A viral toolkit for recording transcription factor–DNA interactions in live mouse tissues

Alexander J. Cammack[a], Arnav Moudgil[b,c,d], Jiayang Chen[c], Michael J. Vasek[c], Mark Shabsovich[a], Katherine McCullough[c], Allen Yen[c], Tomas Lagunas[c], Susan E. Maloney[c], June He[b,c], Xuhua Chen[b,c], Misha Hooda[a], Michael N. Wilkinson[b,c], Timothy M. Miller[a], Robi D. Mitra[b,c], and Joseph D. Dougherty[c,e,1]

[a]Department of Neurology, Washington University School of Medicine, St. Louis, MO 63110; [b]Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO 63110; [c]Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110; [d]Medical Scientist Training Program, Washington University School of Medicine, St. Louis, MO 63110; and [e]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110

Edited by Nathaniel Heintz, Rockefeller University, New York, NY, and approved March 2, 2020 (received for review October 28, 2019)

Transcription factors (TFs) enact precise regulation of gene expression through site-specific, genome-wide binding. Common methods for TF-occupancy profiling, such as chromatin immunoprecipitation, are limited by requirement of TF-specific antibodies and provide only end-point snapshots of TF binding. Alternatively, TF-tagging techniques, in which a TF is fused to a DNA-modifying enzyme that marks TF-binding events across the genome as they occur, do not require TF-specific antibodies and offer the potential for unique applications, such as recording of TF occupancy over time and cell type specificity through conditional expression of the TF–enzyme fusion. Here, we create a viral toolkit for one such method, calling cards, and demonstrate that these reagents can be delivered to the live mouse brain and used to report TF occupancy. Further, we establish a Cre-dependent calling cards system and, in proof-of-principle experiments, show utility in defining cell type-specific TF profiles and recording and integrating TF-binding events across time. This versatile approach will enable unique studies of TF-mediated gene regulation in live animal models.

transcription factor | epigenetics | enhancer | brain | recording

Proper cellular development and function are complex processes established by elaborate gene expression networks. These networks are fundamentally regulated by transcription factors (TFs), which bind to regulatory elements (REs) across the genome and facilitate gene expression through a variety of mechanisms, including recruitment of transcriptional cofactors and modulation of chromatin state (1). Extensive efforts to profile TF genome occupancy and identify active REs across the genome have highlighted the profound diversity of TF binding, providing important insights into TF-mediated gene regulation (2–5). Further, a large portion of genetic variation associated with improper cellular function or disease has been shown to exist in TF-bound REs (3, 6–10), demonstrating the criticality of proper TF binding in maintaining cellular homeostasis.

Several methods exist for profiling TF occupancy across the genome. Antibody-based techniques, such as chromatin immunoprecipitation followed by sequencing (ChIP-seq) and, more recently, Cleavage Under Targets and Release Using Nuclease (CUT&RUN) (11) or Tagmentation (CUT&Tag) (12), are widely used and have provided numerous insights into the cellular functions of TFs (2–4, 7). Notably, however, these methods require the availability and individual optimization of TF-specific antibodies, limiting the throughput and breadth of genome-wide TF profiling. Further, ChIP-seq provides only a snapshot of TF activity at the moment of cell lysis and thus may be inefficient at detecting transient or infrequent TF-binding events. Finally, while robust for non-cell type-selective, tissue-level analyses, it is often challenging to interpret ChIP-seq data obtained from complex tissues such as the brain, which is comprised of many different interconnected cell types. Because of this limitation, efforts have recently been made to modify ChIP-seq for cell type-specific use, either through physical nuclear sorting (10, 13) or conditional expression and subsequent isolation of tagged nuclei (5, 14) or chromatin-associated enzymes (15) prior to ChIP. However, these methods thus far have been limited to highly abundant targets, such as histone modifications (5, 10) and transcriptional coactivators (15), and may be complicated by potential disassociation-related artifacts (16). Therefore, it is unclear if ChIP-seq is feasible from sorted or isolated nuclei for less abundant TFs.

An alternative approach is to record TF-binding events by fusing the TF of interest to DNA-modifying enzymes (17–20). Prominent among these are two techniques: DNA adenine methylation identification (DamID) (17), which records TF binding through local adenine methylation by an *Escherichia coli* Dam methylase fused to a TF of interest, and calling cards (18, 21), in which a TF is fused to a transposase enzyme and binding events are recorded through transposon insertion proximal to the TF-binding site. Importantly, TF-tagging techniques do not require TF-specific antibodies and have the ability to record and integrate occupancy information across time (22), while requiring

Significance

Transcription factors (TFs) play a central role in regulating gene expression. We introduce a method for profiling binding sites of TFs across the genome in living animal tissues. Our approach, termed adeno-associated virus (AAV) calling cards, works by fusing a TF of interest to a transposase, which marks TF-binding events through transposon insertion, and then delivering this construct to animal tissues, such as the mouse brain, via viral vectors. The AAV calling cards approach has several unique utilities for which we demonstrate proof-of-principle experiments, including antibody-free TF profiling, Cre-mediated cell type specificity, and longitudinal TF recording. Moving forward, this toolkit will empower innovative studies of TF-mediated regulation in animal models, thus expanding our ability to understand epigenetic control of gene expression.

Author contributions: A.J.C., A.M., J.C., M.J.V., A.Y., T.L., S.E.M., R.D.M., and J.D.D. designed research; A.J.C., A.M., J.C., M.J.V., M.S., K.M., A.Y., T.L., S.E.M., J.H., X.C., and M.H. performed research; A.J.C., A.M., J.C., M.J.V., A.Y., T.L., J.H., X.C., M.H., M.N.W., R.D.M., and J.D.D. contributed new reagents/analytic tools; A.J.C., A.M., J.C., M.J.V., A.Y., T.L., R.D.M., and J.D.D. analyzed data; and A.J.C., A.M., J.C., M.J.V., A.Y., T.L., T.M.M., R.D.M., and J.D.D. wrote the paper.

Competing interest statement: R.D.M., A.M., and M.N.W. have filed a patent application on SRT technology. No other authors have disclosures to report.

This article is a PNAS Direct Submission.

Data deposition: All raw and processed data are available through Gene Expression Omnibus (GEO) accession no. GSE128493 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128493).

[1]To whom correspondence may be addressed. Email: jdougherty@wustl.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1918241117/-/DCSupplemental.

First published April 16, 2020.

very little starting material (23). Further, these approaches offer the potential for cell type specificity through conditional expression of the TF–enzyme fusion protein. In this way, DamID has been successfully implemented for cell type-specific profiling (24), primarily in *Drosophila* (23) but also with some studies in cultured mammalian cells (25–27) and embryos (26). Meanwhile, calling cards has also been successfully applied to yeast (28) and mammalian cell (18) model systems. However, neither of these methodologies has to date been implemented for TF recording in postnatal mammalian model systems, such as mice.

Here, we adapt calling cards for in vivo use by delivering this system to the mouse brain via adeno-associated virus (AAV). This method, in the mold of traditional calling cards technologies (18), works by first expressing the *hyperPiggyBac* (hypPB) transposase within a cell and providing donor transposons. hypPB inserts donor transposons at TTAA sites throughout the genome, leaving permanent marks, or calling cards, at these loci. These transposons can later be sequenced and mapped to the genome to record the history of hypPB localization across the genome. hypPB-mediated insertions can be used to assess TF binding in two ways: 1) hypPB may be fused to a TF of interest, so that the TF directs the insertion of transposons near its genomic binding sites (18); or 2) unfused hypPB directly interacts with the bromodomain and extraterminal domain (BET) protein, BRD4, and directs transposon DNA into BRD4-associated genomic regions (29, 30), most prominently active super enhancers (7). We establish that calling cards systems can be delivered to the mouse brain via AAV and that these components successfully record TF occupancy without the need for a TF-specific antibody. We then create a conditionally expressed, Cre recombinase-dependent version of AAV calling cards, termed Flip–Excision, or FLEX, calling cards and demonstrate, as a proof of principle, the ability of this system to record cell type-specific TF-occupancy profiles in the brain. Lastly, we provide evidence that through continued transposon insertion, FLEX calling cards can record and integrate TF-binding events over extended time periods following viral delivery, providing insights into transient TF activity that would be otherwise missed with end-point measures such as ChIP-seq.

Results

Intracranial Delivery of Calling Cards via AAV Invokes Widespread Transposon Insertion in the Mouse Cortex. In order to perform transposon calling cards in mammalian cells, two basic components are required: the hypPB transposase (or a TF–hypPB fusion) and donor transposons (18). We sought to develop an in vivo method to efficiently deliver calling cards components throughout the mouse brain to identify TF-associated sites. We first tested AAV as a means for calling cards reagent delivery, as viral *piggyBac* delivery methods have been successful in other organ systems previously (31, 32). We packaged a myc-tagged version of hypPB and donor transposons carrying TdTomato reporter genes into separate AAV serotype 9 (AAV9) vectors, which efficiently transduce neuron and astrocyte populations (33), and intracranially injected these vectors into the cortices of postnatal day 0 to 1 (P0-1) mice. Animals were killed at P21 for analysis (Fig. 1*A*). We analyzed hypPB expression with in situ hybridization (*SI Appendix*, Fig. S1*A*) and transposon-derived TdTomato immunofluorescence (Fig. 1*B*) across the brain and observed widespread viral transduction in the neocortex, hippocampus, and inner brain structures. As expected with the AAV9 serotype (33), the vast majority of transduced cell types were neurons and astrocytes (Fig. 1 *C* and *D*). These results demonstrate that calling cards reagents can be efficiently delivered to the mouse brain by AAV.

Earlier implementations of the calling cards method (e.g., BrokenHeart) mapped transposon insertions by directly sequencing genomically inserted transposon DNA (18, 21) (*SI Appendix*, Fig. S2 *A* and *B*). However, our group recently developed a specialized calling cards donor transposon, termed a "self-reporting transposon" (SRT), which allows for amplification of each insertion via RNA transcription and highly efficient mapping of transposition events through deep sequencing of transposon-derived RNA (34) (*SI Appendix*, Fig. S2 *A* and *C*). We first sought to directly compare traditional DNA calling cards (BrokenHeart) with RNA calling cards (SRT) in AAV systems. To do this, we intracranially injected P0-1 mice with AAV::hypPB and either AAV::BrokenHeart or AAV::SRT. At P21, we isolated DNA or RNA from cortex samples and generated and sequenced calling cards libraries (*SI Appendix*, Fig. S2 *B* and *C*). We found that SRT reads mapped much more consistently to the mouse genome than BrokenHeart, where the majority of alignment was to the original AAV episomal sequence. We then mapped transposon insertions from these reads. While insertions from the two methods reliably mapped to similar genomic locations, we were able to recover an order of magnitude greater total number of insertions from animals receiving AAV::SRT compared with those receiving AAV::BrokenHeart (*SI Appendix*, Fig. S2 *E* and *F*). In summary, we found that SRTs provide a much greater sensitivity for recovering insertion events from AAV than traditional DNA methods. Finally, we tested whether SRT calling cards could also be delivered efficiently via adult intraparenchymal stereotactical cortical injection to precise coordinates, as this is more standard for AAV delivery than P0-1 injection. Indeed, after delivery of AAV::hypPB and AAV::SRT to three P107 adult mice (euthanized 1 mo later, at P136), we observed a near-equivalent read-mapping rate (*SI Appendix*, Fig. S2*D*), insertion total (*SI Appendix*, Fig. S2*E*), and insertion localization (*SI Appendix*, Fig. S2*F*) as in P0-1 SRT delivery. Thus, AAV calling cards systems are functional in vivo and can be delivered to the mouse brain at various time points and to targeted locations.

AAV Calling Cards Delivery to the Mouse Brain Does Not Induce Excess Degeneration, Weight Loss, or Behavioral/Developmental Defects. An important question with all calling cards technologies is whether continued transposon insertion is deleterious, particularly after long-term delivery to a living mammalian system such as the mouse brain. To address this question empirically, we intracranially delivered either calling cards viruses (AAV::hypPB/AAV::SRT) or control viruses (RFP only) to P0-1 mouse pups and then assessed neuronal degeneration, attainment of developmental milestones, anxiety-related behavior, and balance/strength/coordination during the first 4 wk of life. Four weeks after viral delivery, a small population of degenerating cells were found in the cortex, near the injection site; however, there was no significant difference between calling cards and RFP-only injected animals, suggesting that the observed degeneration was likely due to needle injury (Fig. 1 *E* and *F*). Likewise, no significant differences were observed in weight, righting from back, or edge/center dwelling, indicating that calling cards-injected animals develop normally and display no overt anxiety-like behavior (Fig. 1*G* and *SI Appendix*, Fig. S3 *A–C*). Finally, in a sensorimotor battery, calling cards-injected animals displayed a largely normal phenotype, with only one test (60°-inclined screen climbing) having a significant reduction compared with RFP-only controls. Further, these animals performed normally on the 90° inclined screen and inverted screen tests, which are even more difficult tests of balance and strength than the 60°-inclined screen (Fig. 1*H* and *SI Appendix*, Fig. S3 *D–I*). In summary, AAV calling cards reagents did not result in excess degeneration, weight loss, or behavioral/developmental defects, suggesting that genomic transposon insertion does not introduce significant toxicity or deleterious effects to the animal.

Unfused hypPB Delivered with AAV Profiles Active REs in the Brain. BRD4 acts as a "chromatin reader" by binding to acetylated lysine residues on histones (35–37) and regulating gene transcription (38, 39). Accordingly, BRD4 is strongly associated with active, TF-bound REs, most prominently super enhancers (7, 40). Importantly, BRD4 has a known affinity for the unfused hypPB protein (29), and consequently unfused hypPB insertions are greatly enriched at known BRD4-binding sites (29), such as super enhancers (30). Thus, we aimed to test the hypothesis that unfused hypPB transposon insertion can be used to identify active REs in the brain (Fig. 2*A*).

We first analyzed the sensitivity and specificity of unfused hypPB insertions for identification of active super enhancers

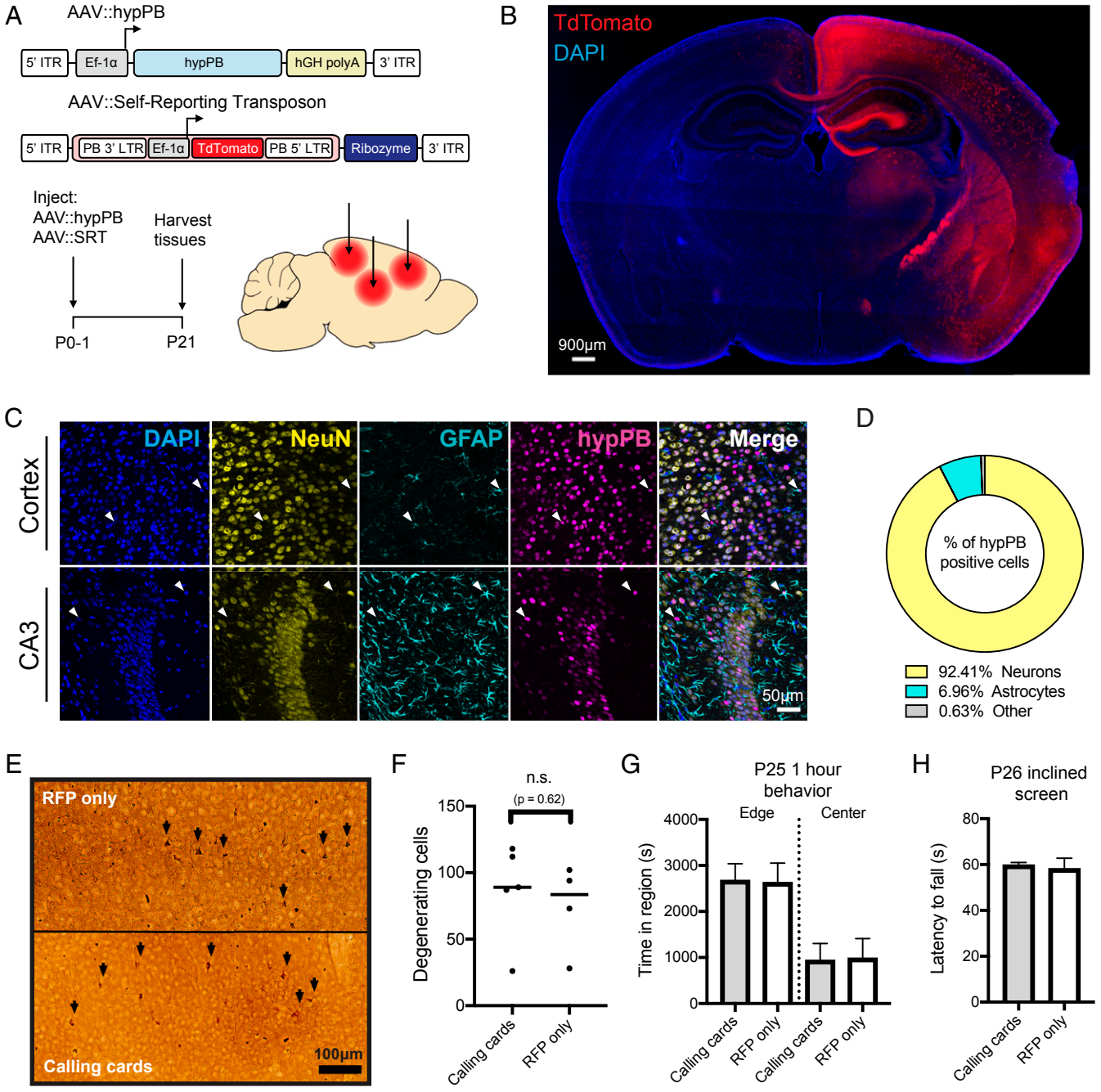

Fig. 1. Co-AAV9 intracranial injection efficiently delivers calling cards to the cortex. (*A*) Experimental paradigm and AAV constructs. Arrows represent approximate AAV injection sites. ITR, inverted-terminal repeat; LTR, long-terminal repeat. (*B*) Coronal section of a brain injected unilaterally at P0-1 with AAV::hypPB and AAV::SRT, displaying widespread expression of SRT-derived TdTomato fluorescence throughout the brain. Abnormality at right edge is tissue damage that occurred during sectioning and artifact has been removed. DAPI, 4′,6-diamidino-2-phenylindole. (*C* and *D*) Coimmunofluorescence showing hypPB expression in neurons and astrocytes. (*C*) Representative images display colocalization of hypPB with neuronal (NeuN) and astrocyte (GFAP) markers in the cortex and hippocampus. hypPB is myc-tagged, allowing for visualization with myc-specific antibodies. Arrowheads show examples of hypPB-positive astrocytes. (*D*) The majority of hypPB(+) cells transduced with AAV9 are NeuN(+) neurons and GFAP(+) astrocytes; $n = 1{,}005$ myc(+) cells, counted across cortical image fields from 5 mice. (*E*) Representative images of silver staining in the dorsal cortex to screen for degenerating cells (black arrows) in mice intracranially injected at P0-1 with red fluorescent protein (AAV::RFP only) (*Top*) or calling cards viruses (AAV::hypPB and AAV::SRT) (*Bottom*) and killed at P28. (*F*) Quantification of silver-positive cells in the dorsal cortex revealed injection with either virus produces limited neurotoxicity that did not significantly differ between groups (two-tailed, unpaired Student's *t* test, $P > 0.05$; n.s., not significant). (*G* and *H*) Mice injected at P0-1 with AAV calling cards ($n = 21$) or control, RFP-only ($n = 24$) viruses displayed no significant differences in anxiety-related behavior (center/edge dwelling) (*G*) or motor skills (inclined screen test) (*H*) relative to control. See *SI Appendix*, Fig. S3 for further behavioral and developmental assessments of these groups. All group comparisons were done with two-tailed, unpaired Student's *t* test, with Bonferroni-corrected $\alpha = 0.05$ as a significance threshold (including all tests in *SI Appendix*, Fig. S3). kB, kilobase.

in vitro in neuroblastoma (N2a) cells. While N2a's do not represent epigenetic regulation in vivo (only 20.9% of super enhancers are shared between N2a cells and P14 mouse cortex), this pure population allows for assessment of sensitivity and specificity without complex cellularity or serotype/transduction variability. To do this, we transfected these cells with hypPB and SRT plasmids, from which we collected a total of 806,653 insertions. We then downsampled this library into randomly selected pools of various insertion totals and used peak calling to identify regions of significantly enriched insertion density in each pool, at a range of significance thresholds. Using a previously published N2a H3K27ac ChIP-seq dataset (12) to independently define active super enhancers in this population, we assayed sensitivity and specificity of calling cards insertion peaks for identifying these regions. From this, we observed that calling cards peaks are highly specific for active super enhancers across a range of sensitivities, with a high area under the receiver–operator characteristic curve (0.82; See *SI Appendix*, Fig. S4*A*). Further, we observed a high sensitivity for super-enhancer identification, even at low insertion totals (e.g., sensitivity of up to 0.8 from only 10,000 insertions), that increases steadily with increasing number of insertions (*SI Appendix*, Fig. S4*B*). Thus, unfused hypPB calling cards profiles can be used to identify active super enhancers in vitro.

We next asked whether AAV::hypPB could identify active REs, including super enhancers, in the brain. To do this, we combined all 3,732,694 insertions collected from two mice injected with AAV::hypPB and AAV::SRT at P0-1 and defined significantly enriched insertion peaks (7,031 peaks; $P < 10^{-30}$), which we predict to be BRD4-bound REs. We observed that insertion density at these peaks was highly correlated between the two animal replicates, indicating a high reproducibility of this method (Fig. 2*B*). To assess whether insertion peaks represented active REs, we compared our calling cards data with Encyclopedia of DNA Elements (ENCODE) ChIP-seq datasets (33) of enhancer-associated histone modifications (4) in the developing mouse cortex. At the 7,031 significantly enriched insertion peaks, we observed a strong enrichment of active enhancer-associated histone modifications H3K27ac and H3K4me1 and a deenrichment of the repressive mark H3K27me3 (Fig. 2 *C*–*F*). We then used a previously published (41) H3K27ac ChIP-seq data from P14 mouse cortex to independently define active enhancers and super enhancers and asked whether calling cards peaks significantly overlapped these regions. We observed that the majority of insertion peaks intersected with H3K27ac-defined enhancers, significantly higher than when insertion peak coordinates are randomized (Fig. 2*G*). Similarly, calling cards peak intersection with super enhancers is also significantly higher than chance (Fig. 2*H*). As expected for a BRD4-mediated mechanism, unfused hypPB calling cards profiles identify only a subset of all enhancers but do intersect the majority of super enhancers (Fig. 2 *I* and *J*). Of note, the reference ChIP-seq dataset used for enhancer and super-enhancer identification encompasses REs from all cortical cell types, while calling cards peaks are derived only from transduced cells (i.e., neurons and astrocytes), thus likely underrepresenting RE sensitivity. Further, this overlap analysis was performed using our standard, highly rigorous significance threshold for peak calling ($P = 10^{-30}$); however, we have also performed these analyses at a range of P value thresholds to confirm the finding is robust to this parameter (*SI Appendix*, Fig. S4 *C* and *D*). Together, these data support that AAV calling cards insertion profiles of unfused hypPB can be used to identify putative REs in the brain.

FLEX Calling Cards System Allows for Cell Type-Specific Profiling of REs in the Brain. Gene-based TF-tagging systems such as calling cards offer the potential for cell type-specific profiling through conditional expression of the TF–enzyme fusion. Thus, we generated a Cre-dependent calling cards system, termed FLEX calling cards, and tested the ability of this system to record cell type-specific RE activity or TF binding in complex tissues without isolation of the cell population of interest. In the FLEX system, the reverse complement of the hypPB or TF–hypPB gene is positioned downstream of a strong, ubiquitous promoter and is flanked by two sets of loxP sites. In the presence of Cre, the transposase gene is flipped to the correct orientation and is expressed. To confirm Cre dependence of the FLEX system, we cotransfected the Cre-dependent hypPB plasmid, hypPB FLEX, into HEK293T cells along with the BrokenHeart reporter plasmid (*SI Appendix*, Fig. S5*A*), which expresses TdTomato only after genomic transposon insertion (42); 24-h posttransfection, we observed BrokenHeart-derived TdTomato fluorescence only in cells that received both the FLEX calling cards constructs and a Cre expression plasmid, demonstrating that this system is Cre-dependent (*SI Appendix*, Fig. S5*A*).

As a proof of principle, we focused on two prominent and well-studied Cre-driver mouse lines, Syn1::Cre and GFAP::Cre, which direct expression to neurons and astrocytes, respectively. We packaged the hypPB FLEX plasmid into the AAV9 vector and intracranially coinjected it along with AAV::SRT into P0-1 mouse pups of either the Syn1::Cre or GFAP::Cre genotype, along with Cre(−) littermates as controls. We euthanized the animals at P28, isolated cortical RNA, and sequenced and mapped insertions across the genome (Fig. 3*A*). Immediately apparent upon sacrifice was that brains of Syn1::Cre-positive animals were noticeably more red than their negative littermates, even to the naked eye (*SI Appendix*, Fig. S5*B*), a result of the transposition-enhanced TdTomato reporter expression derived from AAV::SRT. This change in color was striking for Syn1::Cre brains but not as apparent in GFAP::Cre animals, an observation that is consistent with the relative abundances of transduced neurons and astrocytes (Fig. 1 *C* and *D*). In Syn1::Cre brains, we analyzed TdTomato expression with immunofluorescence and noted a marked increase in neurons of Cre(+) animals but not Cre(−) littermates (*SI Appendix*, Fig. S5*C*). We then sequenced insertions in Cre(+) and Cre(−) littermates from each line and observed a significant increase in insertion events in positive animals as compared with their negative littermates (*SI Appendix*, Fig. S5*D*).

We next sought to test whether FLEX calling cards with unfused hypPB could identify cell type-specific REs. To do this, we identified insertion peaks that were differentially enriched in either Syn1::Cre over GFAP::Cre or GFAP::Cre over Syn1::Cre by a count-based statistical comparison (Fig. 3*B*). We then asked whether genes near these differentially enriched peaks are more likely to be expressed in neurons or astrocytes, using a previously published and widely used cell type-specific RNA-seq dataset (43). As predicted, we found that as our significance threshold for defining differentially enriched insertion peaks became more stringent, the RNA expression of proximal gene sets became more cell type-specific (*SI Appendix*, Fig. S5 *E* and *F*). At a stringent significance threshold of $P = 10^{-7}$, we compared all nearest genes to Syn1::Cre- or GFAP::Cre-enriched insertion peaks and found significant differences in astrocyte versus neuron expression in the expected directionalities (Fig. 3*B* and *SI Appendix*, Fig. S5*G*). Of note, these neuron and astrocyte RNA enrichments were observed despite using proximity as a means for enhancer-gene pairing, which, while widely used for analyses such as these (3), is likely only identifying the correct gene of interest in a subset of pairs (6). Lastly, we inputted these gene sets into an unbiased cell type-identification tool (Cell-Type Specific Expression Analysis [CSEA]) (44) and successfully identified cortical astrocyte and neuron populations for genes near GFAP::Cre- and Syn1::Cre-enriched insertion peaks, respectively (*SI Appendix*, Fig. S5 *H* and *I*). Together, these data indicate that peaks derived from FLEX calling cards insertion profiles recorded by unfused hypPB represent cell type-specific REs responsible for driving expression of cell type-enriched genes.

Lastly, we sought to functionally validate the enhancer activity of a subset of the previously undescribed astrocyte-enriched REs by testing whether these regions could enhance the expression of a dsRed reporter gene in astrocytes in vivo. We chose four candidate astrocyte-enriched REs, based on their size, cell type-specific activity, and astrocyte/neuron RNA expression of their nearest genes (*SI Appendix*, Fig. S6 *A*–*E*). We then cloned these

candidate REs upstream of the hsp68 minimal promoter driving a dsRed reporter gene. As a positive control, we also cloned the canonical GFAP promoter (pGFAP) (45) into the same location upstream of hsp68::dsRed (Fig. 3*C*). To test the functional enhancer activity of these REs in vivo, we delivered these plasmids, along with a separate plasmid carrying a CFP reporter under the GFAP promoter for astrocyte identification, via postnatal astrocyte-labeling electroporation (PALE) (46). At P7, mice were euthanized and brains were collected for immunohistochemistry. This method specifically targeted astrocytes in the cerebral cortex, noting that >96% of dsRed(+) cells in this region were also GFAP::CFP(+) (*SI Appendix*, Fig. S6 *F* and *G*). As expected, the positive control (pGFAP hsp68::dsRed) plasmid exhibited enhanced dsRed fluorescence in astrocytes, relative to a negative control plasmid carrying only hsp68::dsRed, that approached statistical significance ($P = 0.055$; Fig. 3 *D* and *E* and *SI Appendix*, Fig. S6 *F* and *H*). We then quantified enhancer activity of our candidate REs and observed a significant enhancement of dsRed fluorescence for three of the four candidates (Fig. 3 *D* and *E* and *SI Appendix*, Fig. S6 *F* and *H*). Thus, these astrocyte-enriched REs display functional enhancer activity in astrocytes in the mouse brain at P7. Next, we repeated this experiment allowing mice to age to P21 to allow further astrocyte maturation prior to euthanasia. To our surprise, at this time point, we observed a change in localization of dsRed expression in brains receiving the minimal promoter hsp68::dsRed construct, with fewer GFAP::CFP(+) astrocytes and a new population of NeuN(+) neurons labeled with dsRed in the cortex (*SI Appendix*, Fig. S7 *A* and *B*). This suggests that the PALE method does deliver plasmids to neurons or neural progenitors in addition to astrocyte progenitors but that expression via the hsp68 promoter in neurons does not arise until later in postnatal development. Strikingly, however, in animals that received pGFAP hsp68::dsRed or any of the RE candidate plasmids, dsRed expression was contained to GFAP::CFP(+) astrocytes, suggesting that in addition to enhancing expression in astrocytes, these RE

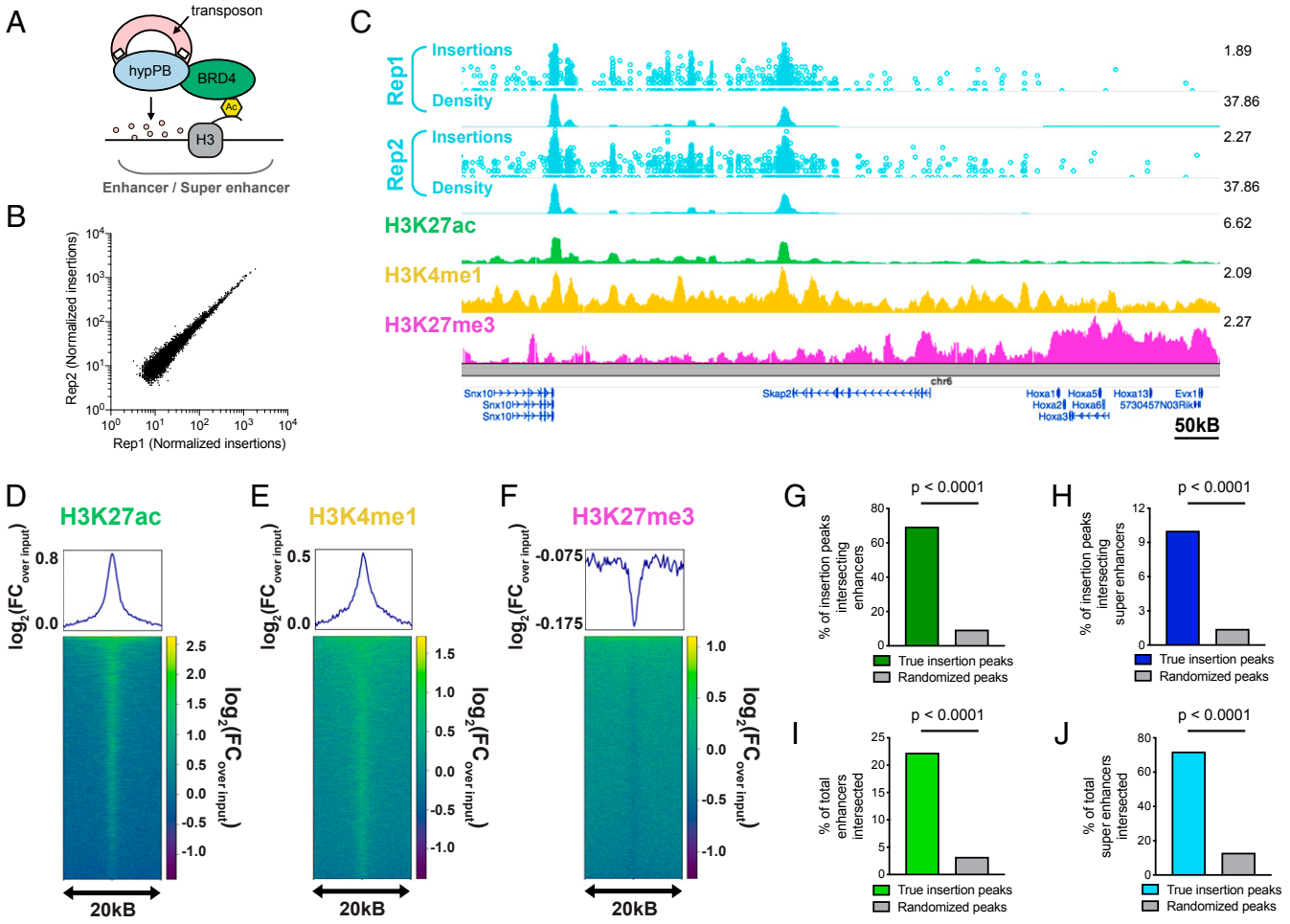

Fig. 2. Unfused hypPB-directed calling cards insertions identify active enhancers and super enhancers in the brain. (*A*) Unfused hypPB endogenously interacts with BRD4 and is redirected toward sites of BRD4 occupancy, i.e., enhancers and super enhancers. (*B*) Normalized insertion totals in two littermate C57BL/6J mice (Rep1 and Rep2) at 7,031 significantly enriched insertion peaks ($P < 10^{-30}$) displaying high correlation between replicates ($R = 0.994$). (*C–F*) Unfused hypPB-directed insertions are highly enriched for the active-enhancer marks H3K27ac and H3K4me1 and depleted for suppressive mark H3K27me3. Representative image (*C*), heat maps, and enrichment plots (*D–F*) of H3K27ac, H3K4me1, and H3K27me3 density at 7,031 significantly enriched insertion peaks in two littermate mice are shown. In *C*, the top track of each insertion replicate displays unique insertions, where each circle represents one unique insertion, the *y* axis represents the number of reads supporting each insertion on a $\log_{10}$ scale, and the bottom track displays normalized local insertion density across the genome (insertions per million per kilobase [kB]). The *y* axis of ChIP-seq data represents read depth with smoothing filter applied. Heat maps and enrichment plots are centered on insertion peaks and extend 10 kB in either direction. Relative enrichment quantifications displayed in $\log_2$(fold change over ChIP-seq input). (*G* and *H*) Percentage of 7,031 significantly enriched insertion peaks with at least 1 base pair (bp) intersection with a H3K27ac-marked enhancer or super enhancer. Gray bar represents intersections after randomizing genomic coordinates of insertion peaks. (χ^2 test with Yates correction: $P < 0.0001$.) (*I* and *J*) Percentage of H3K27ac-marked enhancers and super enhancers with at least 1-bp intersection with a significantly enriched insertion peak. (χ^2 test with Yates correction: $P < 0.0001$.)

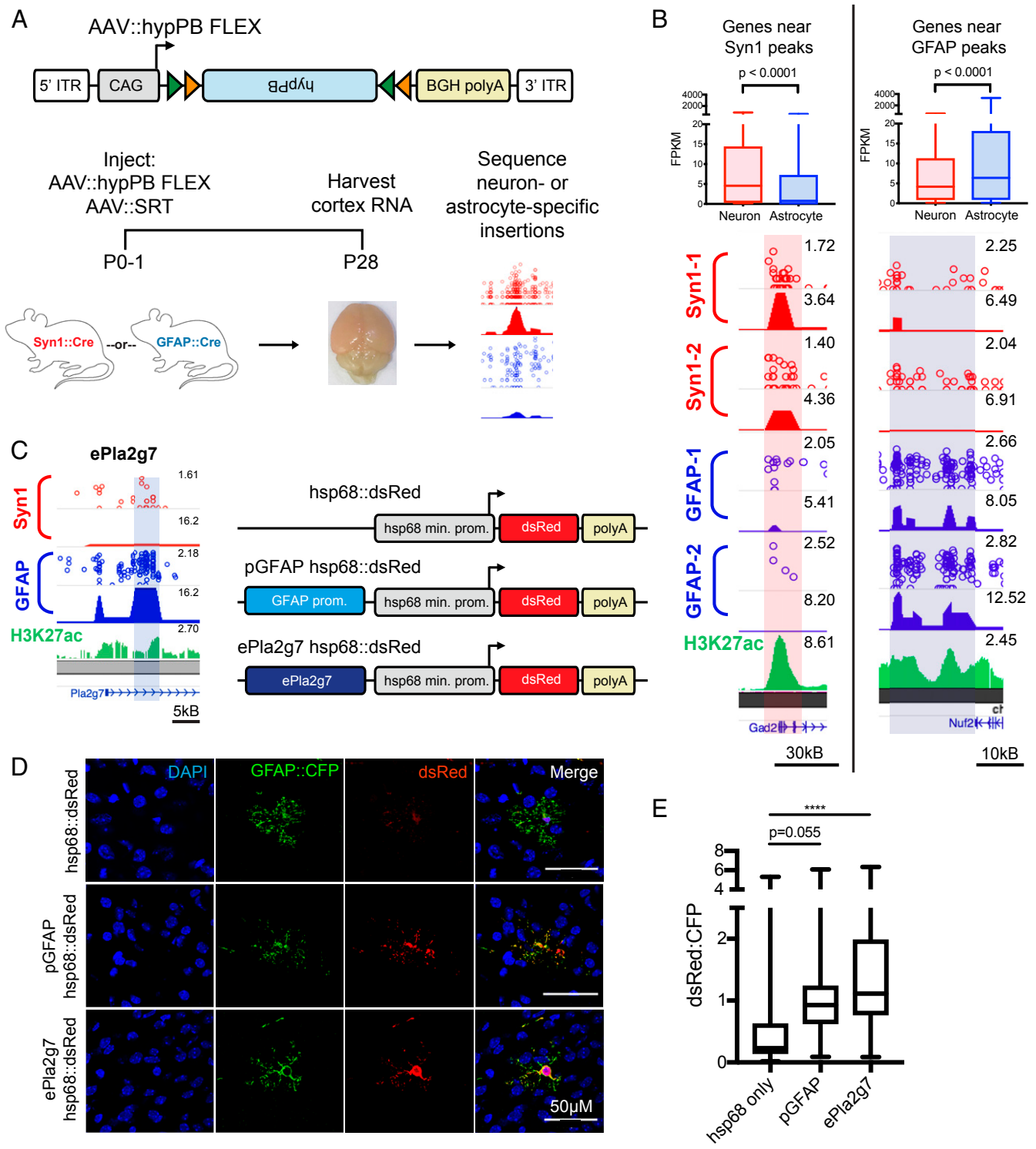

Fig. 3. FLEX calling cards system generates cell type-specific RE profiles. (*A*) AAV::hypPB FLEX construct and experimental design in Syn1::Cre and GFAP::Cre animals. ITR, inverted terminal repeat. (*B*, *Bottom*) Examples of differentially enriched insertion peaks near genes preferentially expressed in neurons (*Right*) or astrocytes (*Left*). (*B*, *Top*) Quantifications of neuron- and astrocyte-specific expression of genes near GFAP::Cre-enriched [$Neuron_{median}$ = 4.16 FPKM, $Astrocyte_{median}$ = 6.38 FPKM, n = 1,180 genes] (*C*) or Syn1::Cre-enriched [$Neuron_{median}$ = 4.55 FPKM, $Astrocyte_{median}$ = 0.78 FPKM, n = 540 genes] (*D*) insertion peaks at a stringent peak-calling significance threshold ($P = 10^{-7}$) showing significant preferential expression in the expected cell type (two-tailed Mann–Whitney U test: $P < 0.0001$). (*C*) A GFAP::Cre-enriched insertion peak proximal to the *Pla2g7* gene (ePla2g7; see *SI Appendix*, Fig. S6*E* for peak coordinates) was cloned into a plasmid upstream of the hsp68 minimal promoter and a dsRed reporter gene and codelivered along with a GFAP::CFP plasmid to ventricle-proximal glia, including astrocytes, with PALE (46). (*D*) Expression of dsRed is enhanced by both the canonical GFAP promoter (pGFAP; positive control) and ePla2g7. DAPI, 4′,6-diamidino-2-phenylindole. (*E*) Quantification of dsRed expression enhancement in CFP(+) astrocytes by pGFAP and ePla2g7; n = 34 to 42 CFP(+) cells from 3 brains per condition (one-way ANOVA with Dunnett's multiple comparisons test; ****$P < 0.0001$); $pGFAP_{mean\ difference}$ = 0.66 (95% $CI_{difference}$ = 0.01 to 1.33), $ePla2g7_{mean\ difference}$ = 1.22 (95% $CI_{difference}$ = 0.58 to 1.86). kB, kilobase.

sequences also repress activity of hsp68 in neurons (*SI Appendix*, Fig. S7 *A* and *B*). Additionally, as in the P7 experiment, we observed increased dsRed expression within GFAP::CFP(+) astrocytes for all RE constructs relative to hsp68::dsRed only, of which two reached statistical significance (*SI Appendix*, Fig. S7*C*). Indeed, even eMms22l, which was not yet active at P7, displays these functional enhancer phenotypes at P21. Overall, these data demonstrate that cell type-specific REs derived from FLEX calling cards functionally regulate cell type-specific gene expression in the brain.

Fusion of hypPB to the Promoter-Binding Transcription Factor SP1 Records SP1 Occupancy. A key feature of calling cards is the ability to record binding of a TF of interest using TF–hypPB fusions. To demonstrate in vivo TF recording with AAV calling cards, we fused hypPB to a sequence-specific DNA-binding general TF, SP1, which binds to gene promoters and is involved in transcription (47, 48), and cloned this fusion gene into the FLEX calling cards system for cell type-specific use (Fig. 4*A*). As full-length SP1 is too large to be efficiently packaged into AAV, we instead used a truncated version of SP1 containing the carboxyl-terminal (C-terminal) 621 amino acids, which includes the DNA-binding domain and has been shown to be sufficient to replicate sequence-specific binding of full-length SP1 (49). To test this system, we intracranially coinjected AAV::SP1(621C)–hypPB FLEX along with AAV::SRT into P0-1 mice of the Syn1::Cre line, killed animals at P28, generated and sequenced SRT libraries from cortical RNA samples, and compared insertion profiles with that of unfused hypPB. Consistent with the affinity of SP1 for proximal promoters, we found that insertions were significantly enriched upstream of the transcription start site (TSS), as compared with unfused hypPB insertion profiles (Fig. 4*C*). Next, we defined differentially enriched insertion peaks in SP1(621C)–hypPB profiles over unfused hypPB ($P < 10^{-15}$) and found that the majority of significant enrichments occur in gene promoters (Fig. 4 *D* and *E*). Finally, at these SP1 peaks, we performed motif discovery and identified the canonical SP1-binding motif, GGGCGGGG (18) (Fig. 4*F*). Thus, fusion of SP1 to hypPB and delivery via AAV identifies SP1-binding sites in the mouse brain.

FLEX Calling Cards Provides Historical TF-Binding Information through Longitudinal TF Recording. An intriguing potential use of calling cards technologies is in the recording of TF binding over an integrated period of time. Such a method, which is not possible with endpoint TF profiling methods such as ChIP-seq or CUT&Tag, could empower novel studies in which historical TF-binding information would be useful, such as during cellular development or differentiation. Further, by integrating signal over time, longitudinal calling cards may report transient binding events which would be otherwise missed with endpoint-only measures.

To test whether FLEX calling cards could report integrated, historical TF occupancy, we asked whether we could recover transient SP1 promoter-binding events and successfully read them out at a later date. Importantly, consistent with the known role of SP1 in regulating gene expression (18, 47, 48), we observed that expression of genes genome-wide was on average correlated with the number of SP1-directed promoter insertions (Fig. 5 *A* and *B*). Thus, we predicted that should a gene be expressed early, but not late, in the lifetime of the animal, this transient event could be marked by SP1 binding and be recoverable via SP1 calling cards at a later time point.

To test this hypothesis, we intracranially coinjected AAV::SP1(621C)–hypPB and AAV::SRT into two separate cohorts of P0-1 mice. The first cohort was euthanized at P10, while the second cohort was allowed to continue to record SP1 occupancy until P28 (Fig. 5*C*). This time period of postnatal brain development involves several key neurodevelopmental processes (50), including substantial hippocampal neurogenesis (51) as well as glial and synaptic maturation (50), development of the extracellular matrix (50), and closing of critical periods (50), which are accompanied by numerous changes in gene and protein expression (52). For these analyses, we utilized a previously published cortical RNA-seq dataset (53) with postnatal time points of 1 wk (Wk1; ~P7) and 4 wk (Wk4; ~P28). From these expression data, we then derived and tested three separate predictions (Fig. 5*C*). First, for genes expressing at Wk1 but not Wk4 (i.e., "early genes"), we would observe near-equivalent SP1 binding at promoters in both cohorts. Second, for constitutive genes that express equally at Wk1 and Wk4, we would observe continued integration of SP1 binding in the P28 cohort, resulting in increased SP1 insertion density at promoters. Third, for genes expressing only at Wk4 (i.e., "late genes"), we would observe SP1 promoter binding only in the P28 cohort. We defined early genes as having a log (Wk4/Wk1 fragments per kilobase of exon model per million reads

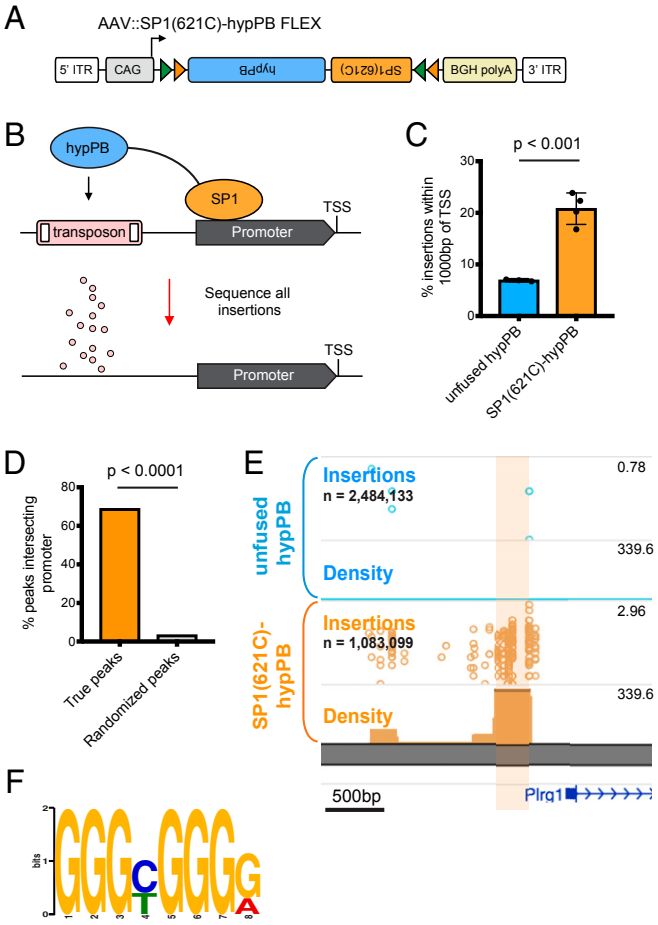

Fig. 4. Fusion of SP1 to hypPB in AAV calling cards system records SP1 occupancy. (*A*) Schematic of AAV::SP1(621C)–hypPB FLEX construct. ITR, inverted-terminal repeat. (*B*) Fusion of the promoter-binding TF SP1 to hypPB directs insertions to promoter-proximal TTAA sites. (*C*) Percentage of total insertions within 1,000 base pairs (bp) of a TSS, displaying increased promoter-directed insertions upon SP1 fusion as compared with unfused hypPB ($n = 3$ to 4 mice per group; two-tailed, unpaired Student's t test: $P < 0.001$, $t = 7.66$, degrees of freedom = 5, 95% $CI_{difference} = 9.22$ to 18.54; unfused $hypPB_{mean} = 6.9\%$, SP1[621C]–$hypPB_{mean} = 20.8\%$). Error bars represent SDs. Total SP1(621C)–hypPB insertions: 1,083,099; total unfused hypPB insertions: 2,484,133. (*D*) Percentage of significant SP1 insertion peaks differentially enriched over unfused hypPB ($P < 10^{-15}$; 1,596 intersecting out of 2,316 total peaks) intersecting promoter-proximal regions (1,000 bp on either side of TSS) compared with randomized peak coordinates (78/2,316). χ^2 test with Yates correction: $P < 0.0001$. (*E*) Representative insertion peak displaying significantly increased insertion density near the TSS of the *Plrg1* gene. (*F*) Highest information-content motif in the sequences flanking the center of significantly enriched SP1(621C)–hypPB insertion peaks ($P < 10^{-15}$) is the canonical SP1-binding motif (GGGCGGGG; $P < 10^{-42}$).

mapped [FPKM]) of less than −0.5 ($n = 292$), late genes as greater than 0.5 ($n = 285$), and all genes in the middle as constitutive ($n = 4{,}413$; Fig. 5 *D* and *E*). Indeed, at the promoters of these gene sets, we observed SP1 promoter occupancy to be consistent with our three predictions (Fig. 5*E*; example early and late genes, *Idh1* and *Gjb6*, displayed in Fig. 5 *F* and *G*, respectively). Importantly, at the promoters of early genes, we observed near-equivalent binding in the P10 and P28 cohorts (mean P28:P10 SP1 promoter insertion ratio = 0.98), despite these genes only being transiently expressed; thus, this system is capable of permanently recording transient TF-binding events for retrospective read out at a later date. Together, these data support that TF–hypPB fusions integrate signal over time and provide a historical, integrated picture of TF occupancy.

FLEX Splicing Variant, "AAV::hypPB Frontflip," Reduces Cre-Independent Transposition. During our analysis of FLEX calling cards in neurons and astrocytes, we observed the presence of some background transposition activity in Cre(−) animals (*SI Appendix*, Fig. S5*D*), which could impinge on our ability to apply FLEX calling cards to analyze more rare cell populations. We hypothesized that this aberrant hypPB expression could be due to cryptic transcription of the hypPB gene on the antisense strand of the AAV::hypPB FLEX sequence, which would result in expression independent of Cre-mediated flip excision. To overcome this, we designed a variant FLEX construct in which only the amino terminus (N terminus) of hypPB is present within the FLEX cassette and is flanked by the 5′ end of a synthetic intron (54), the 3′ end of which is located upstream of the hypPB C terminus. Thus, aberrant transcription in either direction would produce only a nonfunctional, truncated hypPB protein. However, upon Cre recombination, the N-terminal hypPB segment is flipped in frame with its C terminus, split by the newly reconstituted artificial intron, which will then be spliced out of the nascent RNA, producing a full-length, functioning hypPB. We coined this FLEX variant "hypPB Frontflip" (Fig. 6*A*).

We first tested the ability of hypPB Frontflip to reduce Cre(−) background in vitro by cotransfecting it into HEK293T cells along with the BrokenHeart reporter transposon. Previously, we had shown that the original hypPB FLEX construct produces no Cre-independent transposition in vitro at 24-h posttransfection (*SI Appendix*, Fig. S5*A*). However, at 96-h posttransfection, we began to observe some background transposition in the absence of Cre, mimicking our in vivo results. In contrast, the hypPB Frontflip variant eliminated this Cre-independent transposition in vitro (Fig. 6*B*). We then packaged hypPB Frontflip into an AAV9 vector and intracranially coinjected it with AAV::SRT into GFAP::Cre animals at P0-1. At P28, we observed a significant and dramatic reduction in Cre-independent insertions in Cre(−) animals accompanied by a >30-fold increase in Cre(+) littermates (Fig. 6*C*). To confirm that these insertions were indeed astrocyte-derived, as we had previously done with the original AAV::hypPB FLEX, we identified differential BRD4-binding peaks compared with the Syn1:Cre line and analyzed the cell type-expression patterns of proximal genes. Indeed, we found that genes near differential binding peaks again exhibited increased expression in astrocytes, as expected (Fig. 6*D*). In summary, AAV::hypPB Frontflip is a tightly Cre-controlled conditional expression viral system, which will allow for the implementation of AAV calling cards in more rare cell types in the future.

Discussion

In this report, we have successfully developed an in vivo, virally mediated calling cards approach, for which we have demonstrated utility for TF and RE profiling in the live mouse brain. This technology builds on previously developed in vitro calling cards methodologies, adapting this method for investigation of epigenetic regulation in the mammalian brain. Further, in proof-of-principle experiments, we demonstrated effectiveness of this protocol for 1) profiling TFs without an antibody, 2) cell type-specific RE profiling without cellular population purification, and 3) integrative recording of TF-binding events over time. Our use of SRTs in this paradigm now also enables calling cards for single-cell analyses (34), which expands and highlights the versatility of this toolkit in future studies.

Calling cards technologies, as gene-based systems, have several unique features that provide advantages over biochemical TF-profiling methods, such as ChIP-seq for certain applications (18, 21, 42). One such property of the FLEX version of the AAV calling cards methodology is that there is no requirement for physical isolation of cell types or nuclei for cell type-specific analyses. This allows for the same protocol to be used for any cell type of interest, the identify of which is determined by the Cre-driver mouse line used, and avoids potential disassociation-related artifacts (16). Secondly, while not explored here, one could envision simple manipulations of the AAV calling cards system to allow for temporal control of the system (43). Such adaptations could allow for innovative studies in which TF binding is recorded only during defined windows of time (22). In a similar vein, we have demonstrated here the ability of AAV calling cards to integrate TF-binding information over time, which will allow for retrospective analysis of historical TF activity in cells. By applying this unique utility to SP1, we identified promoter regions of genes with accumulating SP1 binding across postnatal development and captured transient SP1-binding events at early expressed genes. Finally, AAV calling cards does not require a TF-specific antibody, allowing for TF profiling for, in theory, any packagable TF, simply by fusing it to hypPB. This being a virally mediated system allows for simple and rapid application to animal models without the need for expensive and time-consuming breeding. Intracranial injection for a standard-sized litter of mice can be completed in under an hour. Finally, the non–Cre-dependent versions of the system should be equally applicable in other species of interest, such as rats and primates. Reagents, cloning strategies, and user-friendly analysis pipelines are available upon request, making AAV calling cards readily available for neuroscience research.

Of course, there are caveats to be considered as well. Most notably, there is potential for induced mutation, given the tendency for transposons to insert into or nearby to critical gene-regulatory regions. Indeed, transposon technologies are often used in mutagenesis screens in which transposon-mediated gene disruption can be deleterious (55). However, in such studies, the transposons are specifically engineered with splice-site gene or enhancer traps, while the SRT used in AAV calling cards only drives expression of a reporter gene and the genomic sequence immediately downstream of its insertion site. Further, the transposition rate of the *piggyBac* transposase is inherently low (<20 to 100 per cell) (56, 57), suggesting that it is highly unlikely for insertions to disrupt regulatory regions on both alleles in the same cell. Consistent with this, we observed no excess degeneration or behavioral/developmental deficits in AAV calling cards-injected animals beyond that induced by needle injury, and in general, calling cards technologies have not exhibited marked deleterious effects in previous reports (18, 21).

The relatively low "per cell" transposition rate of hypPB also helps to alleviate concerns of reduced insertion efficiency at specific loci due to saturation after long periods of calling cards recording. Indeed, with millions of available TTAA sites and thousands of active REs across the genome, the chances of two insertions falling in the same site in the same cell are on the order of less than 10^{-6} to 10^{-12}. Nevertheless, it remains possible that a small subset of calling cards transposition events could perturb RE activity and nascent gene expression, which could, in theory, alter the ability for hypPB to insert a second transposon into a nearby site. Further, while we have demonstrated here the ability to record TF-binding peaks across 28 d of FLEX calling cards expression, it should be noted that calling cards recording may be limited in some cases by transposon availability, which could become exhausted after long periods of recording. Our final experiment with SP1 at P10 and P28 demonstrates the acquisition of new peaks at late gene promoters, despite early genes using many of the available transposons, suggesting that such exhaustion effects are limited here.

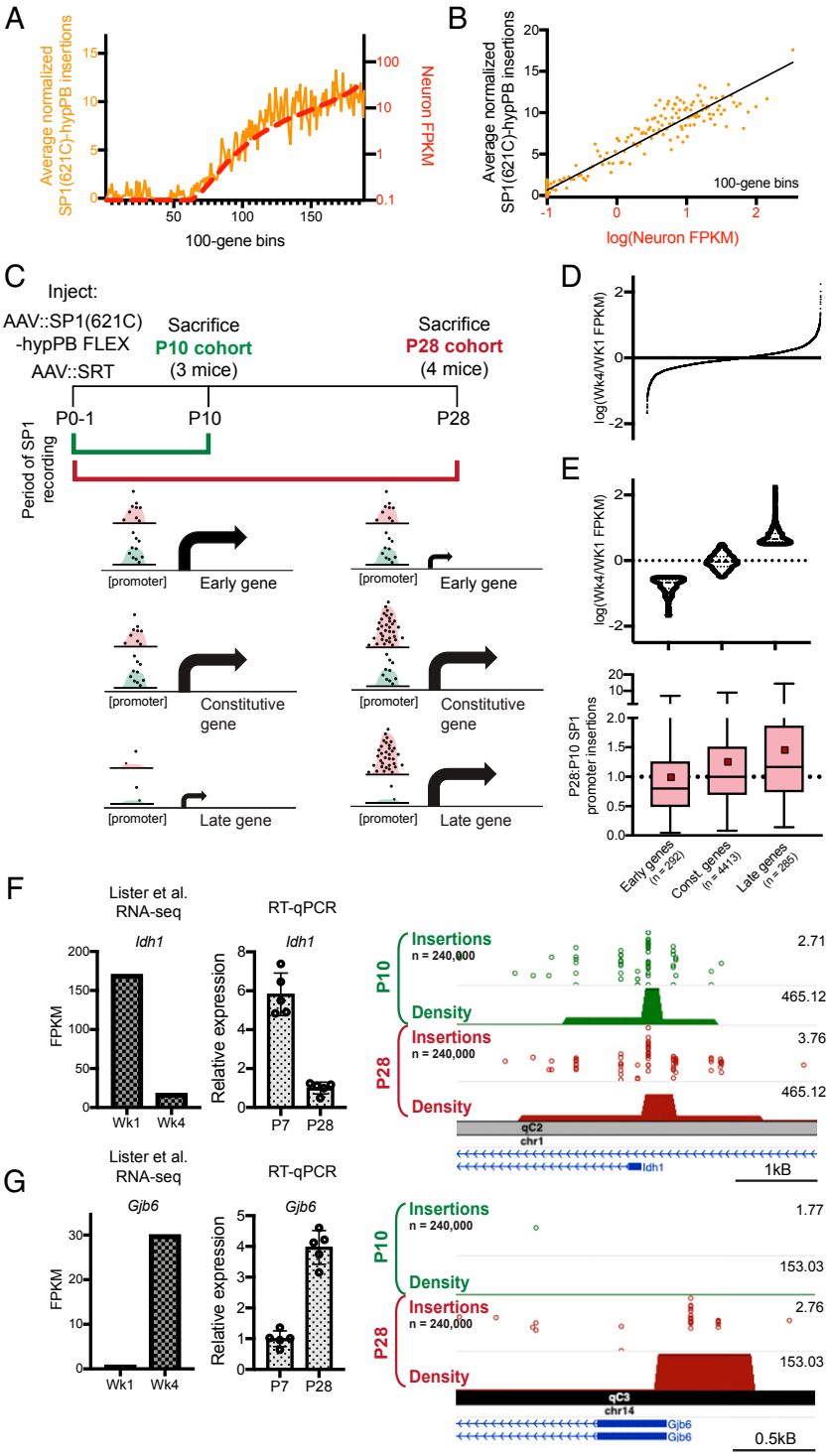

Fig. 5. Longitudinal SP1 profiling reports integrated record of SP1 binding. (*A* and *B*) Normalized number of SP1(621C)–hypPB directed insertions at promoter-proximal regions after subtraction of unfused hypPB insertions, versus neuron-specific gene expression for all genes, binned and averaged into 100-gene bins. In *A*, the left *y* axis represents the number of promoter insertions normalized to 10^6 total insertions in the sample, and the right *y* axis displays neuron-specific RNA expression. Data from ref. 43. *B* displays strong correlation of SP1(621C)–hypPB promoter insertions with gene expression after subtraction of unfused hypPB insertions ($R = 0.96$, $P < 0.0001$). (*C*) Experimental paradigm and predicted temporal SP1 occupancy for early-, constitutive-, and late-expressing genes. (*D*) Distribution of Wk4/Wk1 expression ratios for all expressed and SP1-bound genes. (*E*, *Top*) Categorization of genes into "early" [log(Wk4/Wk1 FPKM) < −0.5], "constitutive" [−0.5 < log(Wk4/Wk1 FPKM) < 0.5], and "late" [log(Wk4/Wk1 FPKM) > 0.5] gene sets. Data from ref. 53. (*E*, *Bottom*) SP1-derived promoter insertions for early, constitutive, and late gene sets, demonstrating efficient capture of transient SP1-binding events at early gene promoters and continued integration of constitutive and late gene promoters in the P28 cohort relative to the P10 cohort. One-way ANOVA [$F(2, 4{,}987) = 16.92$], $P < 0.0001$. Early $\text{genes}_{\text{mean}} = 0.98$, constitutive $\text{genes}_{\text{mean}} = 1.25$, late $\text{genes}_{\text{mean}} = 1.45$. Red squares represent means; solid lines represent medians. (*F* and *G*) Example of an early-expressed gene (*Idh1*) displaying equivalent SP1 binding in both cohorts (*F*) and a late-expressed gene (*Gjb6*) displaying SP1 binding only in the P28 cohort (*G*). The left bar graph displays reference RNA-seq read counts from ref. 53, while the right bar graph displays our own confirmatory RT–quantitative PCR (RT-qPCR) from brains of C57bl6/J mice at P7 and P28. Data from ref. 53.

Next, while useful for profiling enhancers and super enhancers more broadly, the natural affinity of hypPB for BRD4 does necessitate that any experiment using a TF–hypPB be accompanied by a control with unfused hypPB only, such that TF-binding peaks can be identified with differential peak calling, as was done for SP1 in this report. Future versions of AAV calling cards systems may be improved by fusion of TFs to other, non–BRD4-biased transposases (30), although the efficiency of transposon redirection would need to be tested empirically in each case.

Finally, while we do see a clear Cre induction of AAV::hypPB FLEX, providing a proof of principle for its use in cell type-specific profiling, we also observed some background insertion events in the absence of Cre, which could be limiting for profiling of rare cell types in which signal is likely to be reduced. To directly address this, we designed and tested a variant FLEX virus, AAV::hypPB Frontflip, and in doing so, substantially reduced the Cre(−) background observed with the original FLEX construct both in vitro and in vivo. This major improvement will now allow for the application of FLEX calling cards to more rare cell types, where insertion number is likely to be diminished. Of note, "leaky" expression of sensitive enzymes in FLEX cassettes in AAV vectors is a common issue in neuroscience, with a number

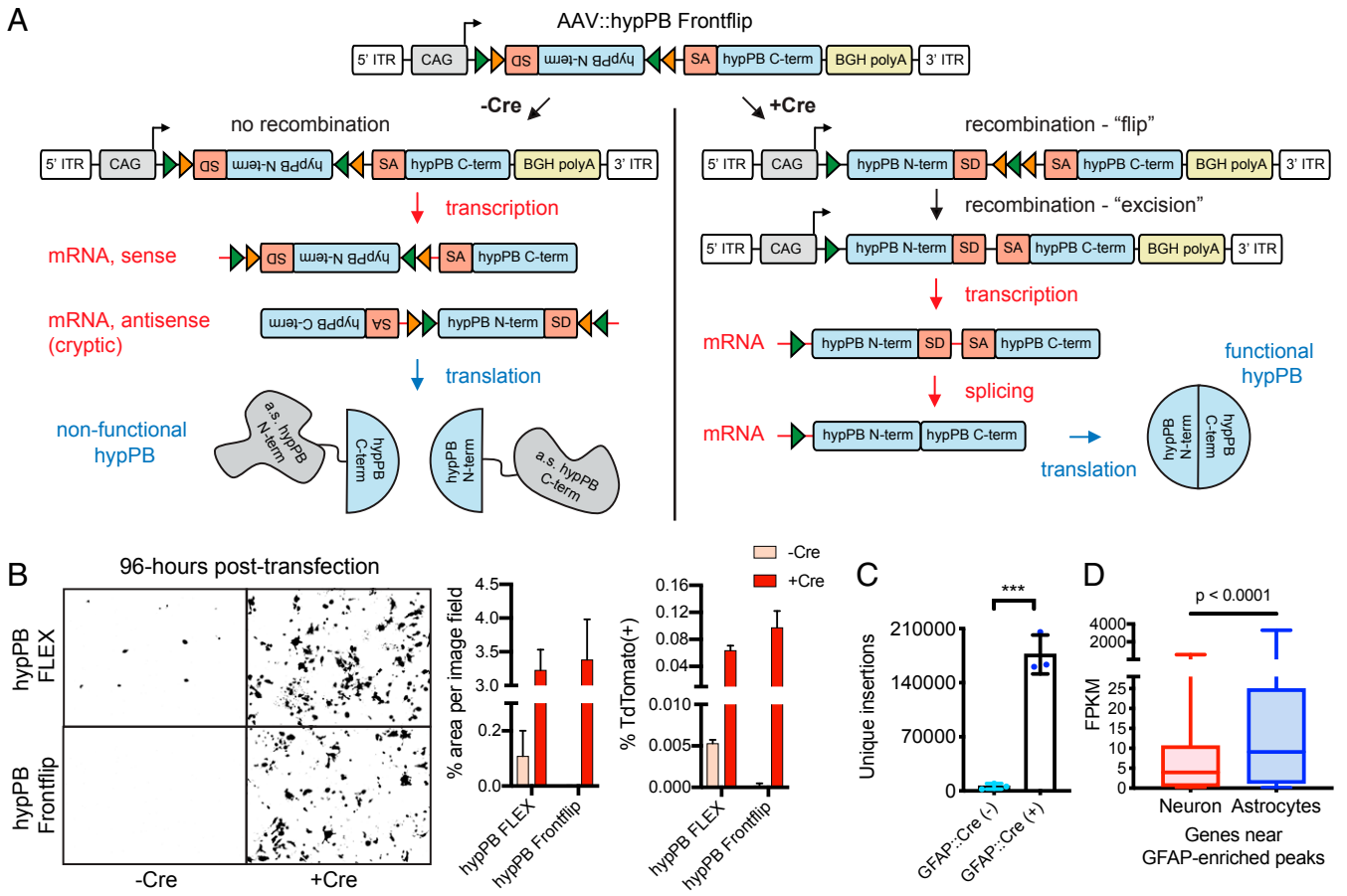

Fig. 6. AAV::hypPB Frontflip reduces Cre-independent transposition in vitro and in vivo. (*A*) Schematic of AAV::hypPB Frontflip. Prior to Cre-mediated recombination, hypPB is split into its N and C termini, with the N terminus in reverse orientation and inside a FLEX cassette. On the 3′ end of the N terminus is the splice donor (SD) and 5′ end of an artificial intron. On the 5′ end of the C terminus is the splice acceptor (SA) and 3′ end of the intron. Upon recombination, the N-terminal fragment is flipped into frame, and the artificial intron is reconstituted. This sequence is then transcribed into mRNA, spliced, and translated into an uninterrupted, functional hypPB protein. ITR, inverted-terminal repeat. (*B*) hypPB FLEX or hypPB Frontflip plasmids were cotransfected into HEK293T cells along with the BrokenHeart fluorescent transposition reporter. At 96-h posttransfection, Cre(−) background is observed from hypPB FLEX but not hypPB Frontflip. For quantification of fluorescent imaging (left graph): $n = 2$ wells per condition, 3 images per well. For flow cytometry (right graph): $n = 2$ wells per condition. (*C*) AAV::hypPB Frontflip significantly reduced transposition in Cre(−) animals (<10,000 insertions per brain), with a >30-fold increase in insertion total in Cre(+) animals (two-tailed, unpaired Student's *t* test: ***$P < 0.001$). GFAP::Cre(+) vs. (−): $P = 0.00031$, $t = 11.63$, degrees of freedom = 4, 95% $CI_{difference}$ = 129,913 to 211,436. (*D*) Quantifications of neuron- and astrocyte-specific expression of genes near GFAP::Cre-enriched [$Neuron_{median}$ = 3.96 FPKM, $Astrocyte_{median}$ = 9.10 FPKM, $n = 615$ genes] insertion peaks, called via AAV::hypPB Frontflip data ($P < 10^{-7}$), showing significant preferential expression in astrocytes (two-tailed Mann–Whitney *U* test: $P < 0.0001$).

of other strategies also having been proposed to reduce Cre(−) background—e.g., start codon/Kozak outside of FLEX cassette (58), mutated loxP sites (59), AAV titration (60)—which may also be useful in reducing FLEX calling cards background. Alternatively, simply changing the viral serotype (33) or promoter (61, 62) could allow for similar analyses in cell types not explored here without the need for Cre-dependent conditional expression.

In summary, we have introduced AAV calling cards as a viable method for recording TF binding and active REs in vivo and demonstrated its effectiveness in profiling cell type-specific and historical TF and RE activity in the brain. Future applications of this technology to animal models of development and disease could unlock important insights into epigenetic gene regulation in a variety of neuroscience disciplines.

Methods

Animals. All animal practices and procedures were approved by the Washington University in St. Louis Institutional Animal Care and Use Committee.

Cell Culture and Transfections. HEK293T and N2a cells used in this study were cultured in 1× Dulbecco's modified Eagle's medium with 10% fetal bovine serum and grown under standard conditions (37 °C; 5% CO_2). Plasmid transfections in HEK293T cells were carried out with Fugene 6 Transfection Reagent (Promega) with the manufacturer's protocol. Calling cards constructs were delivered to N2a cells via either Fugene 6 or Neon Electroporation (no. MPK10025; Thermo Fisher) with the following settings: 1,050 V, 30 ms, and two pulses.

Immunofluorescence and In Situ Hybridization. Ten- or 40-μm-thick fixed-frozen sagittal or coronal brain sections were cut, slide-mounted, and used for immunostaining or in situ hybridization (*SI Appendix, Extended Methods*). Antibodies used for immunostaining included chicken anti-GFP (Aves Labs GFP-1020) at 1:1,000 dilution, mouse rabbit anti-RFP at 1:400 or 1:500 dilution (600-401-379; Rockland), anti-NeuN at 1:100 dilution (MAB377; Millipore-Sigma), rabbit anti-cMyc at 1:250 dilution (C3956; Sigma), and goat anti-GFAP at 1:500 dilution (ab53554; Abcam). Messenger RNA (mRNA) encoding for *hyperPiggyBac* (VF1-20268-01) was detected using a custom probe set designed by Affymetrix (now Thermo Fisher).

Analysis of Brain Degeneration and Mouse Behavior 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 analyzed for anxiety-like behavior, developmental milestones, sensorimotor battery, and brain degeneration. Behavior tests were done as previously described (63, 64) at various time points in the 4 wk following injection. At P28, brains were

silver-stained as described previously (65) and analyzed for brain degeneration. See *SI Appendix, Extended Methods* for behavior and immunohistochemical procedures.

Virus Generation and Injections. Transposase and donor transposon constructs were cloned into Cre-dependent (FLEX) or Cre-independent AAV transfer vectors and used for in vitro transfection or viral packaging. Plasmids were packaged into AAV by the Hope Center Viral Vectors Core at Washington University School of Medicine. For in vivo experiments involving P0-1 delivery, transposase and donor transposon viruses (mixed equally by volume) or undiluted RFP-only virus were intracranially injected into the cortex of P0-1 mice (three sites per hemisphere; 1 μL of viral mix per site). For adult injections, viruses were delivered to P107 animals intraparenchymally with stereotactic surgery, as previously described (66). Two sites were chosen for direct, unilateral cortical injection with coordinates relative to bregma of 1.25-mm rostral, 1.5-mm lateral, and 0.55-mm depth; and 1.06-mm caudal, 1.5-mm lateral, and 0.55-mm depth. Two microliters of viral mix were delivered at a rate of 0.2 μL/min.

Viral titers (viral genomes [vg] per milliliter) were as follows:

AAV::hypPB	1.0×10^{13} to 1.1×10^{13} vg/mL
AAV::hypPB FLEX	8.0×10^{12} to 1.0×10^{13} vg/mL
AAV::SP1(621C)–hypPB FLEX	1.0×10^{13} vg/mL
AAV::SRT	1.6×10^{13} to 2.2×10^{13} vg/mL
AAV::BrokenHeart FLEX	1.4×10^{13} vg/mL
AAV::RFP only	1.6×10^{13} vg/mL
AAV::hypPB Frontflip	8.0×10^{13} vg/mL

SRT and BrokenHeart Library Preparation and Sequencing. 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). 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 (34) and sequenced on Illumina HiSEq 2500, NextSEq 500, or MiniSeq platforms (Illumina). BrokenHeart libraries were prepared from cortical DNA samples, as previously described (18), and sequenced on Illumina NextSEq 500 (Illumina). See *SI Appendix, Extended Methods* for information on read filtering and mapping. 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.

Significant Insertion Peak Calling and Motif Discovery. Significantly enriched insertion peaks were identified via a count-based statistical comparison as previously described (34) (*SI Appendix, Extended Methods*). TF motifs were identified with MEME-ChIP v4.11.2 motif discovery software (67, 68) with -zoops -meme-minw 6 -ccut 250.

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 (*SI Appendix*, Fig. S4). 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 (*SI Appendix, Extended Methods*).

Analysis of Enhancer- and Promoter-Associated Gene Expression. Gene expression has been shown to be preferentially regulated by proximal enhancer elements (4, 69, 70). Thus, since a cell type-specific mapping of enhancers to the genes they regulate is not available, we used proximity as an imperfect (6), albeit widely used (3), proxy (Figs. 3 and 6 and *SI Appendix*, Fig. S5 and *Extended Methods*). Unbiased cell type identification was completed with the CSEA tool (44) (http://genetics.wustl.edu/jdlab/csea-tool-2/) using candidate gene sets near either GFAP::Cre-enriched or Syn1::Cre-enriched insertion peaks (*SI Appendix, Extended Methods*).

For comparison of SP1 binding and gene expression in Fig. 5 *A* and *B*, we utilized the mm10_knownCanonical gene set and mm10_TSS coordinates from the University of California, Santa Cruz Gene Browser genes table, with promoter-proximal regions defined as ±1,000 bases from the TSS. For comparison of P28 and P10 SP1 promoter insertions to RNA expression in Fig. 5 *C–G*, we utilized a previously published RNA-seq dataset (53) with RNA expression data available for Wk1 and Wk4, which correspond to ~P7 and ~P28, respectively. After filtering out genes with low expression and/or no SP1 signal, 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. See *SI Appendix, Extended Methods* for more details on filtering and comparisons.

Validation of Astrocyte Enhancer Candidates with PALE (46). Candidate astrocyte-enriched enhancers were selected from the list of GFAP::Cre-enriched insertion peaks in Fig. 3, 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 (45) 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. Pups were sacrificed at P7 or P21, and brains were collected and analyzed with immunohistochemistry for CFP and dsRed. Individual astrocytes were imaged for intensity analysis in Fig. 3 *E* and *F* and *SI Appendix*, Figs. S6 *F* and *H* and S7 *A* and *C* using equivalent exposure settings (*SI Appendix, Extended Methods*).

eCandidate	Forward primer	Reverse primer
eRasa2	TTCATGAACTCTGTTAC-TAGTTTGT	TTTTAAACAGATGAG-CTGGAGCC
eTaf4b	ATATTGGATCTCACTGG-AGTTGC	TCAAAGTCTAGATTT-AGGCATGAT
ePla2g7	ACAGAACAGACTCACATAAA-CTTGT	CCATTGTCACATCTA-GTCATCAGT
eMms22l	GCTTATTTAAAATGA-AAAGA	AAATTCCCTTAAA-CCCCCCTG

Statistical Analyses. Statistical tests were done with GraphPad Prism v8.1.2 and are detailed in the figure legends. For box-and-whisker plots, central tendency line represents median, box represents 25th to 75th percentiles, and whiskers represent minimum and maximum values.

Data Availability. All raw and processed data are available through GEO accession no. GSE128493 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128493) (71). Figures containing raw data include Figs. 2–6 and *SI Appendix*, Figs. S2 and S4–S6. Supplemental tables containing information about significant peaks and figure-specific analyses are also available upon request.

Code Availability. Calling cards analysis software were developed previously (27) and are available upon request.

ACKNOWLEDGMENTS. We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for genomic analysis. The Center is partially supported by National Cancer Institute Cancer Center Support Grant P30 CA91842 to the Siteman Cancer Center and by Institute of Clinical and Translational Sciences/Clinical and Translational Sciences Award Grant UL1 TR000448 from the National Center for Research Resources (NCRR), a component of the NIH, and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH. We thank the Edison Family Center for Genome Sciences & Systems Biology, specifically Jessica Hoisington-Lopez and MariaLynn Crosby, for assistance with genomic analysis. We thank Bernard Mulvey for providing hsp68::dsRed plasmids. Finally, we thank the Kevin Noguchi, Brant Swiney, and the Intellectual and Developmental Disabilities Research Center at Washington University (National Institute of Child Health and Human Development Grant U54-HD087011) for neuropathological assessments. This work was supported by NIH Grants U01MH10913301 (to J.D.D. and R.D.M.), RF01MH117070-01 (to J.D.D. and R.D.M.), and R21HG009750 (to R.D.M.) and the Hope Center Viral Vectors Core at Washington University School of Medicine. A.J.C. was supported by NIH Grant T32GM008151-32n and the Lucille P. Markey Special Emphasis Pathway in Human Pathobiology. A.M. was supported by NIH Grants T32GM007200, T32HG000045, and F30HG009986. J.C. was supported by a McDonnell Scholarship and the Lucille P. Markey Special Emphasis Pathway in Human Pathobiology. M.J.V. was supported by NIH Grant F32NS105363-02. T.L. was supported by NIH Grant T32GM007067.

1. F. Spitz, E. E. M. Furlong, Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
2. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
3. J. Ernst *et al*., Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
4. N. D. Heintzman *et al*., Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
5. A. Mo *et al*., Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86**, 1369–1384 (2015).
6. M. T. Maurano *et al*., Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
7. D. Hnisz *et al*., Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
8. O. Corradin, P. C. Scacheri, Enhancer variants: Evaluating functions in common disease. *Genome Med.* **6**, 85 (2014).
9. A. Wells *et al*., The anatomical distribution of genetic associations. *Nucleic Acids Res.* **43**, 10804–10820 (2015).
10. K. Girdhar *et al*., Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nat. Neurosci.* **21**, 1126–1136 (2018).
11. P. J. Skene, J. G. Henikoff, S. Henikoff, Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).
12. D. Y. Kwon, Y. T. Zhao, J. M. Lamonica, Z. Zhou, Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nat. Commun.* **8**, 15315 (2017).
13. S. Bonn *et al*., Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. *Nat. Protoc.* **7**, 978–994 (2012).
14. R. B. Deal, S. Henikoff, A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev. Cell.* **18**, 1030–1040 (2010).
15. P. Zhou *et al*., Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. *eLife* **6**, 1–29 (2017).
16. S. C. van den Brink *et al*., Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).
17. B. van Steensel, S. Henikoff, Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424–428 (2000).
18. H. Wang, D. Mayhew, X. Chen, M. Johnston, R. D. Mitra, "Calling cards" for DNA-binding proteins in mammalian cells. *Genetics* **190**, 941–949 (2012).
19. G. E. Zentner, S. Kasinathan, B. Xin, R. Rohs, S. Henikoff, ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat. Commun.* **6**, 8733 (2015).
20. M. Schmid, T. Durussel, U. K. Laemmli, ChIC and ChEC; genomic mapping of chromatin proteins. *Mol. Cell* **16**, 147–157 (2004).
21. H. Wang, D. Mayhew, X. Chen, M. Johnston, R. D. Mitra, Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins. *Genome Res.* **21**, 748–755 (2011).
22. A. C. Mitchell *et al*., Longitudinal assessment of neuronal 3D genomes in mouse prefrontal cortex. *Nat. Commun.* **7**, 12743 (2016).
23. T. D. Southall *et al*., Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: Assaying RNA Pol II occupancy in neural stem cells. *Dev. Cell* **26**, 101–112 (2013).
24. J. van den Ameele, R. Krautz, A. H. Brand, TaDa! Analysing cell type-specific chromatin in vivo with Targeted DamID. *Curr. Opin. Neurobiol.* **56**, 160–166 (2019).
25. M. J. Vogel *et al*., Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res.* **16**, 1493–1504 (2006).
26. L. Tosti *et al*., Mapping transcription factor occupancy using minimal numbers of cells in vitro and in vivo. *Genome Res.* **28**, 592–605 (2018).
27. S. W. Cheetham *et al*., Targeted DamID reveals differential binding of mammalian pluripotency factors. *Development* **145**, dev.170209 (2018).
28. H. Wang, M. E. Heinz, S. D. Crosby, M. Johnston, R. D. Mitra, 'Calling Cards' method for high-throughput identification of targets of yeast DNA-binding proteins. *Nat. Protoc.* **3**, 1569–1577 (2008).
29. A. Gogol-Döring *et al*., Genome-wide profiling reveals remarkable parallels between insertion site selection properties of the MLV retrovirus and the piggyBac transposon in primary human CD4(+) T cells. *Mol. Ther.* **24**, 592–606 (2016).
30. J. Yoshida *et al*., Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Sci. Rep.* **7**, 43613 (2016).
31. R. P. Smith, J. D. Riordan, C. R. Feddersen, A. J. Dupuy, A hybrid adenoviral vector system achieves efficient long-term gene expression in the liver via *piggyBac* transposition. *Hum. Gene Ther.* **26**, 377–385 (2015).
32. A. L. Cooney, B. K. Singh, P. L. Sinn, Hybrid nonviral/viral vector systems for improved piggyBac DNA transposon in vivo delivery. *Mol. Ther.* **23**, 667–674 (2015).
33. S. L. Hammond, A. N. Leek, E. H. Richman, R. B. Tjalkens, Cellular selectivity of AAV serotypes for gene delivery in neurons and astrocytes by neonatal intracerebroventricular injection. *PLoS One* **12**, e0188830 (2017).
34. A. Moudgil *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).
35. T. Kanno *et al*., Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Mol. Cell* **13**, 33–43 (2004).
36. M. Jung *et al*., Affinity map of bromodomain protein 4 (BRD4) interactions with the histone H4 tail and the small molecule inhibitor JQ1. *J. Biol. Chem.* **289**, 9304–9319 (2014).
37. T. Kanno *et al*., BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones. *Nat. Struct. Mol. Biol.* **21**, 1047–1057 (2014).
38. G. LeRoy, B. Rickards, S. J. Flint, The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. *Mol. Cell* **30**, 51–60 (2008).
39. K. Mochizuki *et al*., The bromodomain protein Brd4 stimulates G1 gene transcription and promotes progression to S phase. *J. Biol. Chem.* **283**, 9040–9048 (2008).
40. W. A. Whyte *et al*., Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
41. H. Stroud *et al*., Early-life gene expression in neurons modulates lasting epigenetic states. *Cell* **171**, 1151–1164.e16 (2017).
42. Z. Qi *et al*., An optimized, broadly applicable *piggyBac* transposon induction system. *Nucleic Acids Res.* **45**, e55 (2017).
43. Y. Zhang *et al*., An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* **34**, 11929–11947 (2014).
44. X. Xu, A. B. Wells, D. R. O'Brien, A. Nehorai, J. D. Dougherty, Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J. Neurosci.* **34**, 1420–1431 (2014).
45. K. Sakers *et al*., Astrocytes locally translate transcripts in their peripheral processes. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E3830–E3838 (2017).
46. J. A. Stogsdill *et al*., Astrocytic neuroligins control astrocyte morphogenesis and synaptogenesis. *Nature* **551**, 192–197 (2017).
47. L. Li, S. He, J.-M. Sun, J. R. Davie, Gene regulation by Sp1 and Sp3. *Biochem. Cell Biol.* **82**, 460–471 (2004).
48. K. Beishline, J. Azizkhan-Clifford, Sp1 and the 'hallmarks of cancer'. *FEBS J.* **282**, 224–258 (2015).
49. J. T. Kadonaga, A. J. Courey, J. Ladika, R. Tjian, Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science* **242**, 1566–1570 (1988).
50. B. D. Semple, K. Blomgren, K. Gimlin, D. M. Ferriero, L. J. Noble-Haeusslein, Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Prog. Neurobiol.* **106–107**, 1–16 (2013).
51. R. Khalaf-Nazzal, F. Francis, Hippocampal development - old and new findings. *Neuroscience* **248**, 225–242 (2013).
52. A. Laeremans *et al*., Protein expression dynamics during postnatal mouse brain development. *J. Exp. Neurosci.* **7**, 61–74 (2013).
53. R. Lister *et al*., Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
54. I. Younis *et al*., Rapid-response splicing reporter screens identify differential regulators of constitutive and alternative splicing. *Mol. Cell. Biol.* **30**, 1718–1728 (2010).
55. R. Rad *et al*., A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. *Nat. Genet.* **47**, 47–56 (2015).
56. C. Kettlun, D. L. Galvan, A. L. George, Jr, A. Kaja, M. H. Wilson, Manipulating piggyBac transposon chromosomal integration site selection in human cells. *Mol. Ther.* **19**, 1636–1644 (2011).
57. K. Yusa, L. Zhou, M. A. Li, A. Bradley, N. L. Craig, A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1531–1536 (2011).
58. N. R. Wall, I. R. Wickersham, A. Cetin, M. De La Parra, E. M. Callaway, Monosynaptic circuit tracing in vivo through Cre-dependent targeting and complementation of modified rabies virus. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21848–21853 (2010).
59. K. B. Fischer, H. K. Collins, E. M. Callaway, Sources of off-target expression from recombinase-dependent AAV vectors and mitigation with cross-over insensitive ATG-out vectors. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 27001–27010 (2019).
60. T. K. Lavin, L. Jin, N. E. Lea, I. R. Wickersham, Monosynaptic tracing success depends critically on helper virus concentrations, https://doi.org/10.3389/fnsyn.2020.00006 (2020).
61. J. Jüttner *et al*., Targeting neuronal and glial cell types with synthetic promoter AAVs in mice, non-human primates and humans. *Nat. Neurosci.* **22**, 1345–1356 (2019).
62. L. T. Graybuck *et al*., Prospective, brain-wide labeling of neuronal subclasses with enhancer-driven AAVs. bioRxiv:10.1101/525014 (31 January 2019).
63. D. F. Wozniak *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).
64. R. M. Grady, D. F. Wozniak, K. K. Ohlemiller, J. R. Sanes, Cerebellar synaptic defects and abnormal motor behavior in mice lacking alpha- and beta-dystrobrevin. *J. Neurosci.* **26**, 2841–2851 (2006).
65. K. K. Noguchi, B. Nemmers, N. B. Farber, Age has a similar influence on the susceptibility to NMDA antagonist-induced neurodegeneration in most brain regions. *Brain Res. Dev. Brain Res.* **158**, 82–91 (2005).
66. S. L. DeVos, T. M. Miller, Direct intraventricular delivery of drugs to the rodent central nervous system. *J. Vis. Exp.*, e50326 (2013).
67. T. L. Bailey *et al*., MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
68. P. Machanick, T. L. Bailey, MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
69. M. P. Creyghton *et al*., Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21931–21936 (2010).
70. A. Visel *et al*., ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
71. J. D. Dougherty, A. J. Cammack, A viral toolkit for recording transcription factor-DNA interactions in live mouse tissues. Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128493. Deposited 18 March 2019.