Supplemental information: Extended methods

Animals

All animal practices and procedures were approved by the Washington University in St. Louis Institutional Animal Care and Use Committee (IACUC). Transgenic mouse strains used in this study include Synapsin 1 (Syn1)::Cre (RRID:IMSR_JAX:003966) and glial fibrillary acidic protein (GFAP)::Cre (RRID:IMSR_JAX:024098). All mice were bred to the C57BL/6J background, with the exception of animals used for adult stereotactic injections, which were wild-type animals of the FVB/N6 background (RRID:IMSR_JAX:001800). At indicated endpoints, mice were anesthetized with isoflurane and perfused with 15ml of cold saline (PBS) prior to tissue collection. Animals in the BRD4 Syn1::Cre, BRD4 GFAP::Cre, and P28 SP1 Syn1::Cre cohorts received pentylenetrazole-induced seizures immediately prior to sacrifice. Unless otherwise noted, brains were either dissected and flash frozen in liquid nitrogen (for molecular analyses) or fixed in 4% paraformaldehyde 24-48 hours, exchanged into 30% sucrose, and either directly frozen at -80°C or cryoprotected in O.C.T. (for immunofluorescence).

To collect newborn litters, pregnant females were monitored daily until giving birth. Newborn pups were injected within 24 hours of being found in the cage, i.e. 0-48 hours after birth; thus, these injections are labeled “P0-1.” One exception is AAV::hypPB Frontflip in **Fig 6**, for which 1/3 of both the GFAP::Cre(+) and GFAP::Cre(-) animals were P3-4 when injected.

Flow cytometry analysis

For evaluation of SRT-TdTomato fluorescence in **Fig 6B**, HEK293T cells were collected 96-hours post-transfection for flow cytometry analysis. The cells were lifted using 0.25% trypsin-EDTA (ThermoFisher 25200056) and washed once with PBS. Cell pellets were resuspended at $1x10^6$ cells/ml in PBS/0.5% BSA/2mM EDTA with 0.1µg/ml DAPI (Sigma-Aldrich D9542, St Louis, MO, USA) for 10 min for dead cell and debris exclusion from analysis. Cells were washed once more with PBS/0.5% BSA/2mM EDTA and analyzed on a CytoFLEX S (Beckman Coulter), with each sample being run at a low flow rate (10-30µl/min) for 3 min to collect approximately 20,000 events per condition. A violet 405nm laser with 450/45 bandpass filter was used to detect DAPI and a yellow-green 561nm laser with 585/42 bandpass filter was used to collect tdTomato fluorescence. Flow cytometry data was analyzed with CytExpert software (Beckman Coulter). The gating was performed as outlined in SI Appendix **Fig 8SA**.

Immunofluorescence and imaging

10µm-thick or 40µm-thick fixed-frozen sagittal or coronal brain sections were washed with PBS and permeabilized with 0.1% Triton X-100 (Sigma-Aldrich, St. Louis, MO, USA). Non-specific binding was blocked with 5% normal donkey (Jackson ImmunoResearch, West Grove, PA, USA) or goat (Vector Laboratories, Burlingame, CA, USA) serum for 30-60 minutes at room temperature. After blocking, slides were exposed to primary antibody overnight at 4°C, washed three times with PBS, and then incubated with secondary

antibodies for 1 hour at room temperature. Nuclei were counterstained with DAPI (Sigma-Aldrich, St. Louis, MO, USA) and coverslips were applied with ProLong Gold Antifade (ThermoFisher, Waltham, MA, USA) or Fluoromount-G (SouthernBiotech, Birmingham, AL, USA) mounting media. Immunofluorescent images of brain sections were acquired with a Nikon A1Rsi or Zeiss LSM 700 confocal microscope and imported into ImageJ (v. 1.51s) for manual cell counts and quantification. For analyses of hypPB expression in various cell types, 5 mice were used, and co-localization was quantified in 2 cortical images from a single section per animal. Antibodies used for immunostaining included chicken anti-GFP (Aves Labs GFP-1020) at 1:1000 dilution, mouse rabbit anti-RFP at 1:400 or 1:500 dilution (Rockland 600-401-379), anti-NeuN at 1:100 dilution (Millipore-Sigma MAB377), rabbit anti-cMyc at 1:250 dilution (Sigma C3956), and goat anti-GFAP at 1:500 dilution (Abcam ab53554).

Cells transfected with AAV::hypPB FLEX for testing Cre-dependence were live imaged for TdTomato on a Leica DMI 3000B tissue culture microscope. All images were acquired with equal conditions and exposure times for direct comparison.

In situ hybridization and imaging

10μM-thick, 4% paraformaldehyde fixed-frozen sections were cut and slide-mounted. mRNA encoding for *hyperPiggyBac* (VF1-20268-01) was detected using a custom probe-set designed by Affymetrix (now ThermoFisher) using the Affymetrix ViewRNA ISH Tissue 1-Plex kit (ThermoFisher, QVT0050) and chromogenic signal amplification kit (ThermoFisher VT0200) with the following modifications: Slides were immersed in 4% paraformaldehyde overnight at 4°C prior to *in situ* hybridization, then the baking, deparaffinization, and heat pretreatment steps were omitted (steps 1-3, 5) because sections were not embedded in paraffin. Slides were hybridized either with anti-hypPB or no-probe controls. Following the *in situ* labeling protocol, sections were labeled for 5 minutes with DAPI (1:20,000, Sigma D9542), washed with PBS, and a drop of prolong gold (ThermoFisher P36934) was added while applying the coverslip. Slides were then imaged at 20x magnification on a Zeiss LSM 700 confocal microscope using a Cy3 filterset to detect FastRed fluorescence.

Analysis of brain degeneration following viral injection

Mice were exposed to either the calling cards viruses (AAV::hypPB + AAV::SRT) or RFP-only virus (as a control) via P0-1 injection and sacrificed 28 days later for silver staining (a marker for cells irreversibly committed to cell death). Briefly, animals were heavily sedated and perfused with TRIS fixative in 4% paraformaldehyde followed by vibratome sectioning of brains at 75μM in the coronal plane. Every eighth section across the rostrocaudal extent of the brain was then silver stained as described previously[1]. The number of degenerating neurons was quantified for each animal by a rater blind to treatment who counted the total number of silver positive neurons in the dorsal cortex of each section.

Mouse behavior, developmental milestones, and sensorimotor battery

Mouse behavior and development was monitored and compared between animals injected calling cards AAV reagents (AAV::hypPB and AAV::SRT) or AAV::RFP only. In addition to weight, which was measured at P8, P14, P25, tests were administered to assess attainment of developmental milestones (P14 righting from back), anxiety related behavior (P25 1 hour behavior, recording time spent in the edge or center of cage), and balance/strength/coordination (P25/26 sensorimotor battery). Procedures were done as previously described[2,3], with two trials per animal. A break was allowed after the completion of the first set of test trials to avoid exhaustion effects, and the test order was reversed for the second trial for all animals. Walking initiation, ledge performance, platform performance, and pole performance were all administered at P25, while 60° inclined screen test, 90° inclined screen test, and inverted screen test were administered on P26. Prior to testing, mice were given a routine health check, and from this, two animals were excluded; one for runtiness and one for severe hydrocephaly likely derived from needle stick. Further, one litter of RFP-only animals was not able to complete all timepoints (cage flooding) and was thus also excluded. In total, 21 animals were included in the calling cards group (11M/10F) and 24 in the RFP-only group (10M/14F), all of which were used for downstream analyses. Test administrators were blinded from treatment group identity during all testing.

SRT and BrokenHeart library preparation, sequencing, and mapping

SRT libraries were prepared from cortex RNA samples. Prior to library preparation, cortex samples were dissected into 10 separate pieces, from which RNA was independently isolated with the manufacturer's protocol (Qiagen RNeasy kit, Germantown, MD, USA). This allows for identification of up to 10 independent insertion events into any TTAA site, given that these insertions occur in spatially separate samples. From these RNA samples, transposon sequencing libraries were generated with our bulk SRT protocol[4]. In brief, RNA samples were first reverse transcribed, from which self-reporting transcripts, including flanking genomic sequences, were amplified via PCR. These amplicons were then tagged with universal Illumina sequencing overhangs, with separate indexes for libraries from each dissected piece, allowing for 10 'barcodes' per sample and sequenced on Illumina HiSeq 2500, NextSeq 500, or MiniSeq platforms (Illumina, San Diego, CA, USA).

BrokenHeart libraries were prepared from cortical DNA samples, as previously described[5]. A set of 20 individually barcoded BrokenHeart transposons were pooled and packaged into AAV. Thus, independent insertions into the same TTAA can be uniquely identified via barcode, removing the need to dissect and process tissue samples in separate pools as with SRT. Extracted DNA was self-ligated, amplified with inverse PCR, and sequenced with the Illumina NextSeq 500 platform (Illumina, San Diego, CA, USA).

Sequencing reads obtained from SRT and BrokenHeart libraries were stringently filtered for features of true insertion events (presence of *piggyBac* terminal repeat sequence; intact sequencing adapters, barcodes, and indexes; and a TTAA site) and mapped to build version mm10 of the mouse genome with Novoalign 3 (Novocraft Technologies) (for SRT) or Bowtie2[6] (for BrokenHeart). Reads aligning to the same TTAA with

separate barcodes were considered unique insertions, and all analyses in this report considered all unique insertions equally, independent of read depth. We used BEDtools intersect to count the number of insertions directed to introns, exons, 3'- and 5'-UTRs, and intergenic regions, using annotations from the HOMER package[7] (See SI Appendix, **Fig 2SE**).

Significant insertion peak calling

Significantly enriched insertion peaks were identified via a count-based statistical comparison as previously described[4]. In brief, this pipeline first segments the genome into blocks of constant insertion density. For each block, it calculates the p-value of insertion enrichment relative to a background model assuming uniformly distributed insertions. A user-defined significance threshold is defined, and all blocks surpassing this threshold are considered "significantly enriched insertion peaks". This "background-free" method for unbiased identification of all significantly enriched genomic regions in a single experimental sample, used here in **Fig 2**, is expected to identify all BRD4-bound regions within the parameters of the calling cards system.

Alternatively, we can define differentially-bound regions between two experimental samples, as was done in **Fig 3, Fig 6,** SI Appendix **Fig5S**, and SI Appendix **Fig 6S** for astrocyte- or neuron-specific BRD4 peaks (using the combined insertion pools from three GFAP::Cre [n = 3,072,163 insertions] and three Syn1::Cre animals [n = 2,484,133 insertions]) and **Fig 4** for SP1 peaks over unfused hypPB. In this analysis, the pipeline again segments the genome into blocks, but then assigns a p-value to each block based on the differential enrichment between the two samples. As with the background-free pipeline, a user defined p-value threshold is chosen, below which all blocks are considered significantly enriched. Because of the natural BRD4-bias of hypPB, using this differential peak caller allows for identification of binding sites of specific TFs by selecting genomic locations with re-directed insertion density in a TF-hypPB profile compared to an unfused hypPB background.

Defining enhancers and super enhancers

Since H3K27ac is a known marker of active enhancers[8,9] and super enhancers[10,11], we utilized published P14 mouse cortex[12] and N2a[13] H3K27ac ChIP-Seq datasets to define cortical and N2a enhancers super enhancers, respectively. As previously described[10,11], we used the rank ordering of super enhancers (ROSE) v0.1 pipeline and the model-based analysis for ChIP-Seq (MACS) v1.4.1 peak finding algorithm[14] with a p-value enrichment threshold of 10^{-9} to define enhancers and super enhancers. We then used the BEDtools suite[15] to compare the coincidence of enhancers and super enhancers with our unfused hypPB calling cards insertion peaks (**Fig 2**; See SI Appendix, **Fig 4SC-D**).

Additionally, for qualitative measures of histone modification enrichment at calling cards BRD4 insertion peaks (**Fig 2**), we used publicly available P0 mouse forebrain ChIP-Seq datasets from ENCODE[16]; specifically, H3K27ac (ENCSR094TTT), H3K4me1 (ENCSR465PLB), and H3K27me3 (ENCSR070MOK).

Super enhancer in vitro sensitivity and specificity

Sensitivity and specificity of calling cards peaks were assessed for super enhancer identification in N2a cells that were either transfected or electroporated with AAV::hypPB and AAV::SRT plasmids (See SI Appendix, **Fig 4S**). RNA was separately isolated from a total of 33 wells (i.e. barcodes) from 6-well plates, and SRTs were sequenced, generating 806,653 unique insertions, though of note, the majority (651,631) were derived from 12 barcodes that received plasmids via electroporation. 800,000 unique insertions were randomly selected from the total pool of 806,653 insertions, from which significantly enriched peaks were defined using our background-free peak calling method at a range of significance thresholds. These peaks were intersected (with BEDtools intersect) with known N2a super enhancers defined via a previously published N2a H3K27ac ChIP-seq dataset[13], and sensitivity was defined as the percentage of peaks intersecting super enhancers for each peak calling significance threshold. To then define specificity, we identified the "true negative" space of the genome, and assessed the percentage of true negative peaks intersected by calling cards peaks. To do this, we first identified any possible active enhancer region of the genome with MACS peak finding using a low-stringency significance threshold of 10^{-1} and subtracted these peaks from the mouse genome, creating a "true negative" genome. We then sampled peaks (with BEDtools shuffle) within this true negative genome of the same size distributions as the list of active super enhancers until we collected an average of 1X coverage across the genome. With a true negative space of 2,616,503,093 basepairs and a total super enhancer size of 23,143,876 basepairs, this required 114 random samplings, resulting in 85,158 true negative peaks. Finally, we intersected our calling cards peaks with these true negative peaks, and specificity was defined as the percentage of true negative peaks not intersected by a calling cards peak. Of note, we expect that unfused hypPB is driven to super enhancers via interaction with BRD4; thus, sensitivity and specificity measurements may be higher if compared to BRD4 occupancy rather than H3K27ac.

Analysis of enhancer- and promoter-associated gene expression

Gene expression has been shown to be preferentially regulated by proximal enhancer elements[8,9,17]. Thus, since a cell type-specific mapping of enhancers to the genes they regulate is not available, we used proximity as an imperfect[18] albeit widely used[19] proxy. In our analyses of cell type-specific expression of genes near cell type-enriched BRD4 calling cards peaks (**Fig 3, Fig 6**, and SI Appendix **Fig 5S**), we first defined the nearest gene (or genes, if multiple intersected a calling cards peak) to each significant calling cards peak. These gene sets were then filtered and the remaining genes were used for subsequent analyses. Gene sets were filtered as follows: 1) genes greater than 10,000 bases away from a differential insertion peak were removed, to eliminate low confidence gene-enhancer pairs, 2) genes near or overlapping multiple insertion peaks counted once, and 3) genes for which cell type-specific RNA expression data were unavailable in our comparison dataset were removed.

Unbiased cell type identification was completed with the Cell-type Specific Expression Analysis (CSEA) tool[20] (http://genetics.wustl.edu/jdlab/csea-tool-2/) using candidate gene sets near either GFAP::Cre enriched or Syn1::Cre enriched insertion peaks. For each set, we analyzed genes near the most enriched peaks for

each cell type. For GFAP::Cre, this included 131 genes ($p<10^{-21}$ for associated insertion peaks), of which 114 were present in CSEA reference sets and used for analysis. For Syn1:Cre, this included 123 genes ($p<10^{-11}$), with 110 present in reference sets.

For comparison of SP1 binding and gene expression in **Fig 5A-B**, we utilized the mm10_knownCanonical gene set and mm10_TSS coordinates from the UCSC genes table. We defined promoter-proximal regions as +/-1000 bases from the TSS. We first filtered mm10_knownCanonical gene set to remove duplicates (<3% of total genes) and then intersected gene coordinates with promoter proximal regions. After manually filtering to assign true promoters to each transcript (i.e. immediately upstream from TSS), we generated a list of unique promoter/gene combinations (24,528 unique genes) and compared insertion density and gene expression at these coordinates.

For comparison of P28 and P10 SP1 promoter insertions to RNA expression in **Fig 5C-G**, we utilized a previously published RNA-seq dataset[21] with RNA expression data available for week 1 (Wk1) and week 4 (Wk4), which correspond to ~P7 and ~P28, respectively. Before assessing P28 or P10 insertion density at promoters, insertion profiles were downsampled such that each cohort had exactly 240,000 insertions per library (80,000 per mouse for P10, 3 mice; 60,000 per mouse for P28, 4 mice); thus, insertion totals could be directly compared without any normalization to library size. Further, this downsampling procedure eliminates the possibility that any given observed increase in insertion density at P28 was due to an overall increase in insertion total over time. We then calculated number of insertions at each unique promoter (using the list of unique promoter/gene combinations generated above) and removed any gene with no insertions at either timepoint (19,046 / 24,528 unique genes remaining). A pseudocount of 1 was added to promoter insertion totals for each gene at each timepoint prior to analysis. To eliminate noise due to low RNA expression and/or random low-frequency insertion events, we next removed any gene with <6 insertions combined between the P28 and P10 datasets (including the 2 pseudocounts) or <1 FPKM combined between Wk1 and Wk4 RNA-seq expression, leaving a final total of 4991 unique gene/promoter combinations which were used in subsequent analyses. This list of 4991 genes were divided into three categories, based on their RNA expression at Wk1 and Wk4: (1) early genes; log(Wk4/Wk1 FPKM) < -0.5, (2) constitutive genes; -0.5 < log(Wk4/Wk1 FPKM) < 0.5, and (3) late genes; log(Wk4/Wk1 FPKM > 0.5. Within these categories, SP1 occupancy was compared between the P28 and P10 cohorts.

Validation of astrocyte enhancer candidates with PALE[22]

Candidate astrocyte-enriched enhancers were selected from the list of GFAP::Cre-enriched insertion peaks in **Fig 3** based on three criteria, which select for high-confidence astrocyte enhancers: 1) enhancer size (smaller than 3kB, for cloning purposes), 2) significant GFAP::Cre enrichment (over Syn1::Cre), and 3) astrocyte-specific RNA expression of their nearest genes. These candidates were PCR amplified with primers listed in the table below and with an MluI overhang adapter (TGTAGGACGCGT) on either end, and cloned into the miniP-dsRed plasmid with MluI, upstream of the hsp68 minimal promoter. As a positive control, the canonical GFAP promoter[23] was also cloned into this plasmid, in the same location.

To test efficacy of the candidates for enhancing dsRed expression, each plasmid was electroporated into lateral ventricle-proximal cells, along with a separate plasmid containing CFP driven by the canonical GFAP promoter. P0-1 pups were placed on a wet towel on wet ice for 10 minutes to anesthetize. Then plasmids were delivered to the lateral ventricle via intraventricular injection along with Fast Green FCF dye (Sigma, 2353-45-9), with coordinates approximately equidistant from the lambdoid suture and the eye, 2mm lateral to the sagittal suture, and 2mm depth to ensure lateral ventricle penetration. 1μl of DNA was delivered into one hemisphere for each mouse, at a concentration of 1μg/μl for GFAP::CFP and 0.5μg/μl for dsRed plasmids. Electroporation was induced for 5 square pulses, 50ms per pulse at 100V and 950ms inter-pulse intervals, sweeping the electrodes from the dorsal to lateral using ~25° angle intervals.

Pups were sacrificed at P7 or P21 and brains were collected and analyzed with immunohistochemistry for CFP and dsRed (CFP and dsRed stained with GFP and RFP antibodies, respectively, as indicated in methods). Individual astrocytes were imaged for intensity analysis in **Fig 3E-F**, SI Appendix **Fig 6SF,H**, and SI Appendix **Fig 7SA,C** using equivalent exposure settings. A region of interest (ROI) was defined around the astrocyte based on CFP only, and dsRed and CFP fluorescence was quantified within the ROI using ImageJ (v. 1.51s). This was done from 3 brains (34-42 cells) per condition at P7 and from 2 brains (28-32 cells) per condition at P28 (except pGFAP, which only had 1 brain and 17 cells used in quantification at P28, as this was a positive control sample). The ratio of CFP:dsRed was calculated for each cell and averaged and compared across conditions for final assessment of dsRed enhancement. For P7 co-localization quantifications in SI Appendix **Fig 6SG**, dsRed(+) cells were manually counted from 3 brains (150 total cells) per condition, while P21 quantifications in SI Appendix **Fig 7SB** were taken from 2 brains (287-769 total cells) per condition (again, the positive control, pGFAP had 1 brain, 67 total cells used in quantifications).

Supplemental information: References

1. Noguchi, K. K., Nemmers, B. & Farber, N. B. Age has a similar influence on the susceptibility to NMDA antagonist-induced neurodegeneration in most brain regions. *Dev. Brain Res.* **158**, 82–91 (2005).
2. Wozniak, D. F. *et al.* Apoptotic neurodegeneration induced by ethanol in neonatal mice is associated with profound learning/memory deficits in juveniles followed by progressive functional recovery in adults. *Neurobiol. Dis.* **17**, 403–414 (2004).
3. Grady, R. M., Wozniak, D. F., Ohlemiller, K. K. & Sanes, J. R. Cerebellar Synaptic Defects and Abnormal Motor Behavior in Mice Lacking - and beta-Dystrobrevin. *J. Neurosci.* **26**, 2841–2851 (2006).
4. Moudgil, A. *et al.* Self-reporting transposons enable simultaneous readout of gene expression and transcription factor binding in single cells. bioRxiv:10.1101/538553 (1 February 2019).
5. Wang, H., Mayhew, D., Chen, X., Johnston, M. & Mitra, R. D. 'Calling Cards' for DNA-binding Proteins in Mammalian Cells. *Genetics* **190**, 941–949 (2012).
6. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, (2009).
7. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).
8. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107**, 21931–21936 (2010).
9. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
10. Whyte, W. A. *et al.* Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* **153**, 307–319 (2013).
11. Hnisz, D. *et al.* Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155**, 934–947 (2013).
12. Stroud, H. *et al.* Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States. *Cell* **171**, 1151-1164.e16 (2017).
13. Kwon, D. Y., Zhao, Y. T., Lamonica, J. M. & Zhou, Z. Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat. Commun.* **8**, 1–8 (2017).
14. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, (2008).
15. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
16. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
17. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
18. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associate Variation in Regulatory DNA. *Science.* **337**, 1190–1195 (2012).
19. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**,

43–49 (2011).

20. Xu, X., Wells, A. B., O'Brien, D. R., Nehorai, A. & Dougherty, J. D. Cell Type-Specific Expression Analysis to Identify Putative Cellular Mechanisms for Neurogenetic Disorders. *J. Neurosci.* **34**, 1420–1431 (2014).
21. Lister, R. *et al.* Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science.* **341**, 1237905–1:12 (2013).
22. Stogsdill, J. A. *et al.* Astrocytic neuroligins control astrocyte morphogenesis and synaptogenesis. *Nature* **551**, 192–197 (2017).
23. Sakers, K. *et al.* Astrocytes locally translate transcripts in their peripheral processes. *Proc. Natl. Acad. Sci.* **114**, E3830–E3838 (2017).

Supplemental information: Figure legends

Figure 1S. AAV9 intracranial injection induces hypPB expression in the cortex.
(A) *In situ* hybridization of hypPB RNA and no probe control displaying widespread cortical expression of hypPB after delivery of AAV::hypPB.

Figure 2S. Comparison of neonatal and adult AAV delivery of traditional DNA calling cards and SRT calling cards systems.
(A) Viral donor transposon constructs for DNA (AAV::BrokenHeart) and RNA (Self-Reporting Transposon; SRT) calling cards. (B) DNA calling cards library preparations were carried out as previously described in Wang et al., *Genetics*, 2012. TF-hypPB fusions insert BrokenHeart transposon DNA near TF binding sites. Genomic DNA is then harvested and digested with restriction enzymes that cut near the end of the transposon and in downstream genomic sequence. These fragments are subsequently self-ligated and circularized. From these, transposons and their flanking genomic sequences are amplified with inverse PCR, using primers that contain Illumina sequencing primers and adapters. Final products are sequenced and aligned to the mouse genome to map transposons to genomic locations. (C) Schematic of RNA calling cards library preparation protocol. SRTs contain a ubiquitous promoter that drives the transcription of a TdTomato reporter gene with no 3' poly-A termination signal. When the SRT gene is transcribed episomally in the context of the AAV genome (i.e. prior to transposition), a downstream self-cleaving ribozyme is incorporated into the transcript, destabilizing it and ultimately causing its degradation. However, once the transposon is inserted into the genome, the SRT is no longer affiliated with the ribozyme, and the SRT transcript becomes stabilized and extends to include flanking genomic sequence downstream of the TdTomato gene. RNA is then reverse transcribed and PCR amplified, and Illumina adapters are added for sequencing. (D) Read mapping rates for BrokenHeart DNA calling cards and SRT RNA calling cards libraries after P0-1 or adult cortical delivery, showing increased efficiency of recovery for genome-mapping reads in SRT relative to BrokenHeart. (n=3 mice for P0-1 BrokenHeart [12.3% mapping to mm10]; n=2 mice for P0-1 SRT [80.7%]; n=3 mice for adult SRT [76.5%]) (E) Number of unique insertions observed from BrokenHeart DNA calling cards and SRT RNA calling cards libraries (n=3 mice for P0-1 BrokenHeart [198,981 insertions were recovered at mean read coverage of 8.98 reads/insertion]; n=2 mice for P0-1 SRT [3,732,694 insertions at 3.84 reads/insertion]; n=3 mice for adult SRT [4,114,106 insertions at 5.92 reads/insertion]), displaying greatly increased recovery with SRT. (F) Mapping locations of transposon insertions in various genomic regions from BrokenHeart and SRT calling cards libraries.

Figure 3S. AAV SRT calling cards system does not induce excess degeneration or behavioral/developmental deficits.
(A-I) Behavioral and developmental assessments of mice injected at P0-1 with SRT calling cards (n=21) or control, RFP only (n=24) viruses revealed few or no developmental, sensorimotor, or anxiety-related deficits in

calling cards animals relative to control. All group comparisons were done with two-tailed, unpaired Student's t-test, with Bonferroni corrected α=0.05 as a significance threshold (including all tests in Fig 1). **p<0.01 (Bonferroni corrected).

Figure 4S. In vitro and in vivo sensitivity of unfused hypPB calling cards libraries for active REs.
(A-B) Sensitivity and specificity of super enhancer (SE) identification for unfused hypPB calling cards libraries in N2a cells. (A) Receiver-operator characteristic (ROC) curve for identification of active super enhancers using unfused hypPB peaks called from 800,000 unique insertions. Area under ROC curve: 0.82. (B) Super enhancer sensitivity at various significance thresholds and insertion totals, demonstrating high super enhancer sensitivity at even very low (10^4) insertion totals. (C-D) Intersections between significantly-enriched insertion peaks derived from *in vivo*, cortical unfused hypPB calling cards libraries and H3K27ac-marked (C) enhancers or (D) super enhancers at a range of insertion peak significance thresholds. Red line represents total number of significant peaks at each p-value threshold. Quantifications in Fig. 2F-I represent intersections at p=10^{-30} significance threshold.

Figure 5S. Supplemental to: FLEX calling cards system generates cell type-specific RE profiles.
(A) Transfection of BrokenHeart transposons, hypPB FLEX plasmid, and Cre recombinase into HEK293T cells. n=3 wells per condition, 5 image fields per well, representative images shown. No TdTomato reporter from reconstituted BrokenHeart transposons observed in the absence of Cre or hypPB FLEX at 24-hours post-transfection. (B) Representative images of Syn1::Cre and GFAP::Cre positive and negative littermate brains, displaying increased SRT-derived TdTomato reporter signal in Syn1::Cre positive animals. (C) Representative images displaying preferential expression in Neu(+) neurons in CA3 hippocampal regions of Syn1:Cre(+) mice, but not negative littermates. TdTomato images taken at equal exposure times for direct comparison. (D) Quantifications of unique insertions in Syn1::Cre or GFAP::Cre positive and negative littermate mice (two-tailed, unpaired Student's t-test: * p<0.05, ** p<0.01). Syn1::Cre (+) vs (-): p=0.0032, t=3.86, df=10, 95% $C.I._{diff}$ [268,014 : 1,000,132]; GFAP::Cre (+) vs (-): p=0.015, t=2.92, df=10, 95% $C.I._{diff}$ [103,883 : 771,416] (E) Number of genes near differentially enriched insertion peaks in Syn1::Cre and GFAP::Cre animals at a range of significance thresholds. (F) Normalized neuron-to-astrocyte expression ratio [Neuron FKPM/(Neuron FPKM + Astrocyte FKPM] for genes near Syn1::Cre or GFAP::Cre enriched peaks at a range of significance thresholds for defining differentially enriched insertion peaks. As significance threshold becomes more stringent, expression of nearby genes becomes more cell type-specific. RNA-seq expression from Zhang et al., *Journal of Neuroscience*, 2014. (G) Astrocyte versus neuron expression of all genes near Syn1::Cre or GFAP::Cre enriched insertion peaks at a stringent peak-calling significance threshold (p=10^{-7}). Percent of genes on either side of the y=x midline shown. (H-I) Graphical representation of cortical cell type enrichment based on gene sets near either (H) Syn1::Cre (p<10^{-11}; top 123 genes) or (I) GFAP::Cre (p<10^{-21}; top 131 genes) enriched insertion peaks. Legend displays Benjamini-Hochberg corrected Fisher's Exact Test p-value

for overlap of reference cell type-specific gene sets and Syn1::Cre or GFAP::Cre candidate gene sets. Stringencies for enrichment for each pre-defined reference set are represented by size of each hexagon, with the outer ring being the least stringent set and inner ring being the most stringent set.

Figure 6S. Functional validation of enhancer activity at P7 for GFAP::Cre enriched unfused hypPB peaks.
(A-D) Candidate GFAP::Cre enriched insertion peaks that were chosen for functional enhancer validation based on enhancer size, significant GFAP::Cre enrichment, and astrocyte-specific RNA expression of their nearest genes. Each candidate RE (highlighted in blue) was separately cloned into a plasmid upstream of the hsp68 minimal promoter and a dsRed reporter gene for *in vivo* testing. (E) Chromosomal coordinates and lengths of candidate REs. (F) Candidate RE reporter constructs were co-delivered along with a GFAP::CFP plasmid to ventricle-proximal radial glia, including astrocytes, via PALE[22]. Expression of dsRed was enhanced by both the canonical GFAP promoter (pGFAP; positive control) and three of the four candidate REs (all but eMms22l). (G) Percentage of dsRed(+) cells in cortex co-labeled with GFAP::CFP, demonstrating that >96% of labeled cells are glia at this timepoint in all conditions. n=150 dsRed(+) cells from 3 brains per condition. (H) Quantification of dsRed expression enhancement in GFAP::CFP(+) astrocytes by pGFAP and candidate RE constructs. n=34-42 GFAP::CFP(+) cells from 3 brains per condition (one-way ANOVA with Dunnett's multiple comparisons test; *p<0.05; **p<0.01; ****p<0.0001). $pGFAP_{mean\ diff.}$ = 0.66, 95% $C.I._{diff}$ [0.01 : 1.33]. $eRasa2_{mean\ diff.}$ = 0.67, 95% $C.I._{diff}$ [0.03 : 1.32]. $eTaf4b_{mean\ diff.}$ = 0.84, 95% $C.I._{diff}$ [0.21 : 1.48]. $ePla2g7_{mean\ diff.}$ = 1.22, 95% $C.I._{diff}$ [0.58 : 1.86]. $eMms22l_{mean\ diff.}$ = 0.002, 95% $C.I._{diff}$ [-0.65 : 0.65].

Figure 7S. Candidate astrocyte enhancers identified with FLEX calling cards direct cell type-specific expression in astrocytes by P21.
(A-B) Candidate astrocyte enhancers derived from FLEX calling cards enrichment (see Fig 6S) were electroporated into P0-1 mouse pups via PALE[22] and animals were sacrificed at P21 for IF analysis. In animals receiving a plasmid containing dsRed driven by a minimal promoter only (hsp68::dsRed), dsRed expression in the cortex was evident in a population of NeuN(+) neurons (white arrows) that was not observed at P7 (see Fig 6SF-G). In contrast, in animals receiving either the pGFAP-driven positive control plasmid or plasmids containing the candidate enhancers, dsRed expression was significantly limited to GFAP::CFP(+) cells, indicating that the enhancer candidates facilitate cell type-specificity of gene expression. n=287-769 dsRed(+) cells from two brains per condition for RE candidates and n=67 dsRed(+) cells from one brain for pGFAP (one-way ANOVA with Dunnett's multiple comparisons test; ****p<0.0001). $pGFAP_{mean\ diff.}$ = 74.47%, 95% $C.I._{diff}$ [69.66% : 79.28%]. $eRasa2_{mean\ diff.}$ = 73.57%, 95% $C.I._{diff}$ [68.76% : 78.38%]. $eTaf4b_{mean\ diff.}$ = 74.47%, 95% $C.I._{diff}$ [69.66% : 79.28%]. $ePla2g7_{mean\ diff.}$ = 72.37%, 95% $C.I._{diff}$ [67.56% : 77.18%]. $eMms22l_{mean\ diff.}$ = 74.47%, 95% $C.I._{diff}$ [69.66% : 79.28%]. (C) Quantification of dsRed expression enhancement in GFAP::CFP(+) astrocytes by pGFAP and candidate RE constructs. n=28-32 GFAP::CFP(+) cells from 2 brains per condition for RE candidates and n=17 GFAP::CFP(+) cells from one brain for pGFAP (one-way ANOVA with Dunnett's multiple comparisons test; *p<0.05). $pGFAP_{mean\ diff.}$ = 0.59, 95% $C.I._{diff}$ [-0.03 : 1.21]. $eRasa2_{mean\ diff.}$ = 0.58, 95%

$C.I._{diff}$ [0.04 : 1.13]. $eTaf4b_{mean\ diff.}$ = 0.42, 95% $C.I._{diff}$ [-0.11 : 0.95]. $ePla2g7_{mean\ diff.}$ = 0.40, 95% $C.I._{diff}$ [-0.15 : 0.95]. $eMms22l_{mean\ diff.}$ = 0.62, 95% $C.I._{diff}$ [0.10 : 1.15].

Figure 8S. Gating procedure for flow cytometry in cells with BrokenHeart-derived TdTomato fluorescence. (A) Cells were dissociated into suspension and spiked with 0.1μg/ml DAPI as a viability dye, allowing discrimination of intact from plasma membrane-compromised cells. Gates were drawn based on Cre(-) conditions. The analyzed TdTomato(+) populations were selected from the parent cells/singlets/DAPI(-) subsets.

Figure 1S

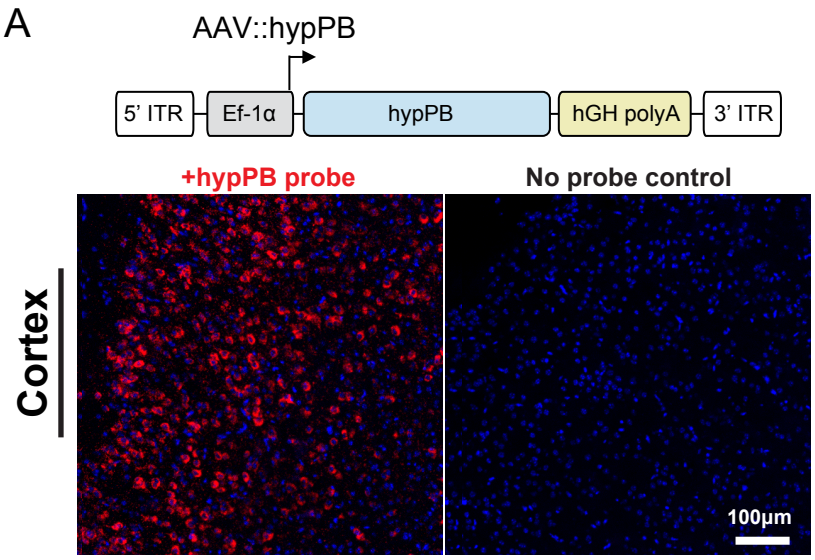

A
AAV::hypPB
5' ITR
Ef-1α
hypPB
hGH polyA
3' ITR
+hypPB probe
No probe control
Cortex
100μm

Figure 2S

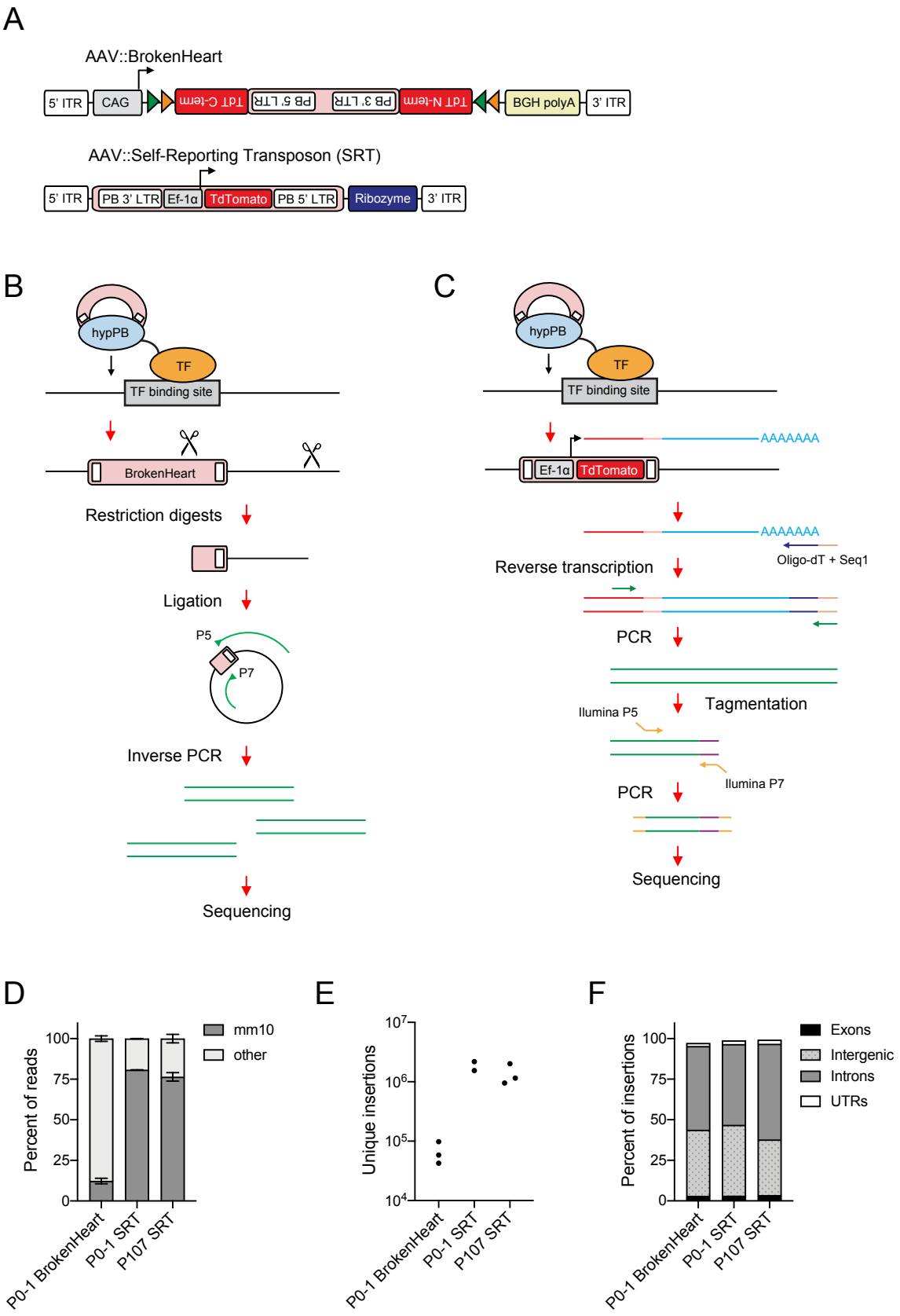

A
AAV::BrokenHeart
5' ITR
CAG
TdT C-term
PB 5' LTR
PB 3' LTR
TdT N-term
BGH polyA
3' ITR
AAV::Self-Reporting Transposon (SRT)
5' ITR
PB 3' LTR
Ef-1α
TdTomato
PB 5' LTR
Ribozyme
3' ITR
B
hypPB
TF
TF binding site
BrokenHeart
Restriction digests
Ligation
P5
P7
Inverse PCR
Sequencing
C
hypPB
TF
TF binding site
AAAAAAA
Ef-1α
TdTomato
AAAAAAA
Oligo-dT + Seq1
Reverse transcription
PCR
Tagmentation
Ilumina P5
Ilumina P7
PCR
Sequencing
D
Percent of reads
100
75
50
25
0
mm10
other
P0-1 BrokenHeart
P0-1 SRT
P107 SRT
E
Unique insertions
10^7
10^6
10^5
10^4
P0-1 BrokenHeart
P0-1 SRT
P107 SRT
F
Percent of insertions
100
75
50
25
0
Exons
Intergenic
Introns
UTRs
P0-1 BrokenHeart
P0-1 SRT
P107 SRT

Figure 3S

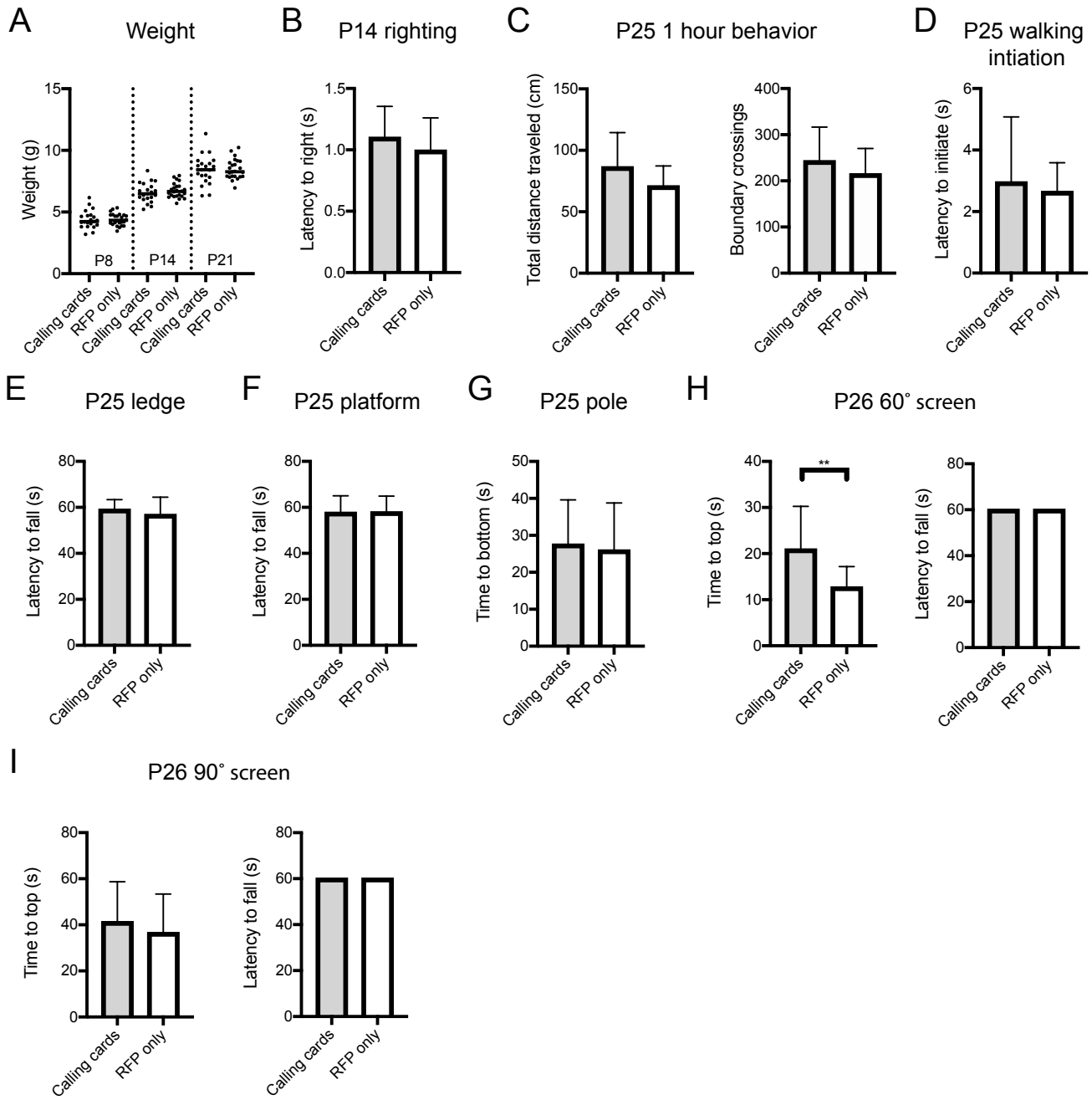

A
Weight
Weight (g)
P8
P14
P21
Calling cards
RFP only
B
P14 righting
Latency to right (s)
C
P25 1 hour behavior
Total distance traveled (cm)
Boundary crossings
D
P25 walking intiation
Latency to initiate (s)
E
P25 ledge
Latency to fall (s)
F
P25 platform
Latency to fall (s)
G
P25 pole
Time to bottom (s)
H
P26 60˚ screen
Time to top (s)
**
Latency to fall (s)
I
P26 90˚ screen
Time to top (s)
Latency to fall (s)

Figure 4S

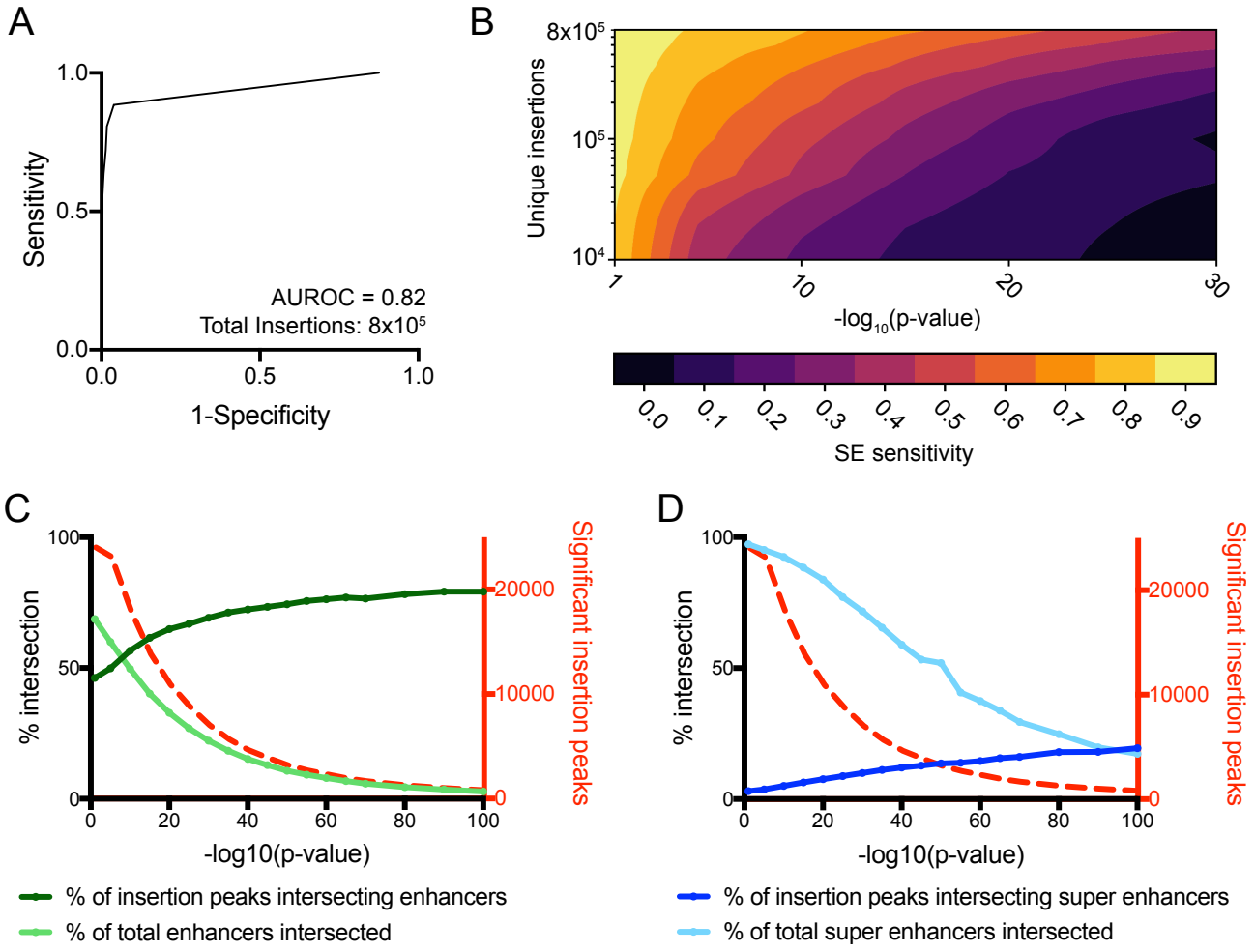

A
1.0
0.5
0.0
Sensitivity
AUROC = 0.82
Total Insertions: 8x10^5
0.0
0.5
1.0
1-Specificity
B
8x10^5
10^5
10^4
Unique insertions
1
10
20
30
-log10(p-value)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
SE sensitivity
C
100
50
0
% intersection
20000
10000
0
Significant insertion peaks
0
20
40
60
80
100
-log10(p-value)
% of insertion peaks intersecting enhancers
% of total enhancers intersected
D
100
50
0
% intersection
20000
10000
0
Significant insertion peaks
0
20
40
60
80
100
-log10(p-value)
% of insertion peaks intersecting super enhancers
% of total super enhancers intersected

Figure 5S

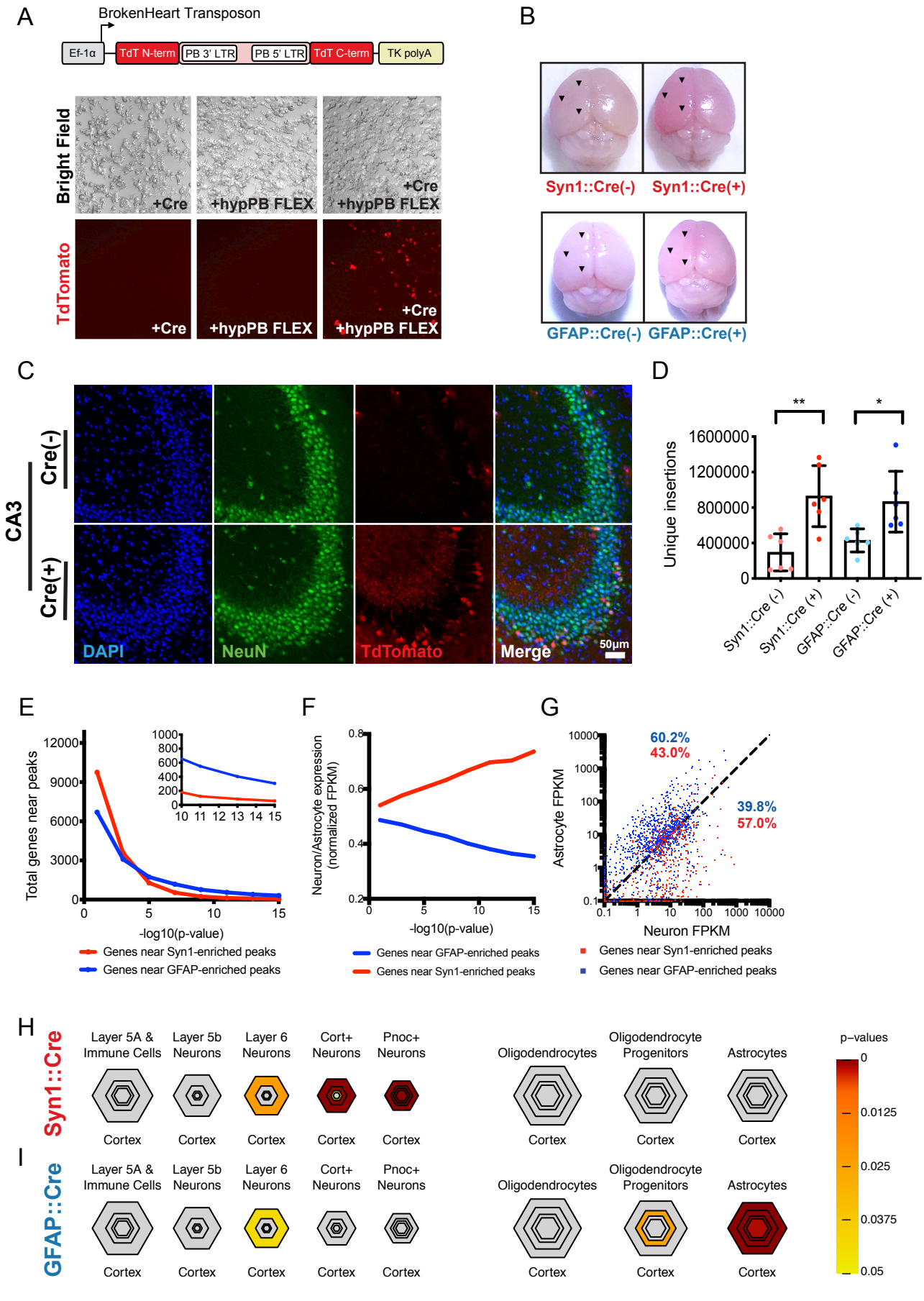
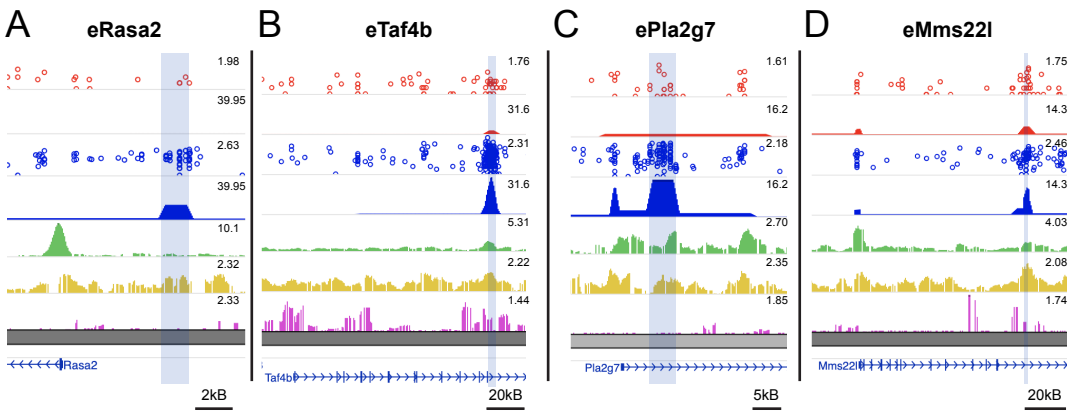
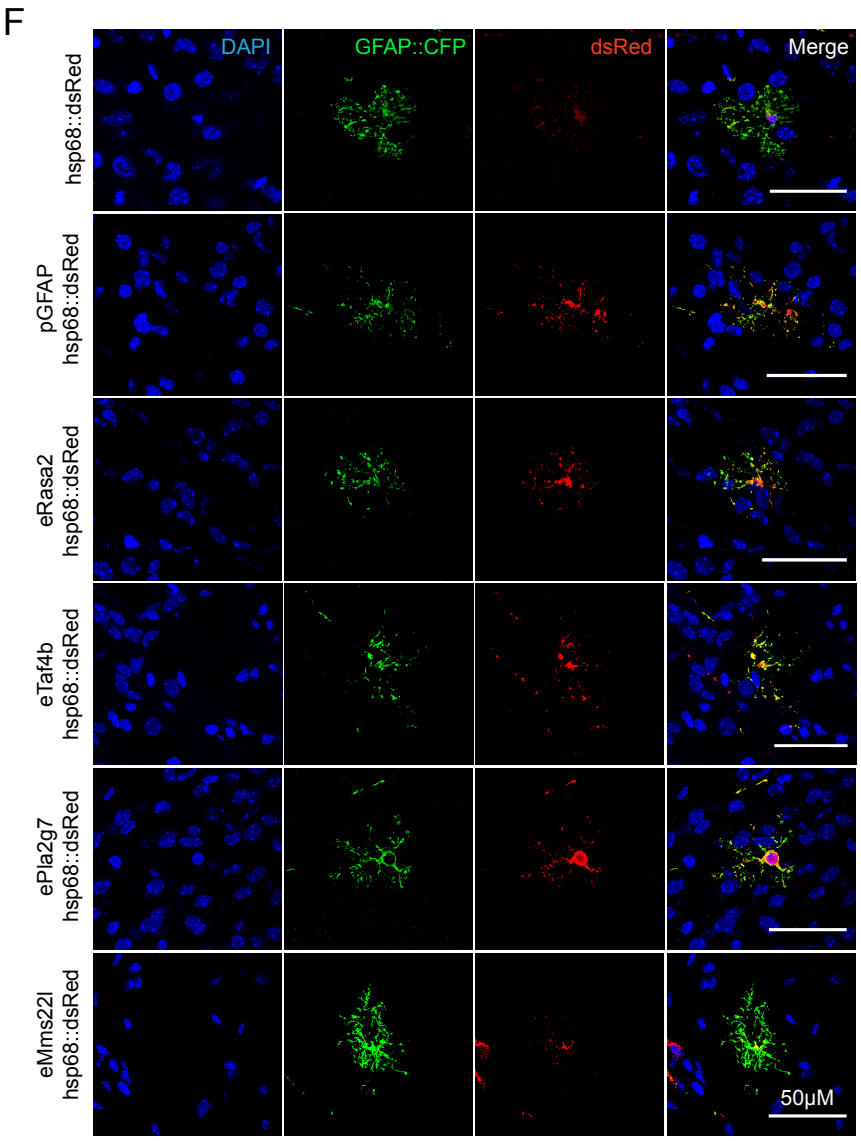

A
BrokenHeart Transposon
Ef-1α
TdT N-term
PB 3' LTR
PB 5' LTR
TdT C-term
TK polyA
Bright Field
TdTomato
+Cre
+hypPB FLEX
+Cre
+hypPB FLEX
B
Syn1::Cre(-)
Syn1::Cre(+)
GFAP::Cre(-)
GFAP::Cre(+)
C
CA3
Cre(-)
Cre(+)
DAPI
NeuN
TdTomato
Merge
50μm
D
Unique insertions
1600000
1200000
800000
400000
0
**
*
Syn1::Cre (-)
Syn1::Cre (+)
GFAP::Cre (-)
GFAP::Cre (+)
E
Total genes near peaks
12000
9000
6000
3000
0
1000
800
600
400
200
0
10 11 12 13 14 15
0 5 10 15
-log10(p-value)
Genes near Syn1-enriched peaks
Genes near GFAP-enriched peaks
F
Neuron/Astrocyte expression (normalized FPKM)
0.8
0.6
0.4
0.2
0 5 10 15
-log10(p-value)
Genes near GFAP-enriched peaks
Genes near Syn1-enriched peaks
G
Astrocyte FPKM
10000
1000
100
10
1
0.1
60.2%
43.0%
39.8%
57.0%
0.1 1 10 100 1000 10000
Neuron FPKM
Genes near Syn1-enriched peaks
Genes near GFAP-enriched peaks
H
Syn1::Cre
Layer 5A & Immune Cells
Layer 5b Neurons
Layer 6 Neurons
Cort+ Neurons
Pnoc+ Neurons
Oligodendrocytes
Oligodendrocyte Progenitors
Astrocytes
Cortex
p–values
0
0.0125
0.025
0.0375
0.05
I
GFAP::Cre
Layer 5A & Immune Cells
Layer 5b Neurons
Layer 6 Neurons
Cort+ Neurons
Pnoc+ Neurons
Oligodendrocytes
Oligodendrocyte Progenitors
Astrocytes
Cortex

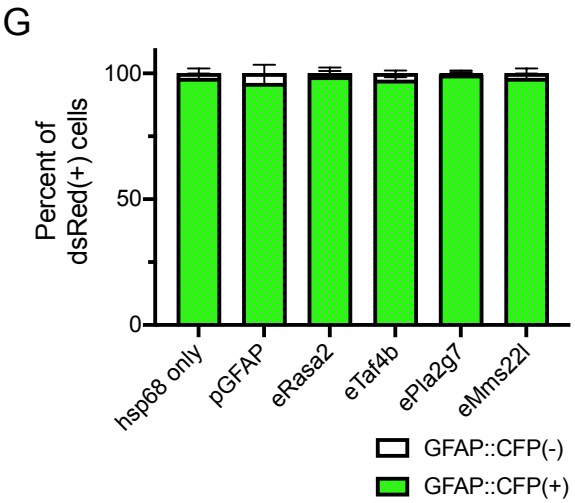
Figure 6S



E

Enhancer region	mm10 chrom. coordinates	Length (bp)
eRasa2	chr9:96636934 -96638463	1530
eTaf4b	chr18:14845190 -14846791	1602
ePla2g7	chr17:43571375 -43574084	2710
eMms22l	chr4:24544409 -24544521	112





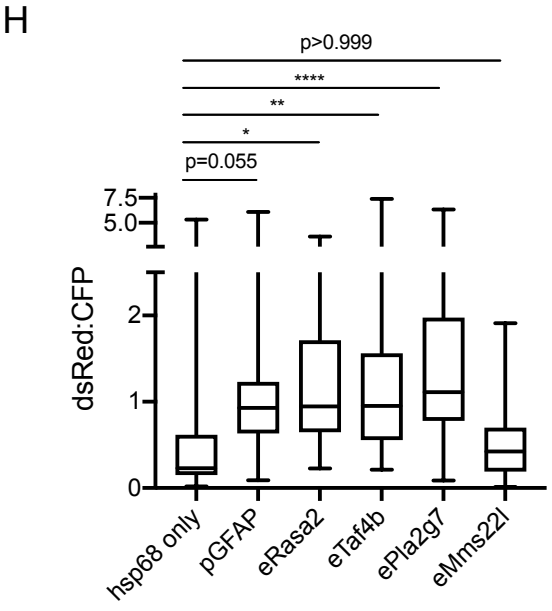


Figure 7S

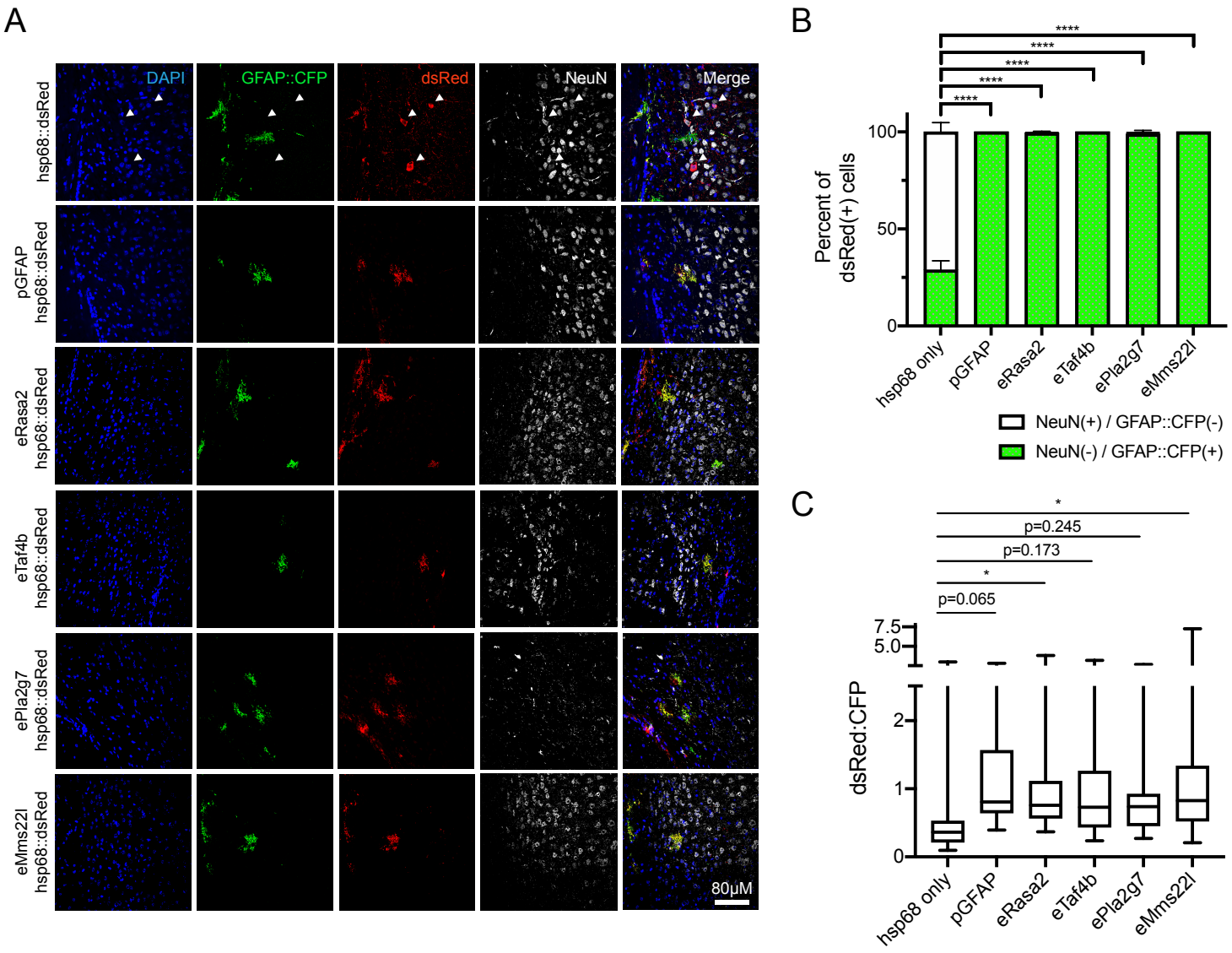
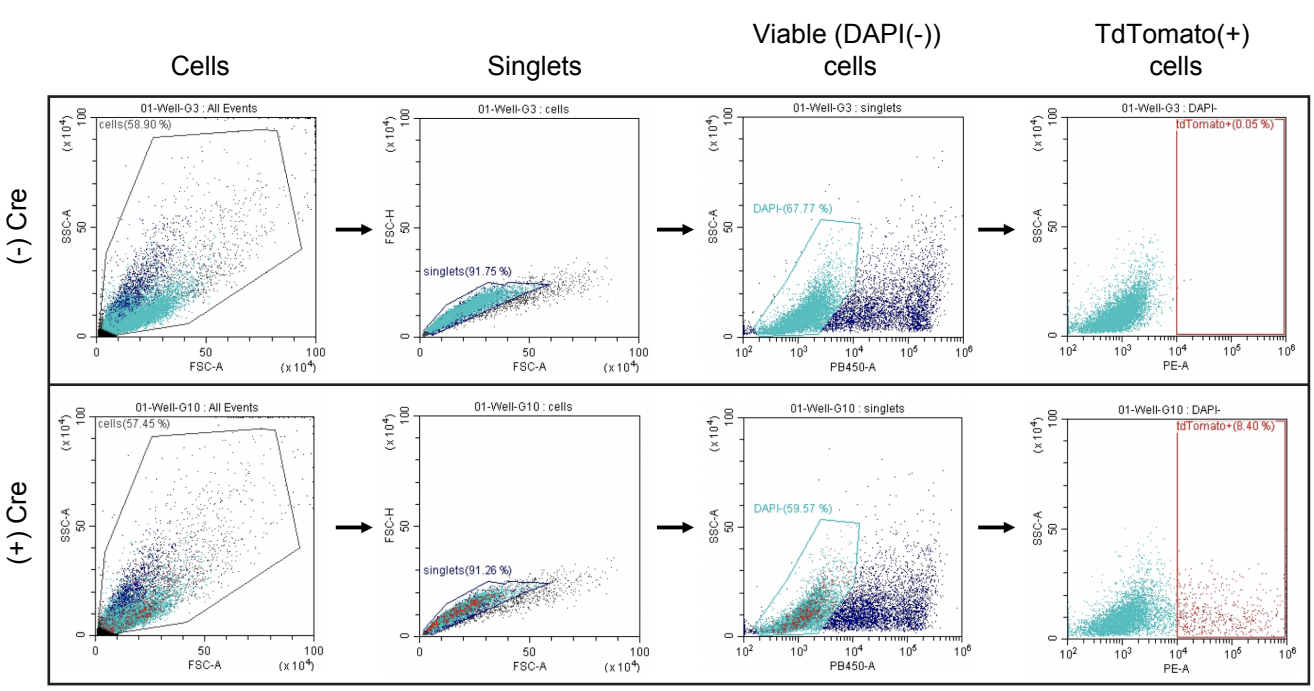

A
DAPI
GFAP::CFP
dsRed
NeuN
Merge
hsp68::dsRed
pGFAP hsp68::dsRed
eRasa2 hsp68::dsRed
eTaf4b hsp68::dsRed
ePla2g7 hsp68::dsRed
eMms22l hsp68::dsRed
80μM
B
Percent of dsRed(+) cells
100
50
0
****
****
****
****
****
hsp68 only
pGFAP
eRasa2
eTaf4b
ePla2g7
eMms22l
NeuN(+) / GFAP::CFP(-)
NeuN(-) / GFAP::CFP(+)
C
dsRed:CFP
7.5
5.0
2
1
0
*
p=0.245
p=0.173
*
p=0.065
hsp68 only
pGFAP
eRasa2
eTaf4b
ePla2g7
eMms22l

Figure 8S

A


Cells
Singlets
Viable (DAPI(-)) cells
TdTomato(+) cells
(-) Cre
(+) Cre
01-Well-G3 : All Events
cells(58.90 %)
01-Well-G3 : cells
singlets(91.75 %)
01-Well-G3 : singlets
DAPI-(67.77 %)
01-Well-G3 : DAPI-
tdTomato+(0.05 %)
01-Well-G10 : All Events
cells(57.45 %)
01-Well-G10 : cells
singlets(91.26 %)
01-Well-G10 : singlets
DAPI-(59.57 %)
01-Well-G10 : DAPI-
tdTomato+(8.40 %)
FSC-A
SSC-A
FSC-H
PB450-A
PE-A