Transposon-mediated, cell type-specific transcription factor recording in the mouse brain

Alexander J. Cammack[1], Arnav Moudgil[2,3,4], Tomas Lagunas[3], Michael J. Vasek[3], Mark Shabsovich[1], June He[2,3], Xuhua Chen[2,3], Misha Hooda[1], Michael N. Wilkinson[2,3], Timothy M. Miller[1], Robi D. Mitra[2,3], Joseph D. Dougherty[3,5]

[1]Washington University School of Medicine, Department of Neurology, St. Louis, MO

[2]Washington University School of Medicine, Edison Family Center for Genome Sciences and Systems Biology, St. Louis, MO

[3]Washington University School of Medicine, Department of Genetics, St. Louis, MO

[4]Washington University School of Medicine, Medical Scientist Training Program, St. Louis, MO

[5]Washington University School of Medicine, Department of Psychiatry, St. Louis, MO

To whom correspondence should be addressed:

Joseph Dougherty, PhD

Washington University School of Medicine

Department of Genetics & Department of Psychiatry

660 S. Euclid Ave, Campus Box 8232

St. Louis, MO 63110

Ph: (314) 286-0752

Fax: (314) 362-7855

Email: jdougherty@wustl.edu

Abstract

Transcription factors (TFs) play a central role in the regulation of gene expression, controlling everything from cell fate decisions to activity dependent transcription. However, widely-used methods for TF profiling *in vivo* (e.g. ChIP-seq) yield only an averaged picture of TF binding across all cell types present within the harvested tissue; thus, it is challenging or impossible to determine how the same TF might bind different portions of the genome in different cell types, or even to identify its binding events at all in rare cell types in a complex tissue such as the brain. Here we present a versatile methodology, FLEX calling cards, for the mapping of TF occupancy in specific cell types from heterogenous tissues. In this method, the TF of interest is fused to a hyperactive *piggyBac* transposase (hypPB), and this bipartite gene is delivered, along with donor transposons, to mouse tissue via a Cre-dependent adeno-associated virus (AAV). The fusion protein is expressed in Cre-expressing cells where it inserts transposon “calling cards” near to TF binding sites. These transposons permanently mark TF binding events and can be mapped using high-throughput sequencing. Alternatively, unfused hypPB interacts with and records the binding of the super enhancer (SE)-associated bromodomain protein, Brd4. To demonstrate the FLEX calling card method, we first show that donor transposon and transposase constructs can be efficiently delivered to the mouse brain with AAV and that insertion profiles report TF occupancy. Then, using a Cre-dependent hypPB virus, we show utility of this tool in defining cell type-specific TF profiles in multiple cell types of the brain. Finally, we demonstrate utility of FLEX calling cards in longitudinal, integrative recording of the promoter-associated TF, SP1, providing a historical record of SP1 occupancy across time. This approach will enable important cell type-specific studies of TF-mediated gene regulation in the brain and could provide valuable insights into brain development, homeostasis, and disease.

Introduction

Proper cellular development and function is a complex process established by elaborate gene expression networks. These networks are fundamentally regulated by transcription factors (TF), which bind to regulatory elements (RE) across the genome and facilitate gene expression through a variety of mechanisms, including recruitment of transcriptional co-factors 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 amongst cell types, providing important insights into cell type-specific gene regulation[2–5]. Indeed, it is of no surprise that a large portion of genetic variation associated with improper cellular function or disease is found in REs that are active only in disease-relevant cell or tissue types (e.g. Alzheimer's disease-associated variants occurring preferentially in brain-specific REs)[3,6–9].

Several methods exist for profiling TF occupancy across the genome in whole tissues. Chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq), which works by crosslinking, isolating, sequencing, and mapping TF-bound DNA, is widely used and has provided numerous insights into the cellular functions of TFs[3,4,7]. Notably however, this method requires the availability and individual optimization of TF-specific antibodies, limiting the throughput and breadth of genome-wide TF profiling. Further, 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. While efforts have been made to manipulate ChIP-seq for cell type-specific use[5,10,11], these approaches are limited by the need for physical cell sorting[11] or breeding to a knock-in mouse line[5,10]. Further, these methods have only been successfully applied to the detection of highly abundant histone proteins, and it is unclear if ChIP-seq is feasible from sorted or purified nuclei for many less abundant TFs. Finally, ChIP-seq is inherently destructive, requiring the lysis and biochemical manipulation of the tissues or cells being analyzed and provides only a snapshot of TF activity at the moment of cell sacrifice. Thus, endpoint measures like ChIP-seq cannot provide historical TF binding information and may be inefficient at detecting transient TF binding events.

Because of these limitations, efforts have recently been made to develop genetically-encoded profiling techniques which are able to record longitudinal, cell type-specific TF occupancy through use of conditional expression systems. Prominent among these is DNA adenine methylation identification (DamID), which records TF binding through targeted adenine methylation by an *E. coli* Dam methylase-TF fusion protein. DamID has been successfully implemented for cell type-specific profiling in Drosophila[12] and recently mammalian cell lines[12], however, overexpression of Dam methylase is highly cytotoxic and thus DamID has yet to be successfully implemented in *in vivo* mammalian model systems. Further, targeted adenine methylation is transient and cannot be recovered after cell division. Thus, there remains a need for cell type-specific TF recording in live mammalian tissues.

Here we present FLEX calling cards, a virally-mediated method for recording longitudinal TF binding profiles in specific cell types from heterogenous living tissues, *in vivo*. FLEX calling cards, in the mold of traditional calling card technologies[13], 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 asses 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[13], or 2) unfused hypPB directly interacts with the bromodomain and extra-terminal domain (BET) protein, Brd4, and directs transposon DNA into Brd4-associated genomic regions[14,15], most prominently active enhancers and super enhancer (SEs)[7]. Importantly, FLEX calling cards is a genetically-encoded system, allowing for cell type-specificity through conditional Cre recombinase-dependent expression without the requirement of TF-specific antibodies. The delivery of FLEX calling cards via adeno-associated virus (AAV) allows for simple and fast adaptation to a variety of mammalian model organisms, and the multitude of readily available Cre-driver mouse lines means that FLEX calling cards can be immediately used to study TFs in a variety of different cell types. Furthermore, through continued transposon insertion FLEX calling cards can record and integrate TF binding events over extended time periods between viral delivery and animal sacrifice, potentially providing insights into transient TF activity that would be otherwise missed with endpoint measures such as ChIP-seq. Finally, FLEX calling cards implements the newly-developed “self-reporting transposon”, or SRT[16], through which transposon insertion events are reported via RNA transcription; thus, this system offers the added advantage of multimodal TF-recording and RNA expression analyses, such as translating ribosome affinity purification (TRAP)[17], in the same cellular populations.

We demonstrate that FLEX calling card systems can be delivered in an efficient and widespread manner to the mouse brain via AAV and that these components successfully record TF occupancy. We then use this system to generate multiple cell type-specific Brd4 occupancy profiles and showed that these cell type-specific REs preferentially regulate known cell type-enriched genes for each respective cell type. Finally, by fusing hypPB to the promoter-binding TF, SP1, we demonstrate *in vivo*, integrated TF occupancy recording across time in mouse cortical neurons.

Results

Generation of the FLEX calling cards viral system

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[13]. We cloned both Cre-independent and dependent (i.e. "FLEX") versions of the transposase as well as donor transposon constructs into adeno-associated virus (AAV) transfer vectors (**Fig 1A**). 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 co-transfected AAV::hypPB FLEX into HEK293T cells with a transposon donor plasmid called "BrokenHeart", which acts as a reporter of transposase activity[18]. BrokenHeart contains a *piggyBac* transposon flanked by a split TdTomato reporter gene; prior to its transposition into the genome, the transposon disrupts the reporter gene, rendering it non-functional. However, upon transposition, the TdTomato gene is reconstituted seamlessly, causing the cell to report functional transposition through red fluorescence. We observed BrokenHeart fluorescence only in cells that received both the FLEX calling card constructs and a Cre expression plasmid, demonstrating that this system is Cre-dependent *in vitro* (**Fig 1B**).

Earlier implementations of the calling cards method (e.g. BrokenHeart) mapped transposon insertions by directly sequencing genomically-inserted transposon DNA[13,19] (**Fig 1SA**). 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 therefore more efficient mapping of transposition events[16]. Briefly, SRTs contain a ubiquitous promoter that drives the transcription of a TdTomato reporter gene with no 3' poly-A termination signal. When the TdTomato 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 (**Fig 1A**). However, once the transposon is inserted into the genome, the SRT is no longer affiliated with the ribozyme, and the TdTomato transcript becomes stabilized and extends to include flanking genomic sequence in its 3' UTR. In this fashion, the transposon "self-reports" its location via RNA. Through subsequent deep sequencing of the transposon-derived RNA, we are then able to record the locations of SRTs across the genome (**Fig 1C**).

Intracranial delivery of calling cards via AAV invokes widespread transposition in the mouse cortex

We sought to develop an *in vivo* method to efficiently deliver calling card components throughout the mouse brain to identify TF-associated sites. We first tested AAV as a means for calling card reagent delivery, as viral *piggyBac* delivery methods have been successful in other organ systems previously[20,21]. We packaged the AAV::hypPB construct into adeno-associated virus serotype 9 (AAV9), which efficiently transduces neuron and astrocyte populations[22], and intracranially injected this vector directly into the cortices of postnatal day 1 (P1) mice. Animals were sacrificed at P21 for analysis (**Fig 2A**). We analyzed the expression pattern of hypPB across the brain with *in situ* hybridization and observed widespread hypPB RNA expression in the neocortex

(**Fig 2SA**). Then, using immunofluorescence of a myc-tag incorporated in the hypPB gene, we confirmed protein expression of the transposase in the cortex and hippocampus (**Fig 2B**). As expected with the AAV9 serotype[22], the vast majority of transduced cell types were neurons and astrocytes (**Fig 2B-C**). Finally, by co-delivering hypPB with an AAV version of the BrokenHeart transposon (AAV::BrokenHeart transposon FLEX; **Fig 1A**), we demonstrated fluorescent reporting of functional calling card transposon insertion *in vivo* (**Fig 2SB**). These results demonstrate that calling card reagents can be efficiently delivered to the mouse brain by AAV, and establish that AAV episomes can provide donor transposons for insertion into the genome.

Next, we directly compared traditional DNA calling cards to SRT calling cards in AAV systems. To do this, we intracranially injected P1 mice with AAV::hypPB and either AAV::BrokenHeart FLEX or AAV::SRT. At P21, we isolated DNA or RNA from cortex samples and generated calling cards sequencing libraries (**Fig 1C and 1SA**). After sequencing, we mapped the reads and found that the majority (80.7%) of SRT reads mapped to the mouse genome (mm10) and had the sequence hallmarks of true insertion events. In contrast, reads from BrokenHeart were much less efficiently recovered from the genome (12.3%; **Fig 2SC**). In both cases, we noted that the many of the reads not mapping to mm10 were aligning to the original AAV episomal sequences, particularly in the BrokenHeart libraries. Thus, the improvement in SRT over BrokenHeart mapping rate likely reflects the efficacy of the ribozyme in suppressing the harvest of non-transposed SRT RNAs, thus reducing background in SRT libraries. Next, we mapped transposon insertions from these reads. From two mice receiving AAV::SRT, this resulted in 3,732,694 insertions at mean read coverage of 3.84 reads/insertion, while only 198,981 insertions were recovered at mean read coverage of 8.98 reads/insertion from three mice receiving AAV::BrokenHeart FLEX (**Fig 2SD**). In summary, we found that SRTs provide a much greater sensitivity for recovering insertion events than traditional DNA methods, and thus we have employed this system for all subsequent *in vivo* work here.

Unfused hypPB delivered with AAV records genome-wide enhancers and SEs

To demonstrate TF recording with AAV-mediated SRT calling cards, we next sought to profile the genome-wide occupancy of the chromatin-interacting TF Brd4. Brd4 acts as a “chromatin reader” by binding to acetylated lysine residues on histones[23–25] and regulating gene transcription[26,27]. Accordingly, Brd4 is strongly associated with REs such as enhancers and SEs[7,28]. Importantly, Brd4 has a known affinity for the unfused hypPB protein[14], and consequently hypPB insertions are greatly enriched at known Brd4 binding sites[14], such as SEs[15]. Thus, we aimed to test the hypothesis that unfused hypPB transposon insertion can be used to identify REs *in vivo*. We combined all 3,732,694 insertions collected from the two mice injected with AAV::hypPB and AAV::SRT at P1 and used peak calling to identify regions of significantly enriched insertion density. This resulted in 7,031 significantly ($p<10^{-30}$) enriched regions, or insertion peaks, which we predict to be biologically important REs. To test reproducibility of this method, we asked how many insertions were derived from each mouse at all 7,031 peaks and found that insertion density was highly correlated (R = 0.994) between the two replicates (**Fig 3A**). To validate that insertion peaks represented enhancers and SEs, we compared our calling card data to ENCODE ChIP-seq datasets[22] 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 de-enrichment of the repressive mark H3K27me3 (**Fig 3B-E**). We then used P14 H3K27ac ChIP-seq data from a previously published dataset[29] to independently define enhancers and SEs and asked whether calling card peaks significantly overlapped these regions. We observed that the majority (69.2%; [4,867/7,031]) of the 7,031 insertion peaks intersected with H3K27ac-defined enhancers, significantly higher than when insertion peak coordinates are randomized (9.2% [647/7,031]); χ^2 test with Yates correction: p<0.0001) (**Fig 3F**). Similarly, calling card peak intersection with H3K27ac-defined SEs is also significantly more likely than by chance (10.0% [702/7,031] v. 1.4% [97/7,031]; χ^2 test with Yates correction: p<0.0001) (**Fig 3G**). As expected for a Brd4-mediated mechanism, unfused hypPB calling card profiles identify only a subset of all H3K27ac-defined enhancers (22.2% [4,774/21,509] v. 3.2% [679/21,509]; χ^2 test with Yates correction: p<0.0001), but do intersect the majority of H3K27ac-defined SEs (71.7% [530/739] v. 12.7% [94/739]; χ^2 test with Yates correction: p<0.0001) (**Fig 3H-I**). It is important to note that these intersections were observed despite that the H3K27ac ChIP-seq dataset was derived from a P14 sample, while calling cards insertions represent signal integrated over the postnatal lifetime of the animal. Further, the ChIP-seq dataset is representative of whole brain tissue, whereas calling cards insertions account for only transduced cell types, due to the viral serotype. It is also noteworthy that 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 (**Fig 3SA-B**). Together, these data support that AAV-mediated calling card insertion profiles of unfused hypPB can be used to identify putative enhancers and SEs in the brain.

FLEX calling cards system allows for Cre-dependent, cell type-specific profiling of TFs in the brain.

We next generated a Cre-dependent calling cards system, termed FLEX calling cards, and tested the ability of this system to record cell type-specific TF binding in complex tissues without cell population purification. 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 intracranially co-injected AAV::SRT with the Cre-dependent hypPB virus, AAV::hypPB FLEX, into P1 mouse pups of either the Syn1::Cre or GFAP::Cre genotype, along with Cre(-) littermates as controls. We sacrificed the animals at P28, isolated cortical RNA, and sequenced and mapped insertions across the genome (**Fig 4A**). 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 (**Fig 4SA**), a result of the transposition-dependent TdTomato reporter expression in the AAV::SRT construct. 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 2B-C**). In Syn1::Cre brains, we analyzed TdTomato expression with immunofluorescence and noted a marked increase in expression in neurons of Cre(+) animals but not Cre(-) littermates (**Fig 4SB**). 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 to their negative littermates (**Fig 4SC**).

We next sought to validate that calling card peaks represented cell type-specific REs in these lines. We first identified significantly enriched insertion peaks across the genome with our background-free peak calling method from the combined profiles of Syn1::Cre and GFAP::Cre animals and asked whether the calling cards enrichment at these peaks would cluster by genotype. Indeed, across 11,393 significantly enriched peaks, we found that calling card insertion density was more similar within genotype than across genotypes (**Fig 4SD**). Next, we identified insertion peaks that were differentially enriched in either Syn1::Cre over GFAP::Cre or GFAP::Cre over Syn1::Cre by count-based statistical comparison and asked whether genes near these differentially active REs are near genes more likely to be expressed in neurons or astrocytes, using a previously published and widely used cell type-specific RNA-seq dataset[30] (**Fig 4B**). As expected, we found that as our significance threshold for defining differentially enriched insertion peaks became more stringent, the RNA expression of nearest gene sets became more cell type-specific (**Fig 4C and 4SE-G**). 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 4D-F**). Finally, we inputted nearest gene sets into an unbiased cell type-identification tool (CSEA[31]) and successfully identified cortical astrocyte and neuron populations for genes near GFAP::Cre and Syn1::Cre enriched insertion peaks, respectively (**Fig 4G-H**). Together, these data indicate that FLEX calling cards profiles are Cre-dependent and that insertion profiles recorded by unfused hypPB represent cell type-specific Brd4-bound REs responsible for driving expression of cell type-enriched genes.

Coupling FLEX calling cards to translating ribosome affinity purification (TRAP) enables parallel cell type-specific profiling of TF binding and actively translated RNA in rare cell types.

A remaining question with FLEX calling cards is the ability to profile TF binding in rare cell types, where insertion number is likely to be diminished and signal-to-background ratios are lower. Indeed, during our analysis of FLEX calling cards in neurons and astrocytes we observed the presence of some background insertions present in Cre(-) animals, particularly in the GFAP::Cre line (**Fig S4B**), which may impinge on our ability to detect TF-bound REs in more rare cell populations. While the cause of these low frequency, Cre-independent insertions is unclear, we reasoned that we could substantially mitigate non-cell type-specific background by simply selecting for the RNA that comes from these rare populations prior to SRT library preparation, thus potentially enhancing sensitivity when profiling rare cell types. To do this, we employed translating ribosome affinity purification (TRAP)[32,33], a method that enriches for cell type-specific transcripts via pull down of a GFP-tagged large ribosomal subunit (Ribosomal Protein L10a; Rpl10a), in conjunction with FLEX calling cards. Further, because both TRAP and SRT-based FLEX calling cards library preparations originate from RNA, this allows for simultaneous analysis of TF binding and actively translated RNA in the same targeted cell populations.

As a proof-of-principle for combinatorial TRAP/FLEX calling cards, we chose to analyze a relatively rare cortical cell population, layer V pyramidal neurons, labeled with Rbp4::Cre. These experiments were carried out in litters from a Rbp4::Cre line that was crossed with a Cre-dependent TRAP reporter mouse[34], which expresses the GFP-tagged Rpl10a subunit in the presence of Cre. As before, we intracranially injected P1

Rbp4::Cre positive and negative littermates with AAV::hypPB FLEX and AAV::SRT, sacrificed animals at P21, and collected brain tissue. As expected, we observed a strong enrichment of SRT-derived TdTomato signal in Rbp4(+) neurons, but some background in Rbp4(-) cell types (**Fig 5SA**). We then performed TRAP immunoprecipitation with anti-GFP coated beads to isolate ribosome-bound mRNA, including SRT transcription products, specifically from Rbp4::Cre(+) cells (**Fig 5A**). As expected, we were able to successfully isolate ribosome-bound RNA from Cre(+) animals but not from Cre(-) littermates, indicating an enrichment of cell type-specific RNAs (**Fig 5SB**). From the Cre(+) TRAP samples, we sequenced insertions with our standard SRT protocol and observed a total of 171,833 insertions across 4 animals, noting that these were collected from Cre(+) RNA only and represent Rbp4(+) layer V pyramidal neuron-specific insertions.

As SE regulation is critical to cell identify and function[7,28], we next aimed to identify Rbp4-enriched SEs. Importantly, because SEs are highly enriched for Brd4 binding, calling cards sensitivity for SE identification is quite high even with lower insertion totals[16]. To profile SEs, we defined significantly enriched insertion peaks with our background-free peak calling method from the 171,833 Rbp4(+) insertions and observed 223 significant peaks ($p<10^{-30}$; **Fig 5B**). As expected, the majority of Brd4 signal was observed at peaks displaying typical SE features[7,28], with the top 25% most enriched peaks harboring >50% of total insertions (top 25%: 9200 insertions; bottom 75%: 7472 insertions) and having a median length more than one order of magnitude larger than the bottom 75% (top 25%: 9.1×10^4 basepairs; bottom 75%: 6.3×10^3 basepairs) (**Fig 5C**). To define SEs specifically enriched in Rbp4(+) cells, we next identified insertion peaks differentially enriched in Rbp4::Cre animals as compared to C57BL/6J wild-type mice profiled with the Cre-independent AAV::hypPB (detailed in **Fig 3**) and identified 65 differentially enriched peaks ($p<10^{-7}$). Of these peaks, 43.1% (28/65) intersected with a known SE identified via whole cortex H3K27ac ChIP-seq data[29] and thus likely represent Rbp4(+) enriched SEs (**Fig 5D**). Similarly to our previous Syn1::Cre and GFAP::Cre analyses, we then asked whether genes near differentially Brd4-bound peaks were more likely to be expressed in Rbp4(+) neurons than other cell types. To do this, we utilized a recently published Rbp4 TRAP dataset generated by our group[35], in which ribosome-bound RNA isolated from Rbp4::Cre(+) cells was profiled and compared to pre-TRAP RNA input from the same mice. As expected, we observed that the majority of peak-proximal genes (29/47; 62%) exhibit increased expression in TRAP RNA over pre-TRAP input (**Fig 5SC**), with a median logFC of 0.287. This mirrors our results in Syn1:Cre and GFAP::Cre (**Fig 4D-F**), and suggests that these peaks represent cell type-specific REs in the rare population of Rbp4(+) layer V pyramidal neurons.

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 any TF of interest using TF-hypPB fusions. To demonstrate flexible TF recording, we fused hypPB to a sequence-specific DNA binding general TF, SP1, which binds to gene promoters and is involved in transcription[36,37], and cloned this fusion gene into the FLEX calling cards system for cell type-specific use (**Fig 6A**). As full length SP1 is too large to be efficiently packaged into AAV, we instead used a truncated version of SP1 containing the 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[38]. To test this system, we intracranially co-injected AAV::SP1(621C)-hypPB FLEX

(see **Fig 1A** for construct design) along with AAV::SRT into P1 mice of the Syn1::Cre line, sacrificed animals at P28, generated and sequenced SRT libraries from cortical RNA samples, and compared insertion profiles to that of unfused hypPB. Consistent with the affinity of SP1 for proximal promoters, we found that insertions were significantly enriched upstream of TSS, as compared to unfused hypPB insertion profiles (**Fig 6B**). 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 6C-D**). At these SP1 peaks, we performed motif discovery and were able to identify enrichment of the canonical SP1 binding motif, GGGCGGGG[13] (**Fig 6E**). Finally, as SP1 regulates expression of its associated genes[13,36,37], we asked whether SP1 binding at gene promoters would correlate with resulting neuronal gene expression. As predicted, we found that expression of genes genome-wide was on average correlated with the number of SP1-directed promoter insertions (**Fig 6F-G**). We also observed a significant correlation between gene expression and unfused hypPB (i.e. Brd4) promoter insertions (**Fig 6SA-B**), which is consistent with role of Brd4 in recruiting protein complexes such as positive transcription elongation factor b (P-TEFb) to gene promoters and promoting transcriptional elongation[39], particularly in the brain, where Brd4 has been shown to regulate transcription of genes critical to learning and memory[40]. However, the overall magnitude of promoter insertion density was substantially lower than in SP1. In summation, these data demonstrate that FLEX calling cards is capable of recording cell type-specific TF binding *in vivo* through TF-hypPB fusions.

FLEX calling cards provides historical TF binding information through longitudinal TF recording

An intriguing potential use of calling card 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, 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 intracranially co-injected AAV::SP1(621C)-hypPB and AAV::SRT into two separate cohorts of mice. After injection, the first cohort was sacrificed at P10, while the second cohort was allowed to continue to record SP1 occupancy until P28 (**Fig 7A**). This time period of postnatal brain development involves several key neurodevelopmental processes[41], including substantial hippocampal neurogenesis[42] as well as glial and synaptic maturation[41], development of the extracellular matrix[41], and closing of critical periods[41], which are accompanied by numerous changes in gene and protein expression[43]. To test whether FLEX calling cards can integrate signal over time, we compared SP1 calling card insertions at gene promoters to RNA expression in a published RNA-seq dataset[44] with postnatal timepoints of 1 week (Wk1; ~P7) and 4 weeks (Wk4; ~P28). We reasoned that if SP1(621C)-hypPB is indeed integrating signal across time, then sets of genes exhibiting increased RNA expression between Wk1 and Wk4 would also display increased SP1 calling card insertion at their promoters in the P28 cohort relative to the P10 cohort (**Fig 7A,** e.g. “Gene Y”). Indeed, we observed a positive correlation between Wk4/Wk1 RNA expression and P28/P10 SP1 promoter insertions (**Fig 7B**). Further, we hypothesized that genes with decreased RNA expression over this time period would not lose SP1 promoter insertions over

time and would thus display nearly equal SP1 binding in both the P10 and P28 cohorts. (**Fig 7A,** e.g. “Gene X”). As expected, we found that the majority of genes with decreased RNA expression between Wk1 and Wk4 exhibited 1:1 or slightly increased P28/P10 SP1 promoter insertion ratios, suggesting that the P28 cohort maintains a record of historical SP1 binding before P10 and continues to integrate signal at lowly-expressed genes over time (**Fig 7B**). Importantly, these analyses were completed on calling card data downsampled to equal insertion total in each cohort; thus, any observed differences are due to insertion distribution and localization and not increased insertion total. To confirm that these correlations were driven by SP1, we combined insertions from the P10 and P28 datasets and called significantly enriched peaks ($p<10^{-30}$) over unfused hypPB (Syn1::Cre cohort from Fig 4) to define all SP1 binding sites. Of the peaks that intersected promoters, we observed a positive correlation between P28/P10 peak insertions and Wk4/Wk1 RNA expression of their associated genes (**Fig 7C-D**). Together, these data support that TF-hypPB fusions integrate signal over time and provide a historical, integrated picture of TF occupancy.

Discussion

This work describes the development of FLEX calling cards, a versatile method for TF mapping in the brain in a cell type-specific manner. This technology builds on previously developed *in vitro* calling card methodologies and represents a substantial step forward in our ability to investigate epigenetic regulation in the brain. We have successfully implemented FLEX calling cards in the mouse brain and further, as a proof-of-principle, demonstrated the effectiveness of this protocol for cell type-specific TF profiling in multiple cell types. Also of note, a parallel report from our group[16] details adaptation of calling cards for single cell analyses, which expands and highlights the versatility of this toolkit in future studies.

Calling cards technologies, as genetically-encoded systems, have several unique features that provide advantages over biochemical TF profiling methods such as ChIP-seq for certain applications[13,18,19]. In FLEX calling cards, we have retained these important advantages and translated them to an *in vivo*, viral system which will be applicable to a range of animal models from brain development to disease. One key feature of our methodology is that there is no requirement for physical isolation of cell types 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. Secondly, while not explored here, one could envision simple manipulations of the FLEX calling cards system to allow for temporal control of the system[30]. Such adaptations could allow for innovative and powerful studies in which TF binding is recorded only during defined windows of time. In a similar vein, we have demonstrated here the ability of FLEX 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.

FLEX 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. Also, 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 size litter of mice can be completed in under an hour. Further, simply changing the viral serotype[22] or promoter[45,46] could allow for similar analyses in cell types not explored here. 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 FLEX 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 near critical gene regulatory regions. Indeed, transposon technologies are often used in mutagenesis screens in which transposon-mediated gene disruption can be deleterious[47]; however in such studies, the transposons are specifically engineered with splice-site gene or enhancer traps, while the SRT used in FLEX calling cards only drives expression of a reporter gene and the genomic sequence immediately downstream of its insertion site. Nevertheless, it remains possible that a subset of calling card transposition events could perturb nearby gene expression. That said, in general, calling cards technologies have not exhibited marked deleterious effects in previous reports[13,19]. Further, the

transposition rate of the *piggyBac* transposase is inherently low (<20 per cell[48]), suggesting that it is highly unlikely for insertions to disrupt regulatory regions on both alleles in the same cell. Finally, it is important to recognize that while we do see a clear Cre-induction of this system, we also observed a population of background insertion events in the absence of Cre, particularly in the GFAP::Cre line, which could be limiting for profiling of rare cell types in which signal is likely to be reduced. One possible solution that we explored in this study is combining the FLEX calling cards system with TRAP[33] to isolate RNAs from specific cell types, adding an extra layer of specificity. Here, we employed TRAP and FLEX calling cards in parallel and demonstrated RE profiling in a rare cell type, Rbp4(+) layer V pyramidal neurons. Perhaps more interestingly, this bimodal approach will allow for profiling of TF binding and RNA expression in tandem from the same cellular populations.

In summary, we have introduced FLEX calling cards as a viable method for recording TF binding events *in vivo* in a cell type-specific manner and demonstrated its effectiveness in profiling cell type-specific and historical TF-bound regions 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.

Acknowledgments

This work was supported by U01MH10913301 (to JDD and RMD), RF01MH117070-01 (to JDD and RMD), R21HG009750 (to RMD), and the Hope Center Viral Vectors Core at Washington University School of Medicine. AJC was supported by T32GM008151-32. AM was supported by T32GM007200, T32HG000045, and F30HG009986. TL was supported by T32GM007067. MJV was supported by F32NS105363-02. We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant# UL1 TR000448 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (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 also thank the Edison Family Center for Genome Sciences & Systems Biology, specifically Jessica Hoisington-Lopez and MariaLynn Crosby, for assistance with genomic analysis.

Author Contributions

Study designed by AJC, AM, TL, MJV, RDM, and JDD. AJC, AM, JH, XC, MNW, and MH designed and generated DNA constructs. AM developed SRT analysis pipelines. AJC, TL, MJV, MS, XC, and RDM generated and analyzed data. Manuscript written by AJC and edited by AM, TL, MJV, TMM, RDM, and JDD.

Declaration of Interests

JDD has received royalties related to TRAP in the past. RDM, AM, and MNW have filed a patent application on SRT technology. No other authors have disclosures to report.

References

1. Spitz, F. & Furlong, E. E. M. Transcription factors: From enhancer binding to developmental control. *Nat. Rev. Genet.* **13,** 613–626 (2012).
2. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
3. Ernst, J. *et al*. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
4. Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459,** 108–112 (2009).
5. Mo, A. *et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86,** 1369–1384 (2015).
6. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associate Variation in Regulatory DNA. *Science* **337,** 1190–1195 (2012).
7. Hnisz, D. *et al*. Super-Enhancers in the Control of Cell Identity and Disease. *Cell* **155,** 934–947 (2013).
8. Corradin, O. & Scacheri, P. C. Enhancer variants: Evaluating functions in common disease. *Genome Med.* **6,** 1–14 (2014).
9. Wells, A. *et al.* The anatomical distribution of genetic associations. *Nucleic Acids Res.* **43,** 10804–10820 (2015).
10. Deal, R. B. & Henikoff, S. A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue. *Dev. Cell* **18,** 1030–1040 (2010).
11. Bonn, S. *et al.* Cell type-specific chromatin immunoprecipitation from multicellular complex samples using bits-chip. *Nat. Protoc.* **7,** 978–994 (2012).
12. Cheetham, S. W. *et al.* Targeted DamID reveals differential binding of mammalian pluripotency factors. *Development* dev.170209 (2018). doi:10.1242/dev.170209
13. 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).
14. Gogol-Doring, A. *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).
15. Yoshida, J. *et al.* Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Sci. Rep.* **7,** 1–18 (2017).
16. Moudgil, A. *et al.* Self-reporting transposons enable simultaneous readout of gene expression and transcription factor binding in single cells. *BioRxiv [Preprint] https://doi.org/10.1101/538553* (2019).
17. Heiman, M., Kulicke, R., Fenster, R. J., Greengard, P. & Heintz, N. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat. Protoc.* **9,** 1282–1291 (2014).
18. Qi, Z. *et al.* An optimized, broadly applicable piggyBac transposon induction system. *Nucleic Acids Res.* **45,** 1–13 (2017).
19. Wang, H., Mayhew, D., Chen, X., Johnston, M. & Mitra, R. D. Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins. *Genome Res.* **21,** 748–755 (2011).
20. Smith, R. P., Riordan, J. D., Feddersen, C. R. & Dupuy, A. J. A Hybrid Adenoviral Vector System Achieves Efficient Long-Term Gene Expression in the Liver via *piggyBac* Transposition. *Hum. Gene Ther.* **26,** 377–385 (2015).
21. Cooney, A. L., Singh, B. K. & Sinn, P. L. Hybrid nonviral/viral vector systems for improved piggyBac DNA transposon in vivo delivery. *Mol. Ther.* **23,** 667–674 (2015).
22. Hammond, S. L., Leek, A. N., Richman, E. H. & Tjalkens, R. B. Cellular selectivity of AAV serotypes for gene delivery in neurons and astrocytes by neonatal intracerebroventricular injection. *PLoS One* **12,** 1–22 (2017).
23. Kanno, T. *et al.* Selective Recognition of Acetylated Histones by Bromodomain Proteins Visualized in Living Cells. *Mol. Cell* **13,** 33–43 (2004).
24. Jung, M. *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).
25. Kanno, T. *et al.* BRD4 assists elongation of both coding and enhancer RNAs by interacting with acetylated histones. *Nat. Struct. Mol. Biol.* **21,** 1047–1057 (2014).
26. LeRoy, G., Rickards, B. & Flint, S. J. The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription. *Mol. Cell* **30,** 51–60 (2008).
27. Mochizuki, K. *et al*. The bromodomain protein Brd4 stimulates g1 gene transcription and promotes progression to S phase. *J. Biol. Chem.* **283,** 9040–9048 (2008).
28. Whyte, W. A. *et al.* Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* **153,** 307–319 (2013).
29. Stroud, H. *et al.* Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States. *Cell* **171,** 1151–1164.e16 (2017).
30. Zhang, Y. *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).
31. 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).

32. Dougherty, J. D. The Expanding Toolkit of Translating Ribosome Affinity Purification. *J. Neurosci.* **37,** 12079–12087 (2017).
33. Doyle, J. P. *et al.* Application of a Translational Profiling Approach for the Comparative Analysis of CNS Cell Types. *Cell* **135,** 749–762 (2008).
34. Zhou, P. *et al.* Interrogating translational efficiency and lineage-specific transcriptomes using ribosome affinity purification. *Proc. Natl. Acad. Sci.* **110,** 15395–15400 (2013).
35. Ouwenga, R., Lake, A. M., Aryal, S., Lagunas, T. & Dougherty, J. The differences in local translatome across distinct neuron types is mediated by both baseline cellular differences and post-transcriptional mechanisms. *eNeuro* **Dec 26,** (2018).
36. Li, L., He, S., Sun, J.-M. & Davie, J. R. Gene regulation by Sp1 and Sp3. *Biochem. Cell Biol.* **82,** 460–471 (2004).
37. Beishline, K. & Azizkhan-Clifford, J. Sp1 and the 'hallmarks of cancer'. *FEBS J.* **282,** 224–258 (2015).
38. Kadonaga, J. T., Courey, A. J., Ladika, J. & Tjian, R. Distinct Regions of Sp1 Modulate DNA Binding and Transcriptional Activation. *Science* **242,** 1566–1570 (1988).
39. Jang, M. K. *et al.* The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol. Cell* **19,** 523–34 (2005).
40. Korb, E., Herre, M., Zucker-Scharff, I., Darnell, R. B. & Allis, C. D. BET protein Brd4 activates transcription in neurons and BET inhibitor Jq1 blocks memory in mice. *Nat. Neurosci.* **18,** 1464–1473 (2015).
41. Semple, B. D., Blomgren, K., Gimlina, K., Ferrieroe, D. M. & Noble-Haeussleina, L. J. Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Prog. Neurobiol.* **106–107,** 1–16 (2013).
42. Khalaf-Nazzal, R. & Francis, F. Hippocampal development - Old and new findings. *Neuroscience* **248,** 225–242 (2013).
43. Laeremans, A. *et al.* Protein Expression Dynamics During Postnatal Mouse Brain Development. *J. Exp. Neurosci.* **7,** 61–74 (2013).
44. Lister, R. *et al.* Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* **341,** 1237905–1:12 (2013).
45. Juettner, J. *et al.* Targeting neuronal and glial cell types with synthetic promoter AAVs in mice, non-human primates, and humans. *bioRxiv [Preprint]* doi.org/10.1101/434720 (2018).
46. Graybuck, L. T. *et al.* Prospective, brain-wide labeling of neuronal subclasses with enhancer-driven AAVs. *bioRxiv [Preprint]* doi.org/10.1101/525014 (2019).
47. Rad, R. *et al.* A conditional piggyBac transposition system for genetic screening in mice identifies oncogenic networks in pancreatic cancer. *Nat. Genet.* **47,** 47–56 (2014).
48. Kettlun, C., Galvan, D. L., George Jr, A. L., Kaja, A. & Wilson, M. H. Manipulating piggybac transposon chromosomal integration site selection in human cells. *Mol. Ther.* **19,** 1636–1644 (2011).
49. Beltramo, R. *et al*. Layer-specific excitatory circuits differentially control recurrent network dynamics in the neocortex. *Nat. Neurosci.* **16,** 227–234 (2013).
50. 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).
51. Bailey, T. L. *et al.* MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37,** 202–208 (2009).
52. Machanick, P. & Bailey, T. L. MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **27,** 1696–1697 (2011).
53. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107,** 21931–21936 (2010).
54. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** (2008).
55. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
56. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457,** 854–858 (2009).
57. Sakers, K. *et al.* Astrocytes locally translate transcripts in their peripheral processes. *Proc. Natl. Acad. Sci.* **114,** E3830–E3838 (2017).

Methods

Animals

All animal practices and procedures were approved by the Washington University in St. Louis Institutional Animal Care and Use Committee (IACUC) in accordance with National Institutes of Health (NIH) guidelines. Transgenic mouse strains used in this study include β-actin(ActB)::Cre (RRID:IMSR_JAX:019199), Synapsin 1 (Syn1)::Cre (RRID:IMSR_JAX:003966), glial fibrillary acidic protein (GFAP)::Cre (RRID:IMSR_JAX:024098), retinol binding protein 4 (Rbp4)::Cre (RRID:MMRRC_037128-UCD)[49], and TRAP reporter mice CAG::EGFP/Rpl10a (RRID:IMSR_JAX:022367). All mice were bred to the C57BL/6J background. 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. 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).

Cell culture and transfections

HEK293T cells used in this study were cultured in 1X DMEM with 10% fetal bovine serum (FBS) and grown under standard conditions (37°C; 5% CO_2). Plasmid transfections were carried out with Fugene® 6 Transfection Reagent (Promega, Madison, WI, USA) with the manufacturer's protocol.

Immunofluorescence and imaging

10μm-thick (for co-localization studies) or 40μm-thick (for imaging of AAV::BrokenHeart) 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 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.

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 all *in vivo* experiments, transposase and donor transposon viruses were mixed equally by volume and then intracranially co-injected into the cortex of postnatal day 1 (P1) mouse pups, 3 sites per hemisphere, 1ul viral mix per site. Viral titers (viral genomes per milliliter) were as follows:

AAV::hypPB	$1.1x10^{13}$ vg/ml
AAV::hypPB FLEX	$8.0x10^{12}$ - $1.0x10^{13}$ vg/ml
AAV::SP1(621C)-hypPB FLEX	$1.0x10^{13}$ vg/ml
AAV::SRT	$1.6x10^{13}$ - $2.2x10^{13}$ vg/ml
AAV::BrokenHeart FLEX	1.4×10^{13} vg/ml

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[16]. 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 or NextSeq 500 platforms (Illumina, San Diego, CA, USA).

BrokenHeart libraries were prepared from cortical DNA samples, as previously described[13]. 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[50] (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.

Significant insertion peak calling and motif discovery

Significantly enriched insertion peaks were identified via a count-based statistical comparison as previously described[16]. 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 3, Fig 4S,** and **Fig 5**, 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 4** for astrocyte- or neuron-specific Brd4 peaks, **Fig 5** for Brd4 peaks specific to Rbp4(+) layer V projection neurons, and **Fig 6-7** 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.

TF motifs were identified with MEME-ChIP motif discovery software[51,52] with -zoops -meme-minw 6 -ccut 250.

Defining enhancers and SEs

Since H3K27ac is a known marker of active enhancers[4,53] and SEs[7,28], we utilized a published P14 mouse cortex H3K27ac ChIP-Seq dataset[29] to define cortical SEs. As previously described[7,28], we used the rank ordering of super-enhancers (ROSE) pipeline and the model-based analysis for ChIP-Seq (MACS) version 1.4.1 peak finding algorithm[54] with a p-value enrichment threshold of 10^{-9} to define enhancers and SEs.

We then used the BEDtools suite[55] to compare the coincidence of enhancers and SEs with our Brd4-directed calling cards insertion peaks (**Fig 3** and **Fig 5**).

Additionally, for qualitative measures of histone modification enrichment at calling cards Brd4 insertion peaks (**Fig 3**), we used publicly available P0 mouse forebrain ChIP-Seq datasets from ENCODE[55]; specifically, H3K27ac (ENCSR094TTT), H3K4me1 (ENCSR465PLB), and H3K27me3 (ENCSR070MOK).

Analysis of enhancer- and promoter-associated gene expression

Gene expression has been shown to be preferentially regulated by proximal enhancer elements[4,53,56]. 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. In our analyses of cell type-specific expression of genes near cell type-enriched Brd4 calling cards peaks (**Fig 4** and **Fig 5**), 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, 2) genes near 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[31] (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 6**, 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/P10 promoter insertions to RNA expression in **Fig 7B**, we utilized a previously published RNA-seq dataset[44] 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.

A similar analysis was carried out in **Fig 7C**, wherein SP1 peaks enriched over unfused hypPB were defined with our differential peak caller (778 peaks at $p<10^{-30}$) and filtered as follows: 1) included only those intersecting promoters from the unique promoter/gene list, 2) excluded duplicate genes intersecting multiple peaks, leaving 637 genes for subsequent analyses.

Translating Ribosome Affinity Purification (TRAP)

TRAP was used to isolate cell type-specific, ribosome-bound RNAs from Rbp4(+) layer V pyramidal neurons in **Fig 5** as previously described[33,57]. In brief, Cre positive and negative animals were processed by harvesting brains at P21 and splitting down the midline, creating two “barcodes” per animal. Tissue samples were homogenized according to modified protocols previously described[33]. Immunoprecipitation (IP) of transcripts was achieved via biotinylated-protein L (ThermoFisher Scientific 29997)-coated streptavidin magnetic beads (ThermoFisher Scientific 65602) that were coupled with two custom monoclonal anti-EGFP antibodies (Htz-19f7, Htz-19c8)[33]. Input samples were collected prior to IP and remaining homogenate was incubated on an inverted mixer for 4 hours at 4°C. Beads were then washed several times with buffers and RNA purification was performed on eluted beads (Zymo RNA Clean & Concentrator -5, R1013, Zymo Research, Irvine, CA, USA). RNA quality and concentration were determined via Agilent BioAnalyzer prior to sequencing.

Statistical analyses

Statistical tests were done with GraphPad Prism v 8.0 and are detailed in figure legends.

Figure legends

Figure 1. Constructs and SRT library preparation.

(A) Constructs used in AAV calling cards system, including Cre-dependent ("FLEX") and independent hypPB and TF-hypPB as well as donor transposon expression plasmids. (B) Transfection of BrokenHeart transposons (Cre-independent version), 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. (C) Schematic of Self-Reporting Transposon (SRT) library preparation protocol. HypPB inserts SRTs near TF binding sites, which then transcribe and report their genomic locations via RNA. RNA is reverse transcribed and PCR amplified, and Illumina adapters are added for sequencing.

Figure 2. Co-AAV9 intracranial injection efficiently delivers calling cards to the cortex.

(A) Experimental paradigm. Arrows represent approximate AAV injection sites. (B,C) Co-immunofluorescence showing hypPB expression in neurons and astrocytes. (B) Representative images display co-localization of hypPB with neuronal (NeuN) and astrocyte (GFAP) markers. Insets show examples of hypPB-positive astrocytes. (C) Majority of hypPB-positive cells transduced with AAV9 are NeuN-positive neurons and GFAP-positive astrocytes (n=1005 myc(+) cells, counted across cortical image fields from 5 independent mice).

Figure 3. Unfused hypPB-directed calling cards insertions identify enhancers and super enhancers in the brain.

(A) Normalized insertion depth in two littermate C57BL/6J mice (Rep1 and Rep2) at 7031 significantly-enriched insertion peaks ($p<10^{-30}$) displaying high correlation between replicates (R=0.994). (B-E) Unfused hypPB-directed insertions are highly enriched for the active enhancer marks H3K27ac and H3K4me1 and depleted for suppresive mark H3K27me3. Representative image (B), heatmaps, and enrichment plots (C-E) of H3K27ac, H3K4me1, and H3K27me3 density at 7031 significantly-enriched insertion peaks in two littermate mice. In (B), top track of each insertion replicate displays unique insertions, where each circle = 1 unique insertion and y-axis represents number of reads supporting each insertion on a log10 scale, and bottom track displays normalized local insertion density across the genome (insertions per million per kB). Y-axis of ChIP-seq data represents read depth with smoothing filter applied. Heatmaps and enrichment plots are centered on insertion peaks and extend 10kB in either direction. Relative enrichment quantifications displayed in log2(fold-change over ChIP-seq input). (F,G) Percentage of 7031 significantly-enriched insertion peaks with at least 1 basepair (bp) intersection with a H3K27ac-marked enhancer or super enhancer. Gray bar represents intersections after randomizing genomic coordinates of insertion peaks. Chi^2 test with Yates correction: $p<0.0001$. (H,I) Percentage of H3K27ac-marked enhancers and super enhancers with at least 1 bp intersection with a significantly-enriched insertion peak. Chi^2 test with Yates correction: $p<0.0001$.

Figure 4. FLEX calling cards system generates cell type-specific TF profiles.

(A) Experimental design in Syn1::Cre and GFAP::Cre animals. (B) Examples of differentially enriched insertion peaks near genes preferentially expressed in neurons (right) or astrocytes (left). RNA-seq expression from Zhang et al., *Journal of Neuroscience*, 2014. (C) 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 signficance thresholds for defining differentially enriched insertion peaks. As significance threshold becomes more stringent, expression of nearby genes becomes more cell type-specific. (D) Astrocyte vs 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. (E,F) Quantifications of neuron and astrocyte specific expression of genes near GFAP::Cre (E) or Syn1::Cre (F) enriched insertion peaks displayed in (D), showing significant preferential expression in the expected cell type (Mann-Whitney U test: $p<0.0001$). (G,H) Graphical representation of cortical cell type enrichment based on gene sets near either (G) Syn1::Cre ($p<10^{-11}$; top 123 genes) or (H) 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 5. Combinatorial FLEX calling cards/TRAP allows for TF profiling in Rbp4::Cre(+) layer V pyramidal neurons.

(A) Experimental design of combinatorial FLEX calling cards and TRAP for cell type-specific TF profiling in Rbp4(+) neurons. (B) Representative differentially enriched peak in Rbp4::Cre mice over wild-type. (C) Characteristic "hockey stick" graph of all Rbp4(+) background-free insertion peaks sorted by insertion density. Top 25% most insertion-enriched peaks (highlighted in dark purple) display increased median peak length relative to bottom 75% (highlighted in light purple) and harbor >50% of total insertions. (D) Out of 65 differential insertion peaks in Rbp4::Cre over wild-type, 28 intersect a known H3K27ac super enhancer (SE).

Figure 6. Fusion of SP1 to hypPB in FLEX calling cards system records SP1 occupancy.

(A) Fusion of the promoter-binding TF SP1 to hypPB directs insertions to promoter-proximal TTAA sites. (B) Percentage of total insertions within 1000 basepairs (bp) of a transcription start site (TSS), displaying increased promoter-directed insertions upon SP1 fusion as compared to unfused hypPB (n = 3-4 mice per group; unpaired Student's t-test: $p<0.001$). Total SP1(621C)-hypPB insertions: 1,083,099. Total unfused hypPB insertions: 2,484,133. (C) Percentage of significant SP1 insertion peaks differentially enriched over unfused hypPB ($p<10^{-15}$; 1596 intersecting out of 2316 total peaks) intersecting promoter-proximal regions (1000bp on either side of TSS) compared to randomized peak coordinates (78/2316). Chi^2 test with Yates correction: $p<0.0001$. (D) Representative insertion peak displaying significantly increased insertion density near the TSS of the *Plrg1* gene. (E) Highest information content motif in the sequences flanking the center of significantly enriched SP1(621C)-hypPB insertion peaks ($p<10^{-15}$) displays the canonical SP1 binding motif

($p<10^{-42}$). (F-G) 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 (F), left y-axis represents number of promoter insertions normalized to 10^6 total insertions in the sample and right y-axis displays neuron-specific RNA expression from Zhang et al., *Journal of Neuroscience*, 2014. (G) displays strong correlation of SP1(621C)-hypPB promoter insertions with gene expression after subtraction of unfused hypPB insertions (R=0.96, p <0.0001).

Figure 7. Longitudinal SP1 profiling reports integrated, historical record of SP1 binding.
(A) Experimental paradigm. Genes expressed before P10 but not after (e.g. Gene X) display similar SP1 promoter insertion profiles in both cohorts. Genes expressed after P10 but not before (e.g. Gene Y) display increased SP1 promoter insertion density in the P28 cohort relative to the P10. (B,C) Correlation of SP1 promoter insertion ratios at P28 to P10 for each gene versus relative expression ratio at week4 (Wk4) to week 1 (Wk1) for (B) all expressed genes (R=0.681, p<0.0001) and (C) genes with SP1-enriched promoter peaks (R=0.673, p=0.0006). Genes were sorted by Wk4/Wk1 expression and binned by 100 or 25 gene bins. Dashed red line represents 1:1 ratio of promoter insertions between cohorts. In general, gene promoters gain insertions over time due to continued SP1 binding, even with decreased RNA expression by Wk4. RNA-seq from Lister et al., *Science*, 2013. (D) Examples of genes with P28-enriched SP1 promoter peaks and increased Wk4/Wk1 RNA expression.

Figure 1

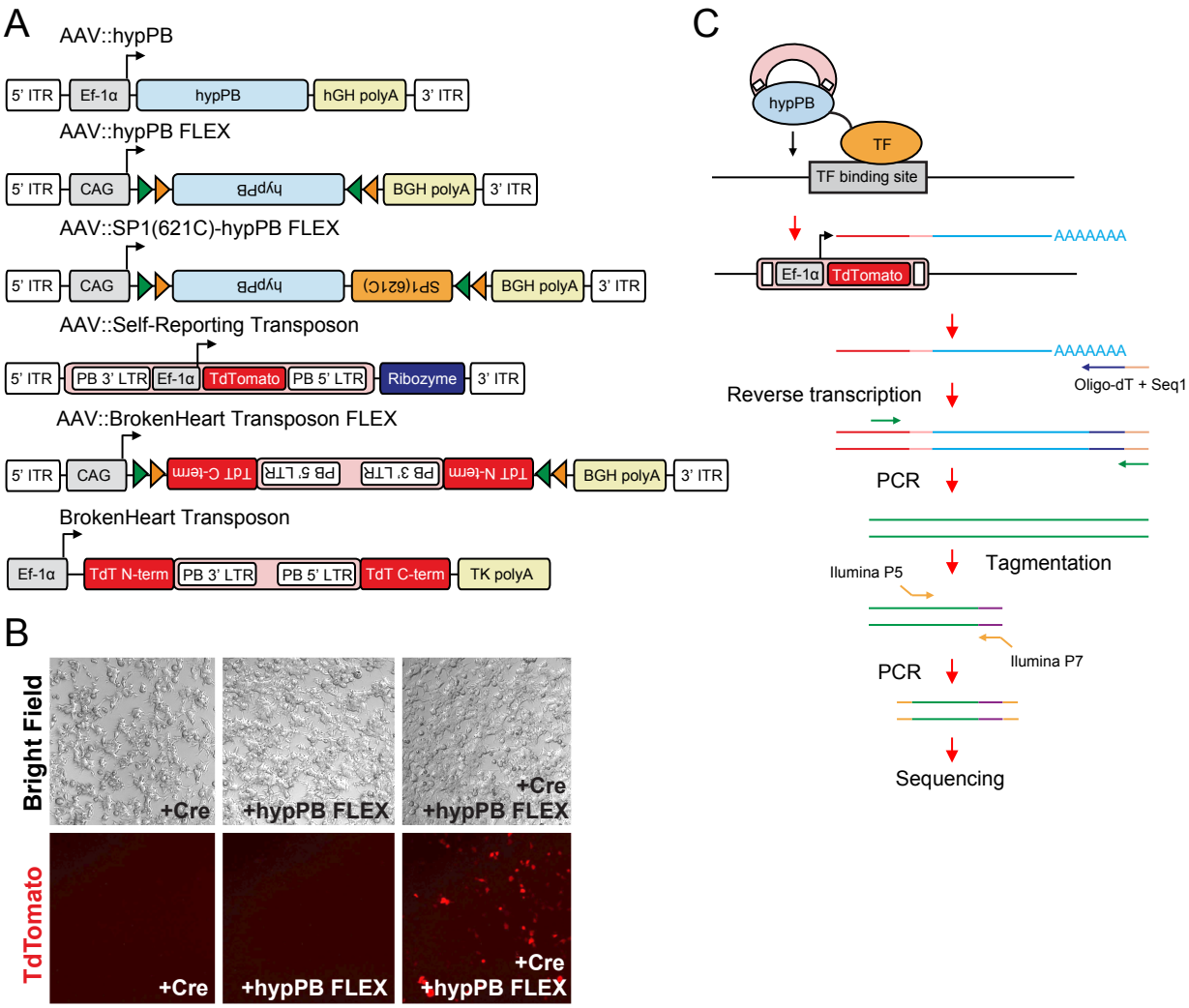

Figure 2

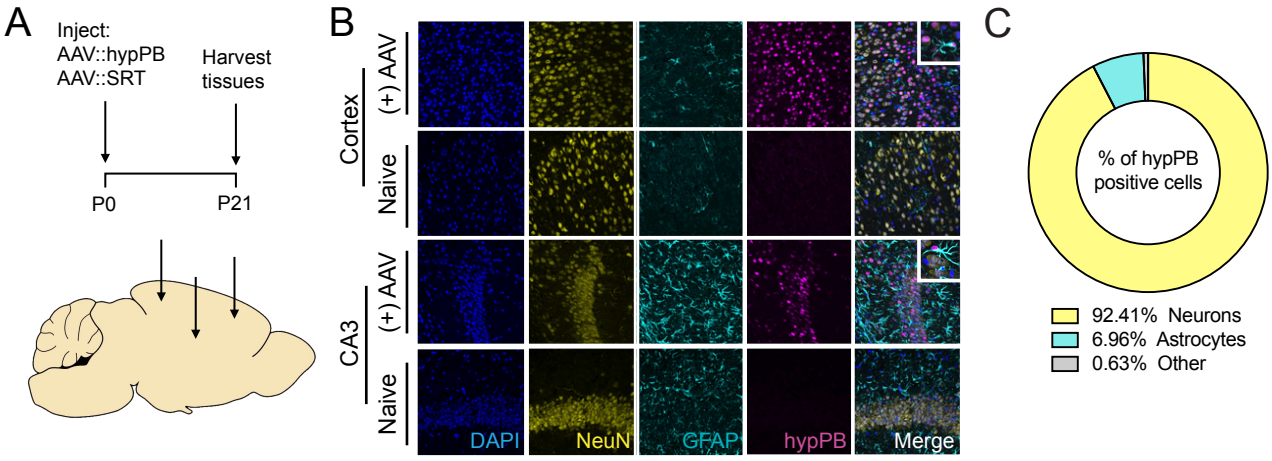

Figure 3

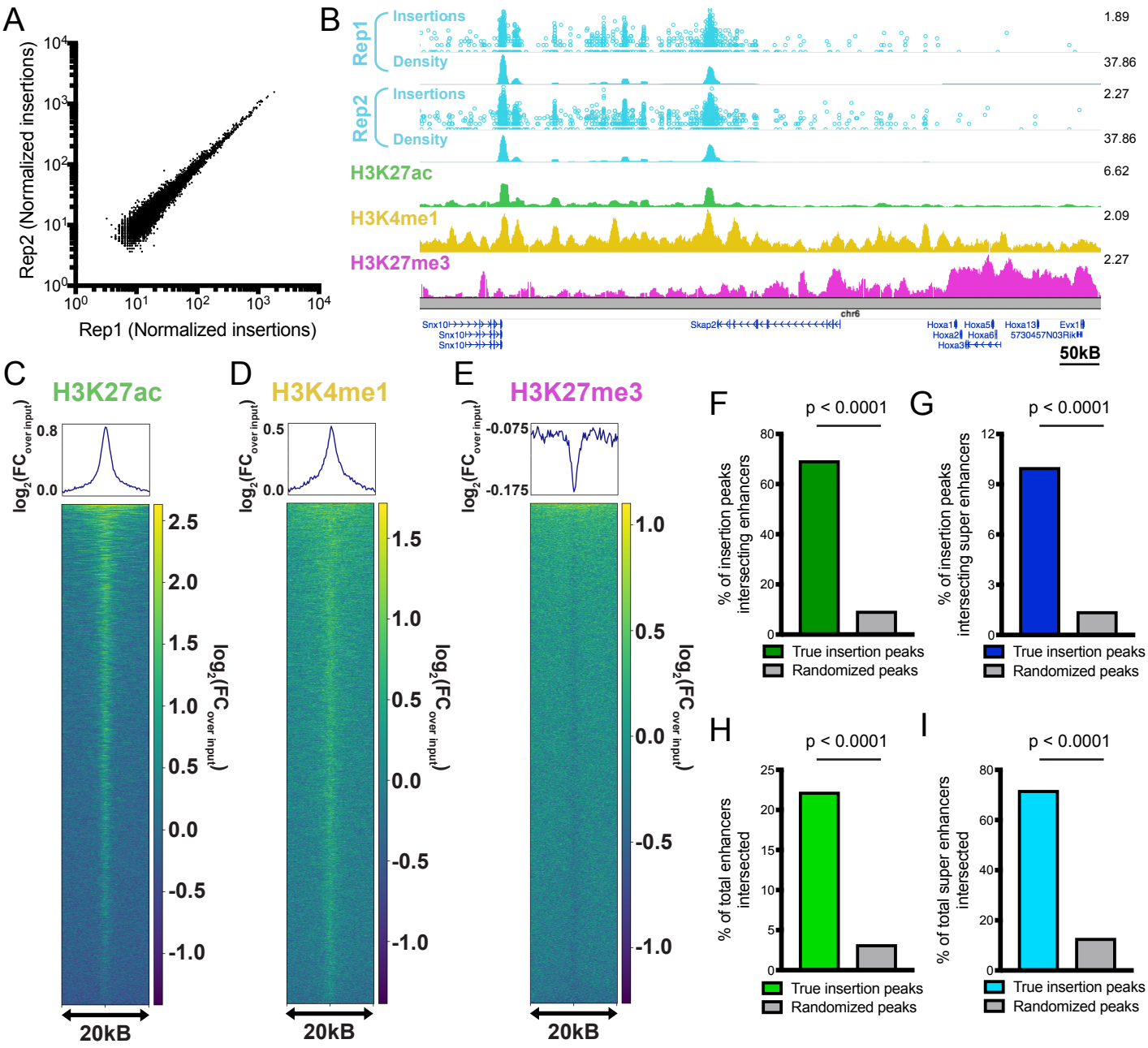

Figure 4

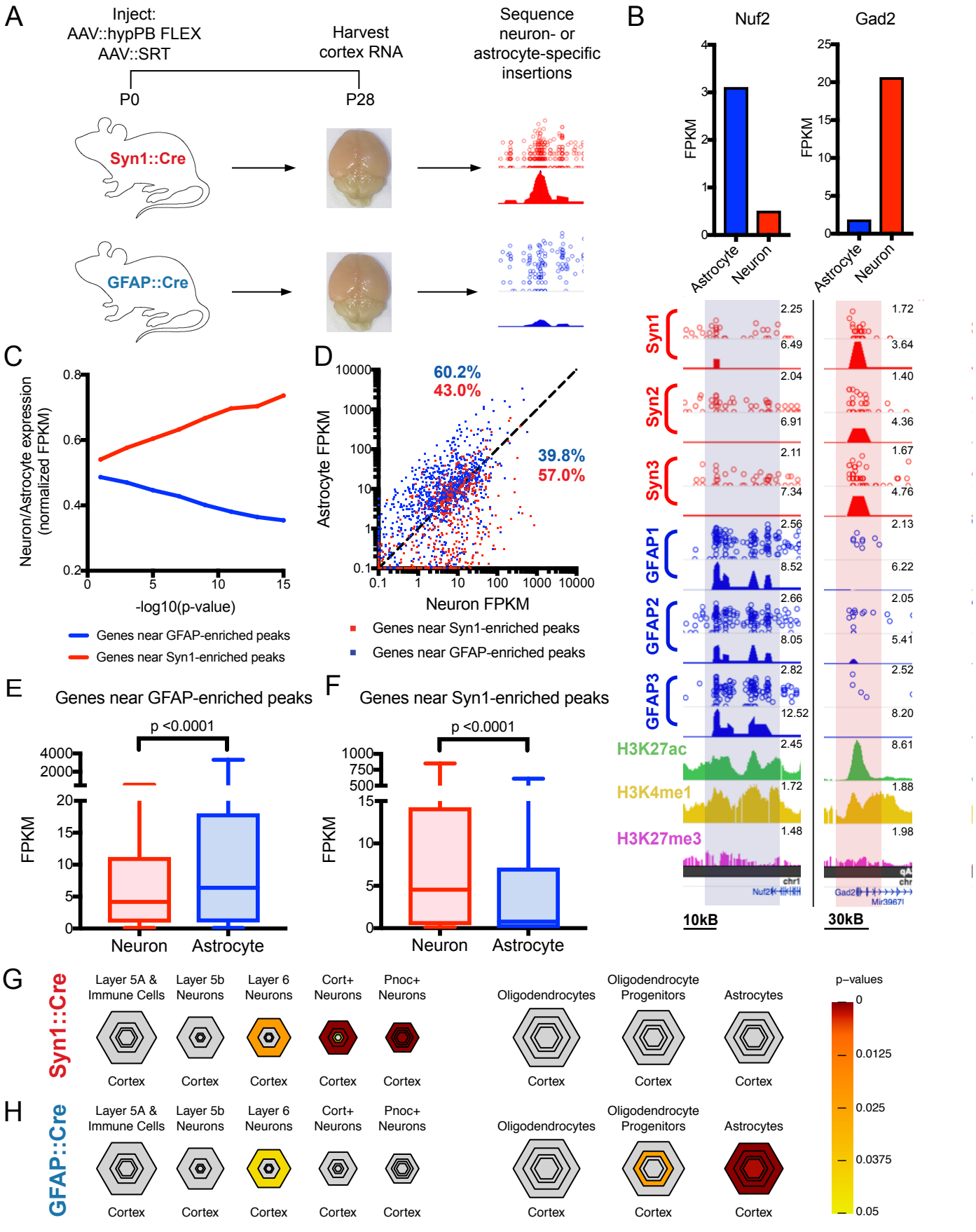

Figure 5

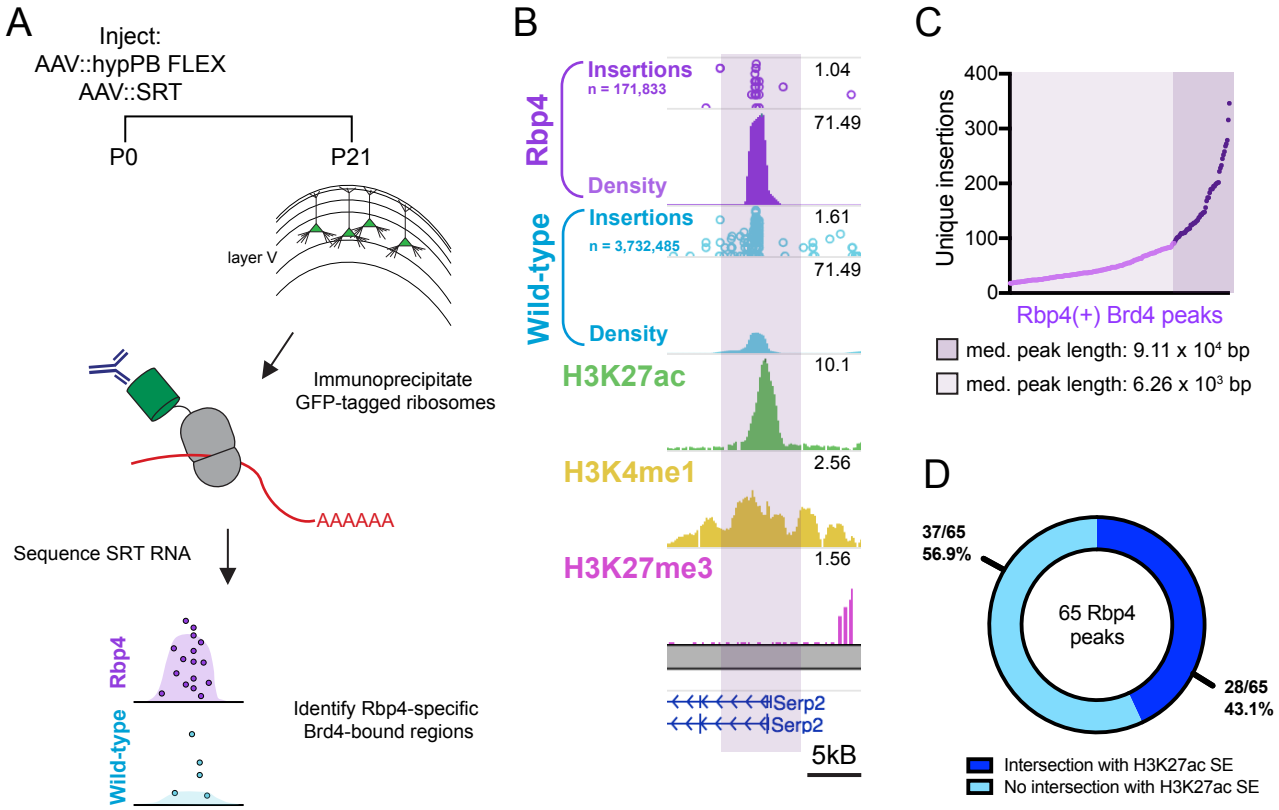

Figure 6

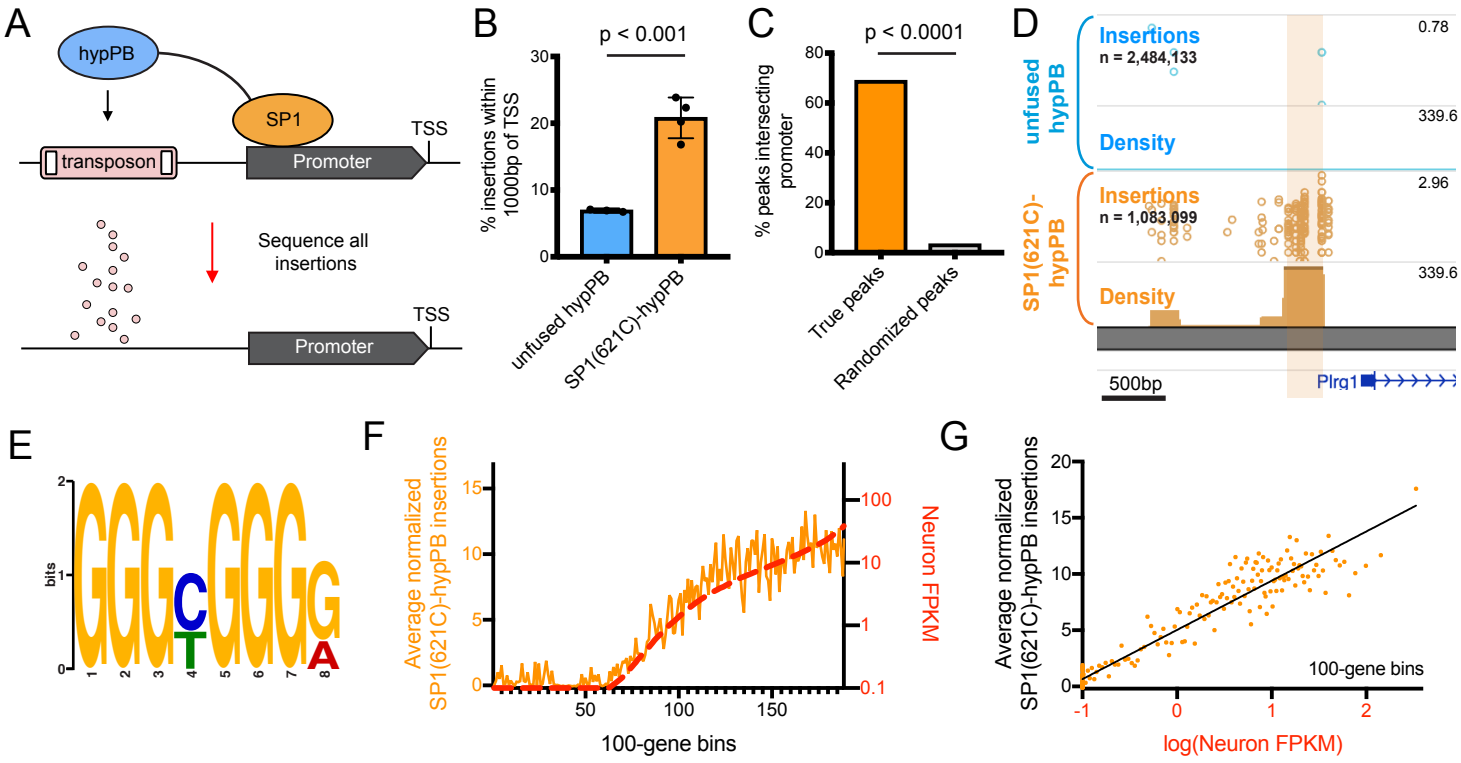

Figure 7