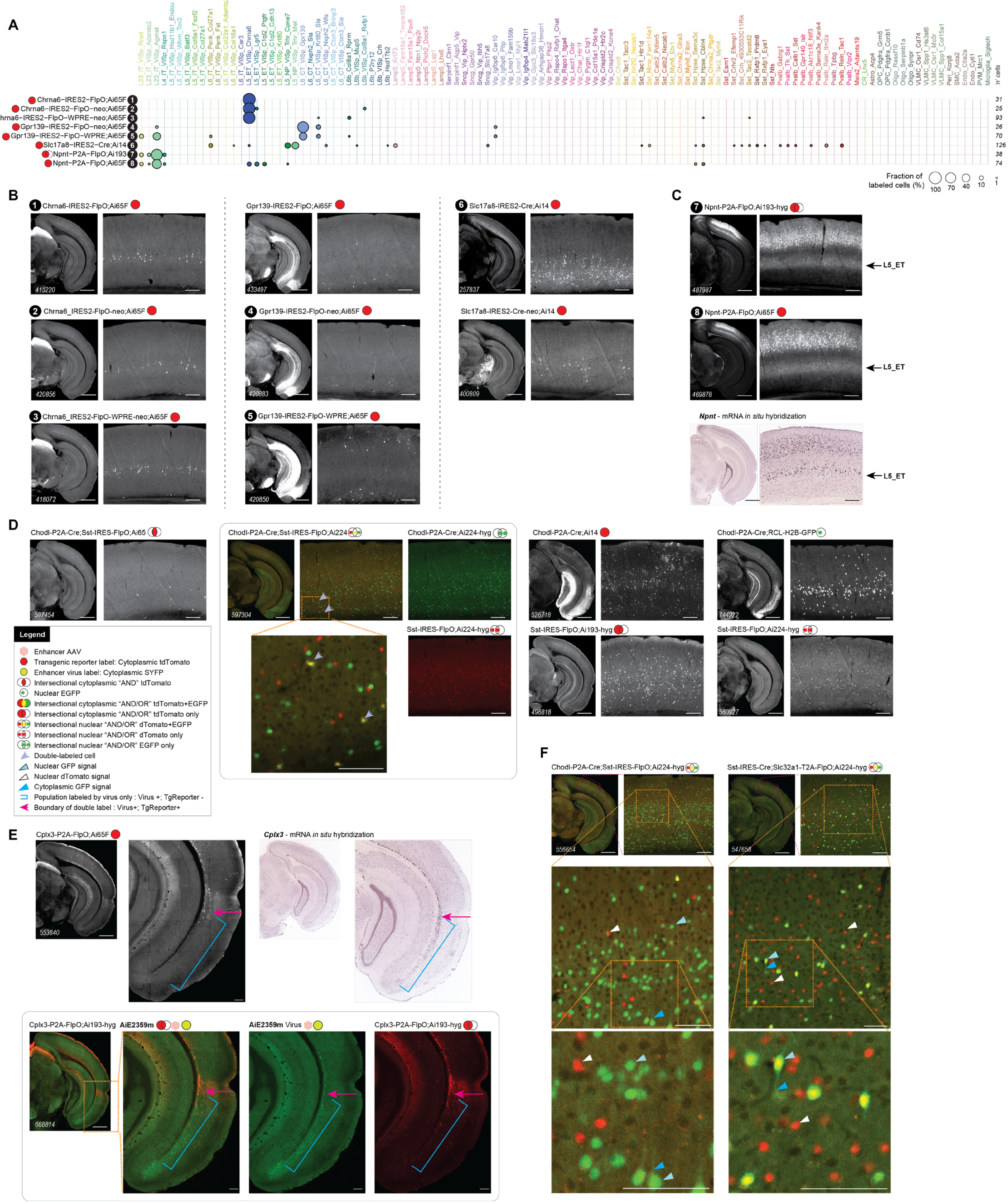

A
L23_IT_VISp_Rrad
L23_IT_VISp_Adamts2
L23_IT_VISp_Agmat
L4_IT_VISp_Rspo1
L5_IT_VISp_Hsd11b1_Endou
L5_IT_VISp_Whrn_Tox2
L5_IT_VISp_Batf3
L5_IT_VISp_Col6a1_Fezf2
L5_IT_VISp_Col27a1
L6_IT_VISp_Penk_Col27a1
L6_IT_VISp_Penk_Fst
L6_IT_VISp_Col23a1_Adamts2
L6_IT_VISp_Col18a1
L6_IT_VISp_Car3
L5_ET_VISp_Chrna6
L5_ET_VISp_Lgr5
L5_ET_VISp_C1ql2_Ptgfr
L5_ET_VISp_C1ql2_Cdh13
L5_ET_VISp_Krt80
L5_NP_VISp_Trhr_Cpne7
L5_NP_VISp_Trhr_Met
L6_CT_VISp_Gpr139
L6_CT_Nxph2_Sla
L6_CT_VISp_Krt80_Sla
L6_CT_VISp_Nxph2_Wls
L6_CT_VISp_Ctxn3_Brinp3
L6_CT_VISp_Ctxn3_Sla
L6b_Col8a1_Rprm
L6b_VISp_Mup5
L6b_VISp_Col8a1_Rxfp1
L6b_P2ry12
L6b_VISp_Crh
L6b_Hsd17b2
Lamp5_Krt73
Lamp5_Fam19a1_Tmem182
Lamp5_Fam19a1_Pax6
Lamp5_Ntn1_Npy2r
Lamp5_Plch2_Dock5
Lamp5_Lsp1
Lamp5_Lhx6
Serpinf1_Clrn1
Serpinf1_Aqp5_Vip
Sncg_Vip_Nptx2
Sncg_Gpr50
Sncg_Vip_Itih5
Sncg_Slc17a8
Vip_Igfbp6_Car10
Vip_Igfbp6_Pltp
Vip_Lmo1_Fam159b
Vip_Lmo1_Myl1
Vip_Igfbp4_Mab21l1
Vip_Gpc3_Slc18a3
Vip_Arhgap36_Hmcn1
Vip_Ptprt_Pkp2
Vip_Rspo4_Rxfp1_Chat
Vip_Rspo1_Itga4
Vip_Lect1_Oxtr
Vip_Chat_Htr1f
Vip_Pygm_C1ql1
Vip_Col15a1_Pde1a
Vip_Crispld2_Htr2c
Vip_Crispld2_Kcne4
Sst_Chodl
Sst_Tac1_Tacr3
Sst_Nr2f2_Necab1
Sst_Tac1_Htr1d
Sst_Mme_Fam114a1
Sst_Calb2_Pdlim5
Sst_Calb2_Necab1
Sst_Myh8_Etv1
Sst_Chrna2_Glra3
Sst_Myh8_Fibin
Sst_Hpse_Sema3c
Sst_Hpse_Cbln4
Sst_Chrna2_Ptgdr
Sst_Tac2_Myh4
Sst_Esm1
Sst_Crhr2_Efemp1
Sst_Crh_4930553C11Rik
Sst_Tac2_Tacstd2
Sst_Rxfp1_Prdm8
Sst_Rxfp1_Eya1
Sst_Nts
Pvalb_Gabrg1
Pvalb_Th_Sst
Pvalb_Calb1_Sst
Pvalb_Gpr149_Islr
Pvalb_Akr1c18_Ntf3
Pvalb_Sema3e_Kank4
Pvalb_Reln_Itm2a
Pvalb_Tpbg
Pvalb_Reln_Tac1
Pvalb_Vipr2
Meis2_Adamts19
CR_Lhx5
Astro_Aqp4
OPC_Pdgfra_Grm5
OPC_Pdgfra_Ccnb1
Oligo_Rassf10
Oligo_Serpinb1a
Oligo_Synpr
VLMC_Osr1_Cd74
VLMC_Spp1_Hs3st6
VLMC_Osr1_Mc5r
VLMC_Spp1_Col15a1
Peri_Kcnj8
SMC_Acta2
Endo_Ctla2a
Endo_Cytl1
PVM_Mrc1
Microglia_Siglech
'n' cells
Chrna6–IRES2–FlpO;Ai65F 1 31
Chrna6–IRES2–FlpO–neo;Ai65F 2 25
Chrna6–IRES2–FlpO–WPRE–neo;Ai65F 3 93
Gpr139–IRES2–FlpO–neo;Ai65F 4 26
Gpr139–IRES2–FlpO–WPRE;Ai65F 5 70
Slc17a8–IRES2–Cre;Ai14 6 126
Npnt–P2A–FlpO;Ai193 7 38
Npnt–P2A–FlpO;Ai65F 8 74
Fraction of labeled cells (%)
100
70
40
10
1
B
1 Chrna6-IRES2-FlpO;Ai65F
415220
2 Chrna6_IRES2-FlpO-neo;Ai65F
420856
3 Chrna6_IRES2-FlpO-WPRE-neo;Ai65F
418072
Gpr139-IRES2-FlpO;Ai65F
433497
4 Gpr139-IRES2-FlpO-neo;Ai65F
420883
5 Gpr139-IRES2-FlpO-WPRE;Ai65F
420850
6 Slc17a8-IRES2-Cre;Ai14
257837
Slc17a8-IRES2-Cre-neo;Ai14
400809
C
7 Npnt-P2A-FlpO;Ai193-hyg
487987
L5_ET
8 Npnt-P2A-FlpO;Ai65F
469878
L5_ET
Npnt - mRNA in situ hybridization
L5_ET
D
Chodl-P2A-Cre;Sst-IRES-FlpO;Ai65
597454
Chodl-P2A-Cre;Sst-IRES-FlpO;Ai224
597304
Chodl-P2A-Cre;Ai224-hyg
Sst-IRES-FlpO;Ai224-hyg
Chodl-P2A-Cre;Ai14
526718
Chodl-P2A-Cre;RCL-H2B-GFP
744372
Sst-IRES-FlpO;Ai193-hyg
496818
Sst-IRES-FlpO;Ai224-hyg
560927
Legend
Enhancer AAV
Transgenic reporter label: Cytoplasmic tdTomato
Enhancer virus label: Cytoplasmic SYFP
Intersectional cytoplasmic "AND" tdTomato
Nuclear EGFP
Intersectional cytoplasmic "AND/OR" tdTomato+EGFP
Intersectional cytoplasmic "AND/OR" tdTomato only
Intersectional nuclear "AND/OR" dTomato+EGFP
Intersectional nuclear "AND/OR" dTomato only
Intersectional nuclear "AND/OR" EGFP only
Double-labeled cell
Nuclear GFP signal
Nuclear dTomato signal
Cytoplasmic GFP signal
Population labeled by virus only : Virus +; TgReporter -
Boundary of double label : Virus+; TgReporter+
E
Cplx3-P2A-FlpO;Ai65F
553840
Cplx3 - mRNA in situ hybridization
Cplx3-P2A-FlpO;Ai193-hyg
AiE2359m
668814
AiE2359m Virus
Cplx3-P2A-FlpO;Ai193-hyg
F
Chodl-P2A-Cre;Sst-IRES-FlpO;Ai224-hyg
556654
Sst-IRES-Cre;Slc32a1-T2A-FlpO;Ai224-hyg
547658

Figure S8. Factors influencing tool expression and evaluation. A. scRNA-seq (SSv4) data showing distribution of labeled cells from tools **1-8** mapped to mouse VISp taxonomy and displayed at the cluster level. **B.** Select STPT images for tools **1-6**, and additional related tools. **C.** Representative *Npnt* mRNA *in situ* hybridization and STPT images of *Npnt-P2A-FlpO* with two different reporters showing labeling of cells in L5, whereas the SSv4 data for the cross to *Ai193* (tool **7** in A) do not show L5 cells. This could be due to L5_PT cells not surviving FACS for this experiment. **D**. Representative STPT data for *Chodl-P2A-Cre; Sst-IRES-FlpO* crossed with previously characterized reporters (*Ai14* and *Ai65F*) and the new AND/OR reporters (*Ai193* and *Ai224*) both independently and as a triple transgenic. **E**. Representative STPT images showing *Cplx3-P2A-FlpO* with different reporters and *Cplx3* mRNA expression (blue brackets) by RNA *in situ* hybridization (https://mouse.brain-map.org/experiment/show/70928340). The expression pattern for the enhancer AAV, AiE2359m, mirrors *Cplx3* expression (blue brackets) by RNA *in situ* hybridization, whereas the expression of the transgenic line, *Cplx3-P2A-FlpO;Ai193* does not include *Cplx3*+ cells in the entorhinal area. **F.** Expression of nls-EGFP (Cre-dependent) and nls-dTomato (Flp-dependent) is faithful in the *Ai224* reporter line; however, nuclear localization is imperfect. The GFP appears mostly nuclear, but weak signal can be observed in the cytoplasm (light blue arrow) and processes (blue arrow). In comparison, nls-dTomato appears nuclear (white arrow). Scale bars: 1.0 and 0.2 mm for full section and expanded view; 0.1 mm for further expanded view in (C).

Supplemental tables.

Table S1. Summary of labeling patterns produced by transgenic mouse lines, *related to Figures 7 and 8*.

Table S2. Transgenic mouse lines used in the manuscript, *related to STAR methods*.

Table S3. AAV plasmids used in the manuscript, *related to STAR methods*.

Table S4. Unique enhancers evaluated in this manuscript, *related to STAR methods*.

Table S5. “Hall-of-fame” enhancer AAVs, *related to Figures 3-6*.