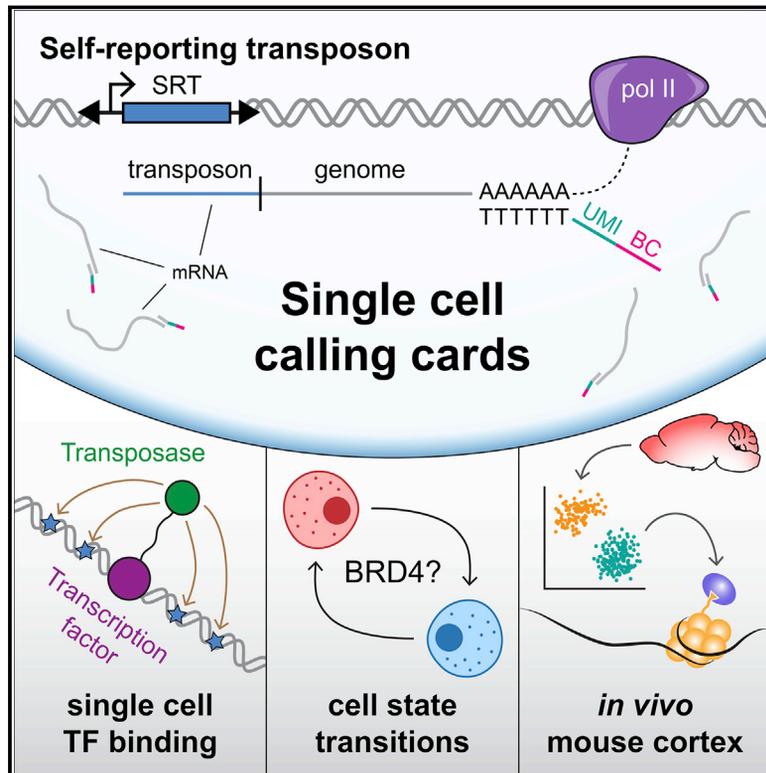


# Self-Reporting Transposons Enable Simultaneous Readout of Gene Expression and Transcription Factor Binding in Single Cells

## Graphical Abstract



## Authors

Arnav Moudgil, Michael N. Wilkinson, Xuhua Chen, ..., Samantha A. Morris, Joseph D. Dougherty, Robi D. Mitra

## Correspondence

rmitra@wustl.edu

## In Brief

Moudgil et al. present a single-cell method for simultaneously capturing gene expression and transcription factor binding site data from the same cells, first in cell lines and then in the mouse brain.

## Highlights

- Novel genetic element can be mapped from mRNA and directed by DNA-binding proteins
- Combined gene expression and transcription factor binding data from single cells
- Mapped multiple transcription factors in several cell lines and the mouse cortex
- Discovered bromodomain-dependent cell-state transitions in leukemic cells

Resource

# Self-Reporting Transposons Enable Simultaneous Readout of Gene Expression and Transcription Factor Binding in Single Cells

Arnab Moudgil,<sup>1,2,3</sup> Michael N. Wilkinson,<sup>1,2</sup> Xuhua Chen,<sup>1,2</sup> June He,<sup>1,2</sup> Alexander J. Cammack,<sup>4</sup> Michael J. Vasek,<sup>1,5</sup> Tomás Lagunas, Jr.,<sup>1,5</sup> Zongtai Qi,<sup>1,2</sup> Matthew A. Lalli,<sup>1</sup> Chuner Guo,<sup>1,3,6</sup> Samantha A. Morris,<sup>1,6,7</sup> Joseph D. Dougherty,<sup>1,5</sup> and Robi D. Mitra<sup>1,2,8,\*</sup>

<sup>1</sup>Department of Genetics, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA

<sup>2</sup>Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA

<sup>3</sup>Medical Scientist Training Program, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA

<sup>4</sup>Department of Neurology, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA

<sup>5</sup>Department of Psychiatry, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA

<sup>6</sup>Department of Developmental Biology, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA

<sup>7</sup>Center of Regenerative Medicine, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA

<sup>8</sup>Lead Contact

\*Correspondence: [rmitra@wustl.edu](mailto:rmitra@wustl.edu)

<https://doi.org/10.1016/j.cell.2020.06.037>

## SUMMARY

Cellular heterogeneity confounds *in situ* assays of transcription factor (TF) binding. Single-cell RNA sequencing (scRNA-seq) deconvolves cell types from gene expression, but no technology links cell identity to TF binding sites (TFBS) in those cell types. We present self-reporting transposons (SRTs) and use them in single-cell calling cards (scCC), a novel assay for simultaneously measuring gene expression and mapping TFBS in single cells. The genomic locations of SRTs are recovered from mRNA, and SRTs deposited by exogenous, TF-transposase fusions can be used to map TFBS. We then present scCC, which map SRTs from scRNA-seq libraries, simultaneously identifying cell types and TFBS in those same cells. We benchmark multiple TFs with this technique. Next, we use scCC to discover BRD4-mediated cell-state transitions in K562 cells. Finally, we map BRD4 binding sites in the mouse cortex at single-cell resolution, establishing a new method for studying TF biology *in situ*.

## INTRODUCTION

Transcription factors (TFs) regulate the gene expression patterns that specify cell state (Gurdon, 2016; Hafler et al., 2012; Mizuguchi et al., 2001; Zhu et al., 2012). They are central to a number of critical developmental processes including the maintenance of pluripotency (Liu et al., 2008; Takahashi and Yamanaka, 2006), fate choice (Mizuguchi et al., 2001; Zhu et al., 2012), and embryogenesis (Fogarty et al., 2017). Perturbing TF activity can disrupt cellular development, homeostasis, or function, resulting in altered morphology (Gonen et al., 2018; Kvon et al., 2016), cellular transdifferentiation (Davis et al., 1987), or increased susceptibility to disease (Lee and Young, 2013). A better understanding of TF binding during development and homeostasis would provide insights into how cellular diversity arises and is maintained under normal and abnormal biological conditions.

Single-cell RNA sequencing (scRNA-seq) has emerged as the *de facto* approach for characterizing cellular diversity in

complex tissues and organisms (Campbell et al., 2017; Cao et al., 2017; Fincher et al., 2018; Han et al., 2018; Karaikos et al., 2017; Zeisel et al., 2015). Recently, multi-modal scRNA-seq technologies have emerged (Angermueller et al., 2016; Cao et al., 2018; Clark et al., 2018; Dey et al., 2015; Macaulay et al., 2015; Peterson et al., 2017; Stoeckius et al., 2017) linking transcriptional information to other genomic assays. These methods address the fact that, while scRNA-seq can describe the current state of a biological system, it alone cannot explain how that state arose. A notable lacuna in the single-cell repertoire is a method for jointly assaying transcriptome and TF binding. Such a method would lead to the genome-wide identification of TF binding sites across multiple cell types in complex tissues. Chromatin immunoprecipitation sequencing (ChIP-seq) is the most popular approach to studying TF binding (Johnson et al., 2007), and, while a number of antibody-based single-cell methods to detect DNA-protein contacts have been reported (Ai et al., 2019; Carter et al., 2019; Grosselin et al., 2019; Hainer et al., 2019; Harada et al.,

2019; Kaya-Okur et al., 2019; Rotem et al., 2015; Wang et al., 2019), these techniques have generally mapped highly abundant proteins, such as modified histones and CTCF. DamID can recover TF binding sites by detecting nearby exogenously methylated adenines (Greil et al., 2006; Vogel et al., 2007), but in single cells it has only been used to study lamina-associated domains (Kind et al., 2013, 2015; Rooijers et al., 2019). A combined single-cell assay of DamID and transcriptome (scDam&T-seq) has been described (Rooijers et al., 2019) but is a plate-based assay that limits throughput. None of the other single-cell techniques that measure DNA-protein interactions simultaneously capture mRNA, restricting their use to predetermined cell types. Single-cell assays for transposase-accessible chromatin using sequencing (ATAC-seq; Buenrostro et al., 2015; Cao et al., 2018) could be used to identify nucleosome-free regions that may be bound by TFs, though they rely on motif inference to identify potential DNA binding proteins. These assays do not directly measure TF occupancy nor can they be used to study transcriptional regulators that bind DNA indirectly or non-specifically, such as chromatin remodelers.

We have previously developed transposon calling cards to assay TF binding (Wang et al., 2007, 2011, 2012a). This system relies on two components: a fusion between a TF and a transposase and a transposon carrying a reporter gene. The fusion transposase deposits transposons near TF binding sites, which are subsequently amplified from genomic DNA and sequenced. Thus, the redirected transposase leaves “calling cards” at the genomic locations it has visited, which can be identified later in time. The result is a genome-wide assay of all binding sites for that particular TF. In mammalian cells, we have heterologously expressed the *piggyBac* transposase (Ding et al., 2005) fused to the TF SP1 and shown that the resulting pattern of insertions reflects SP1’s binding preferences (Wang et al., 2012a). However, this method was only feasible in bulk preparations of thousands of cells.

Here, we present single-cell calling cards (scCC), an extension of transposon calling cards that simultaneously profiles mRNA content and TF binding at single-cell resolution. The key component of our work is the self-reporting transposon (SRT), a novel element whose genomic location can be mapped from mRNA. We show that the RNA-based calling card method is more efficient than our standard DNA-based protocol and can be used to map TF binding sites with a directed transposase. We also demonstrate that the unfused *piggyBac* transposase, through its native affinity for the bromodomain TF BRD4, can be used to identify BRD4-bound super-enhancers (SEs). We then present the scCC method, which allows cell-type-specific mapping of SRTs from scRNA-seq libraries. Thus, in one experiment, we can cluster cells by transcriptional identity and identify TF binding sites within those cell types. We highlight the range of this technology using a breadth of TFs in a variety of cell lines. We then use scCC to discover bromodomain-dependent cell-state dynamics in K562 cells. Finally, we identify cell-type-specific BRD4 binding sites *in vivo* in the postnatal mouse cortex. These results demonstrate that scCC could be a broadly applicable tool to study specific TF binding interactions across multiple cell types within heterogeneous systems.

## RESULTS

### SRTs Can Be Mapped from mRNA Instead of Genomic DNA

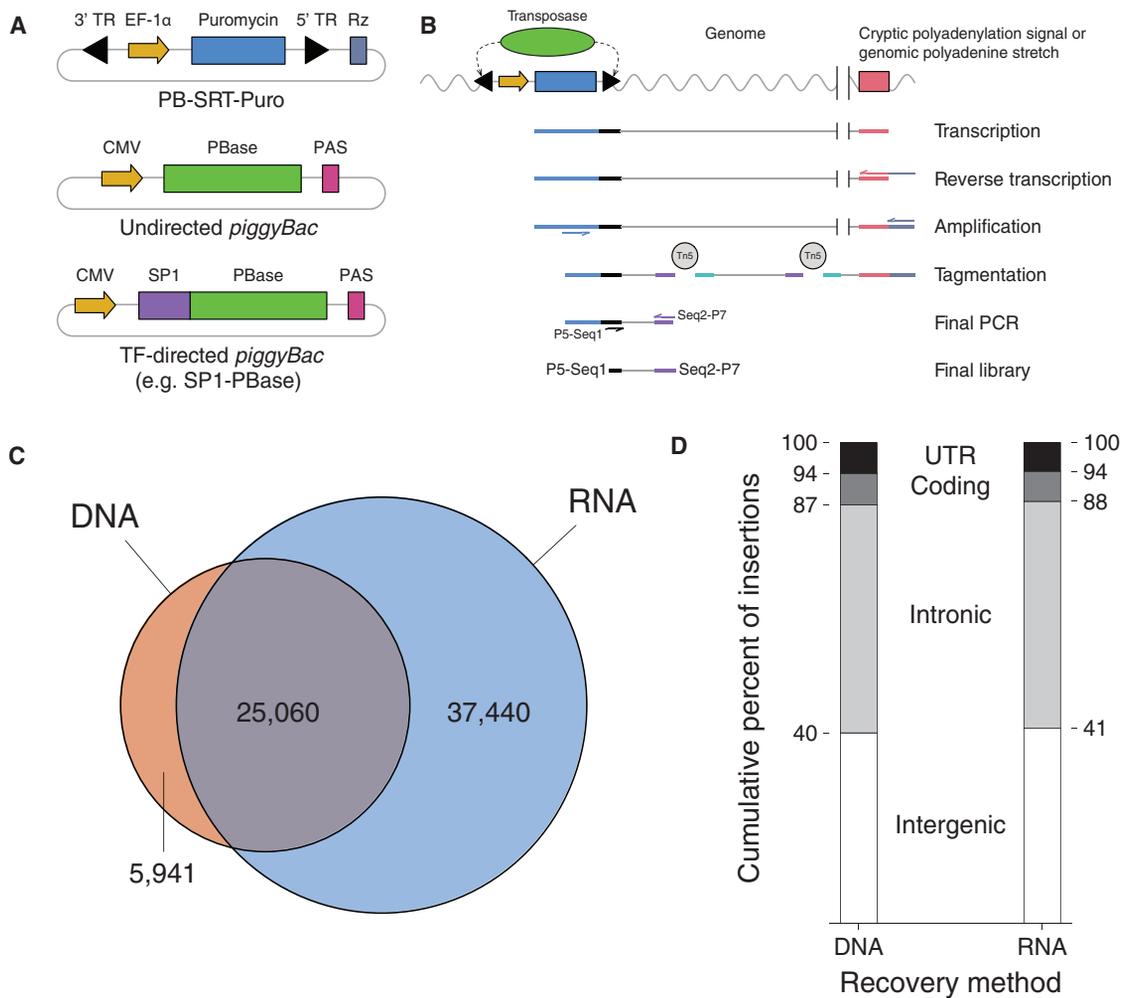
To combine scRNA-seq with calling cards, we first developed a transposon whose genomic location could be determined from mRNA. We created a *piggyBac* SRT by removing the polyadenylation signal (PAS) downstream of the reporter gene (Figure 1A; Methods S1) in the transposon. RNA polymerase II (Pol II) transcribes the SRT reporter and continues through the terminal repeat (TR) into the flanking genomic sequence. Thus, SRTs “self-report” their locations through the unique genomic sequence found in the 3′ untranslated regions (UTRs) of the reporter gene transcripts. While previously published gene- or enhancer-trap transposons (Cadiñanos and Bradley, 2007) could also encode local positional information in RNA, they are resolution-limited to the nearest gene or enhancer, respectively. In contrast, SRT-derived transcripts contain the transposon-genome junction, so insertions can be mapped with base-pair precision.

SRTs are mapped following reverse transcription (RT) and PCR amplification of self-reporting transcripts. These transcripts contain stretches of adenines that are derived from either cryptic PASs or templated polyadenine tracts in genomic DNA downstream of the SRT insertion site (Figure 1B). We then use a modified tagmentation protocol to enrich for the transposon-genome junction (STAR Methods). We confirmed SRTs generate reproducible libraries, require a functional transposase, and can be recovered from virtually any chromatin state (Methods S1).

To compare how the new RNA-based approach fares against our standard DNA-based method (Wang et al., 2012a), we tested both protocols on the same population of cells. Our DNA-based library yielded 31,001 insertions, while the RNA-based protocol recovered 62,500 insertions (Table S1). Importantly, 80% of the insertions found by DNA calling cards were also recovered in the RNA-based library (25,060 insertions; Figure 1C). Thus, at the level of individual transpositions, RNA-based mapping is highly sensitive. Moreover, the RNA protocol recovered a further 37,440 insertions that were not found in the DNA-based library. We analyzed the distribution of insertions by both genetic annotation (Figure 1D) and chromatin state (Methods S1) and found no appreciable differences in either case between the DNA and RNA libraries. Finally, we also confirmed that SRTs could still be used to study TF binding using established TF-*piggyBac* constructs and quantified the redirectability of these fusions (Methods S1). Thus, RNA-based recovery of transposons appears to be unbiased with respect to our established, DNA-based protocol.

### Clustering of Undirected *piggyBac* Insertions Identifies BRD4-Bound SEs

Previous studies have shown that undirected *piggyBac* preferentially inserts transposons near SEs (Yoshida et al., 2017), unique regulatory elements involved in regulating cell identity (Hnisz et al., 2013). SEs are enriched for the histone modification H3K27ac as well as Pol II and transcriptional coactivators like the mediator element MED1 and the bromodomain protein



**Figure 1. Self-Reporting Transposons Are Mapped More Efficiently from RNA Compared to DNA**

(A) Schematics of a self-reporting *piggyBac* transposon with puromycin reporter gene (PB-SRT-Puro) and undirected (PBase) and SP1-directed (SP1-PBase) *piggyBac* transposases.

(B) Molecular workflow for mapping SRTs from bulk RNA libraries.

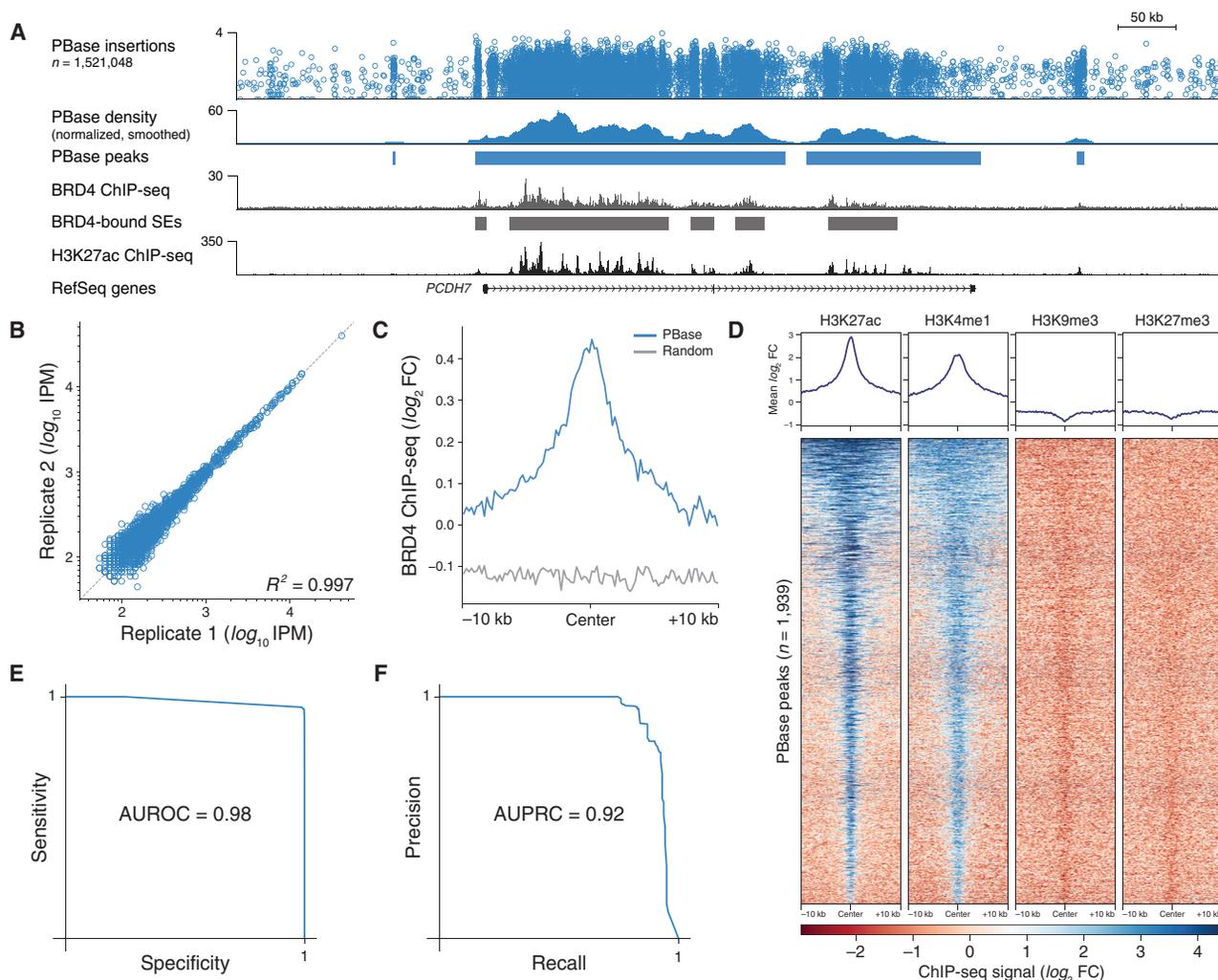
(C) Overlap of SRTs recovered by DNA- or RNA-based protocols in HCT-116 cells.

(D) Distribution of insertions with respect to genetic annotation between SRT libraries prepared from either DNA or RNA. TR, terminal repeat; Puro, puromycin; PAS, polyadenylation signal.

BRD4 (Hnisz et al., 2013; Lovén et al., 2013; Whyte et al., 2013). *piggyBac* has a strong biophysical affinity for BRD4, as these proteins can be co-immunoprecipitated (Gogol-Döring et al., 2016). Thus, we hypothesized that, given the millions of insertions recoverable by SRTs (Table S1), we would be able to identify BRD4-bound SEs simply from the localization of undirected *piggyBac* transpositions.

In HCT-116 cells, undirected *piggyBac* showed non-uniform densities of insertions at BRD4-bound loci (Figure 2A; for guidance on interpreting a calling card track, see STAR Methods). At statistically significant peaks of *piggyBac* calling cards, *piggyBac* showed high reproducibility of normalized insertions between biological replicates ( $R^2 > 0.99$ ; Figure 2B). We calculated the mean BRD4 enrichment, as assayed by ChIP-seq (McClelland et al., 2016), over all *piggyBac* peaks, which

showed significantly increased BRD4 signal compared to a permuted control set (Figure 2C; Kolmogorov-Smirnov [KS] test  $p < 10^{-9}$ ). Maximum BRD4 ChIP-seq signal was observed at calling card peak centers and decreased symmetrically in both directions. Moreover, *piggyBac* peaks showed striking overlap with ChIP-seq profiles for several histone modifications (Sloan et al., 2016; ENCODE Project Consortium, 2012), in particular, an enrichment for H3K27 acetylation (Figure 2D). Since bromodomains bind acetylated histones, this observation further supports the notion that undirected *piggyBac* insertions can be used to map BRD4 binding. Peaks were also enriched in H3K4me1, another canonical enhancer mark, and depleted for H3K9me3 and H3K27me3, modifications associated with heterochromatin (Lawrence et al., 2016). In all, *piggyBac* insertion density is highly correlated with BRD4



**Figure 2. Undirected *piggyBac* SRTs Mark BRD4-Bound Super-enhancers**

(A) Browser view of an undirected PBBase insertions in HCT-116 cells at a SE alongside BRD4 and H3K27ac ChIP-seq data.

(B) Reproducibility of normalized insertions at PBBase peaks.

(C) Mean BRD4 ChIP-seq signal at PBBase peaks compared to permuted control set.

(D) Heatmap of H3K27ac, H3K4me1, H3K9me3, and H3K27me3 ChIP-seq signal at PBBase peaks.

(E) Receiver-operator characteristic curve for SE detection using PBBase peaks.

(F) Precision-recall curve for SE detection using PBBase peaks.

See also [Figure S1](#). SE, super-enhancer; IPM, insertions per million mapped insertions; AUROC, area under receiver-operator curve; AUPRC, area under precision-recall curve; FC, fold change.

binding throughout the genome and that regions enriched for undirected *piggyBac* insertions share features common to enhancers.

We next assessed whether undirected *piggyBac* peaks can be used to identify BRD4-bound SEs. We constructed receiver-operator characteristic curves based on our ability to detect SEs from *piggyBac* (Figure 2E). The high area under the curve (0.98) indicates that we can robustly identify BRD4-bound SEs from *piggyBac* transpositions. Across a range of sensitivities, calling card peaks are highly specific and have high positive predictive value (AUPRC = 0.92; Figure 2F). These trends also hold true for the hyperactive *piggyBac* mutant (Methods S1).

Thus, undirected *piggyBac* transpositions can accurately assay BRD4-bound SEs.

We also investigated how similar *piggyBac* transposition is to that of Tn5, the transposase used in ATAC-seq (Buenrostro et al., 2013, 2015) to identify open chromatin. Since BRD4 and H3K27ac co-occur at accessible loci, it may be that undirected calling cards and ATAC-seq provide redundant information. If that were the case, we should be able to identify BRD4-bound SEs with high sensitivity from ATAC-seq data alone, much as we have shown for *piggyBac*. We called SEs using publicly available ATAC-seq data from HCT-116 cells (Ponnaluri et al., 2017) in the same manner that we did for BRD4 ChIP-seq. We found

almost no overlap between BRD4-bound SEs and these so-called SEs from ATAC-seq data (Figure S1A). Moreover, there are a small number (4.3%) of *piggyBac* peaks that are not found in accessible chromatin (Figure S1B), suggesting that there may be regulatory elements in closed chromatin that calling cards are better able to detect. Globally, over 20% of Tn5 insertions are directed to accessible sites, starkly higher than undirected *piggyBac* but comparable to TF-*piggyBac* fusions (Figure S1C). That *piggyBac*'s preference for targetting open chromatin can be markedly increased by a covalently linked TF highlights both *piggyBac*'s baseline insensitivity for accessible sites and the efficacy of TF redirection. Finally, we find that *piggyBac* peaks are an order of magnitude larger than ATAC-seq peaks and, as a result, capture more BRD4 binding (Figure S1D). We conclude that unfused *piggyBac* reflects BRD4's binding preferences whereas Tn5 reports on all accessible chromatin; as a result, undirected calling cards are not equivalent to ATAC-seq.

### scCC Enables Simultaneous Identification of Cell Type and Cell-Type-Specific BRD4 Binding Sites

We next sought to recover SRTs from scRNA-seq libraries, which would let us identify cell types from transcriptomic clustering and, using the same source material, simultaneously profile TF binding in those cell types. We adopted the 10x Chromium platform due to its high efficiency of cell and transcript capture as well as its ease of use (Zheng et al., 2017) but with a modified protocol (Methods S1). We split the first-strand synthesis product in two: one half is used to generate a scRNA-seq library, while the other half undergoes specific amplification for SRTs followed by circularization. The circularization step brings the cell barcode and unique molecular index (UMI), found at the 3' ends of each transcript, next to the transposon-genome junction. In this way, SRTs can be mapped and assigned to single cells using high-throughput short read sequencing (Figure 3A). After sequencing, the cell barcodes shared between both libraries are used to connect individual insertions to specific cell types. We call this protocol scCC.

We first validated scCC by performing a species-mixing experiment with human HCT-116 cells and mouse N2a cells transfected with hyperactive *piggyBac* (HyPBBase) and PB-SRT-Puro. The resulting scRNA-seq library showed strong species separation with an estimated multiplet rate of 3.2% (Figure S2A). We restricted our calling card analysis to those insertions whose cell barcodes were observed in the scRNA-seq library (Table S2). The distribution of insertions across these cells reflected a continuum from pure mouse to pure human (Figures S2B and S2C). Since intramolecular ligation in the circularization step or the subsequent PCR may introduce artifacts, such as the mis-assignment of a barcode from a mouse cell to an insertion site in a human cell, we required that a given insertion in a given cell must have at least two different UMIs associated with it. This filter greatly improved the number of pure mouse and human cells (Figure S2D), yielding clear species separation with an estimated multiplet rate of 7.9% (Figure 3B). Thus, scCC can accurately map SRT insertions in single cells.

We then asked whether scCC could discern cell-type-specific BRD4 binding. We transfected two human cell lines, HCT-116 and K562, with HyPBBase and PB-SRT-Puro and mixed them

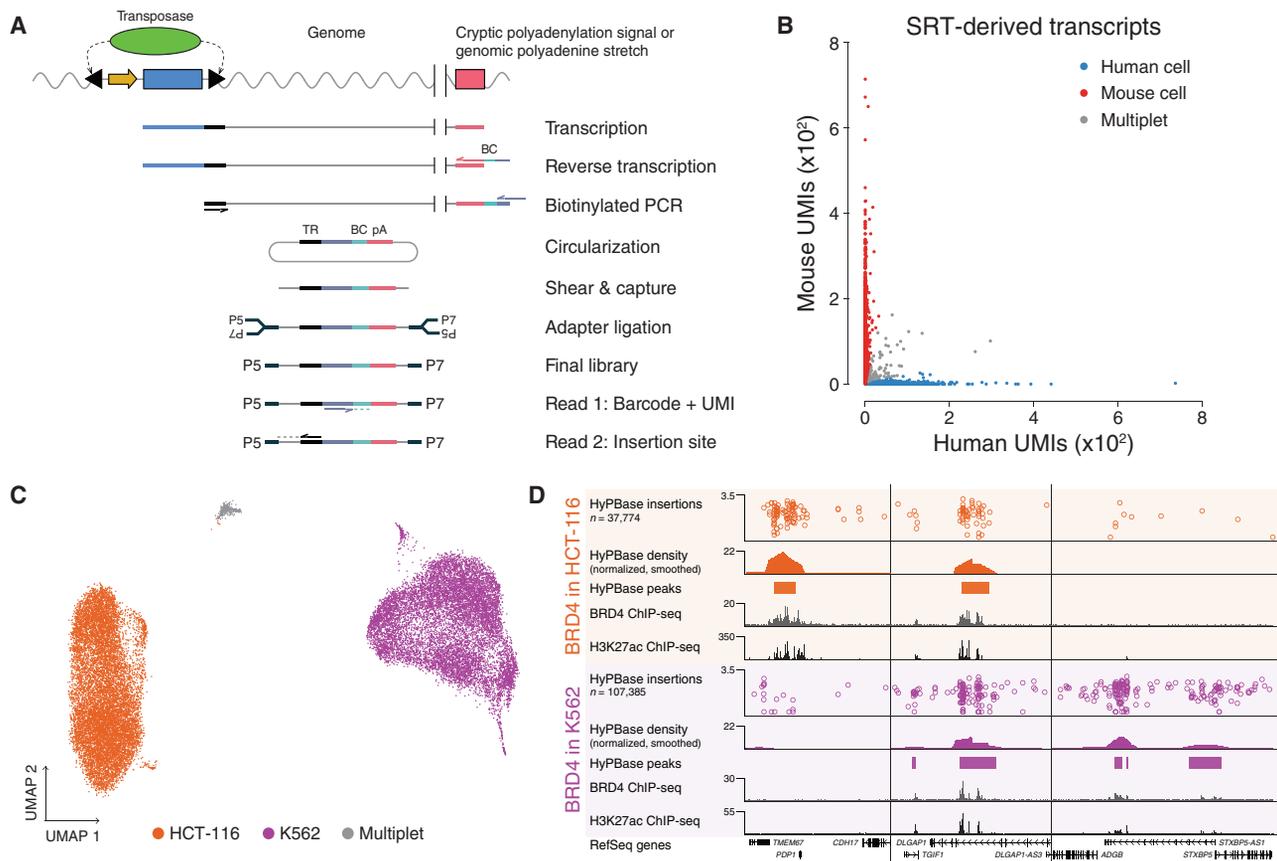
together. The resulting scRNA-seq libraries clearly distinguished these two cell types (Figures 3C and S3A). We prepared scCC libraries from these cells and used the cell barcodes from the HCT-116 and K562 clusters to assign insertions to the two different cell types (Table S2). The distribution of insertions per cell varied by cell type (Figure S3D) and was not explained by differences in total RNA content (Figures S3B and S3C). Over 93% and 96% of HCT-116 and K562 cells, respectively, had at least one insertion event (Table S2). Using the scCC insertion data, we called peaks and successfully identified BRD4-bound loci that were specific to HCT-116 cells, shared between HCT-116 and K562, and specific to K562 cells, respectively (Figure 3D). Both HCT-116 and K562 peaks showed statistically significant enrichment for BRD4 ChIP-seq signal over randomly permuted peaks (Figures S3E and S3F; KS test  $p < 10^{-9}$  in both instances). Furthermore, 57% of HCT-116 peaks and 81% of K562 peaks were specifically bound in their respective cell type. We estimated that with a p value cutoff of  $10^{-9}$ , our sensitivity for detecting BRD4-bound SEs would be approximately 60% (Methods S1), while the actual sensitivity at this level was 67%. Finally, at statistically significant peaks, normalized insertion counts were highly concordant between biological replicates in both cell types ( $R^2 = 0.91$  and  $0.94$ , respectively; Figures S3G and S3H). In all, these experiments demonstrate that scCC can be used to identify and deconvolve cell-type-specific BRD4 binding sites.

### scCC Identifies Binding Sites across a Spectrum of TFs and in a Variety of Cell Types

Our success mapping BRD4 SEs in single cells gave us confidence that we would also be able to map TF binding with scCC. We transfected HCT-116 and K562 cells with an SP1 fusion construct (SP1-HyPBBase) and performed scCC (Table S2). As was observed in bulk (Methods S1), SP1-HyPBBase-directed insertions recovered from single cells localized to SP1 binding sites in both HCT-116 and K562 cells (Figures 4A and 4E). In both cell lines, we observed significant enrichment of SP1 ChIP-seq signal at scCC peaks (Figures 4B and 4C and Figures 4F and 4G) and motif analysis identified the SP1 DNA binding motif (Figure 4D and Figure 4H) ( $p < 10^{-30}$  in each instance). SP1 is known to preferentially bind near transcription start sites (TSSs) and is also thought to play a role in demethylating CpG islands (Brandeis et al., 1994; Macleod et al., 1994; Philipson and Suske, 1999). Accordingly, we observed significant enrichments for insertions near TSSs, CpG islands, and unmethylated CpG islands in particular (Figures S4A and S4B; G test of independence  $p < 10^{-9}$  in each instance).

We next performed scCC in HepG2 cells with the pioneer factor FOXA2 (Table S2), which has been shown to be required for normal liver development and drives core transcriptional networks in cancer cells (Fournier et al., 2016; Lee et al., 2005). As with SP1, we observed a specific enrichment of insertions at FOXA2 binding sites (Figure 4I). Peaks called from scCC FOXA2 data were enriched in FOXA2 ChIP-seq signal (Figures 4J and 4K) and motif analysis was able to infer the core FOXA2 DNA binding motif (Figure 4L).

Last, we mapped the binding of BAP1 in the uveal melanoma cell line OCM-1A (Yen et al., 2018) using scCC (Table S2). Unlike SP1 and FOXA2, BAP1 does not bind DNA directly; instead, it is



**Figure 3. scCC Maps BRD4 Binding in Single Cells**

(A) Schematic of the scCC library preparation strategy from scrRNA-seq libraries.

(B) Barnyard plot of scCC on a mixture of human HCT-116 and mouse N2a cells.

(C) UMAP of scrRNA-seq of a mixture of human HCT-116 and K562 cells.

(D) Browser view of BRD4 peaks specific to HCT-116 and K562 cells deconvolved using scCC.

See also [Figures S2](#) and [S3](#). TR, terminal repeat; BC, barcode; pA, poly(A) sequence; UMI, unique molecular index.

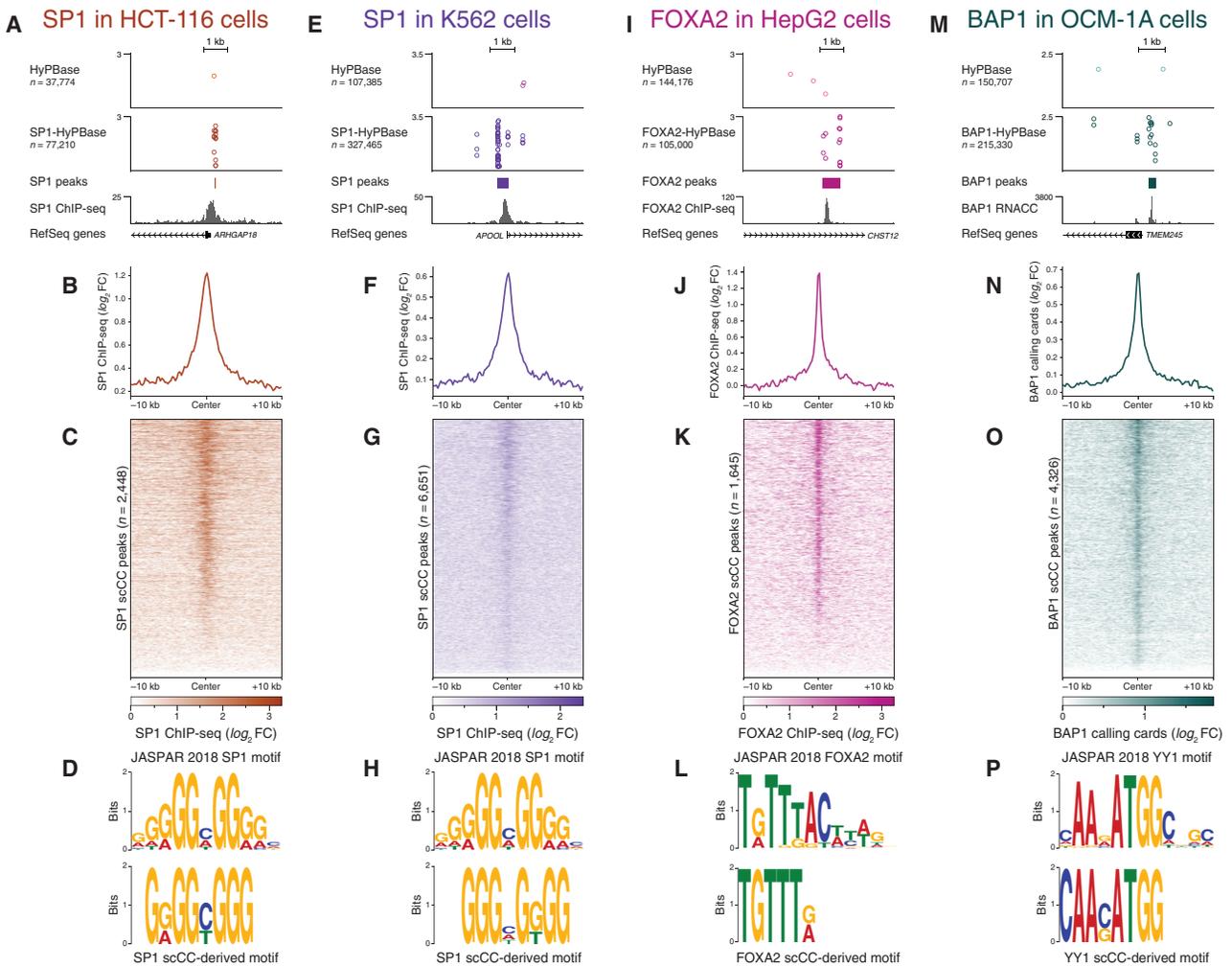
drawn to chromatin in a complex (Carbone et al., 2013; Yu et al., 2010) where it acts as a histone deubiquitinase. Despite this indirect interaction, we were able to resolve sharp BAP1-directed peaks (Figure 4M). These peaks showed high concordance with bulk RNA calling card data that we also generated in this system (Figures 4N and 4O; Table S1). Sequence analysis elicited the motif of YY1 (Figure 4P), a DNA binding TF and known member of the BAP1 complex (Yu et al., 2010). BAP1 is known to preferentially bind promoters (Dey et al., 2012), and, as such, we observed a significant enrichment for BAP1-directed insertions near TSSs (Figure S4C; G test of independence  $p < 10^{-9}$ ). While BAP1 is a member of the Polycomb repressive complex, there are conflicting reports as to its direct effects on gene expression (Campagne et al., 2019; Matatall et al., 2013; Yu et al., 2010). We cross-referenced our single-cell BAP1 peaks against published RNA-seq data in unperturbed and BAP1 knockdown OCM-1A cells (Yen et al., 2018). Genes where BAP1 is bound at the promoter, as opposed to in the gene body or at a nearby intergenic locus, are significantly more likely to have increased expression upon BAP1 knockdown (Figure S4D; Fisher's exact

test  $p < 10^{-9}$ ). This suggests that, in this model of uveal melanoma, promoter-bound BAP1 primarily acts as a repressor of gene expression.

Collectively, these results indicate that scCC can successfully map DNA-protein interactions for a range of TFs and in a variety of cell types. Furthermore, scCC showed high reproducibility in all four tested conditions ( $R^2$  between 0.71 and 0.95; Figures S4E–S4H). Although TF-piggyBac fusions have been previously reported to decrease transposase activity (Wu et al., 2006), our findings were more equivocal: some fusions showed less activity per cell than undirected HypBase, while others were more efficient (Figures S4I–S4L). Thus, there may be some variability in the number of recovered insertions depending on the TF and cell type of interest. Overall, however, the method is robust.

### scCC Reveal Bromodomain-Dependent Cell-State Dynamics in K562 Cells

SEs and BRD4 are thought to mark genes important for specifying cell identity, and, while the strongest evidence for this comes from comparisons between organ systems or between

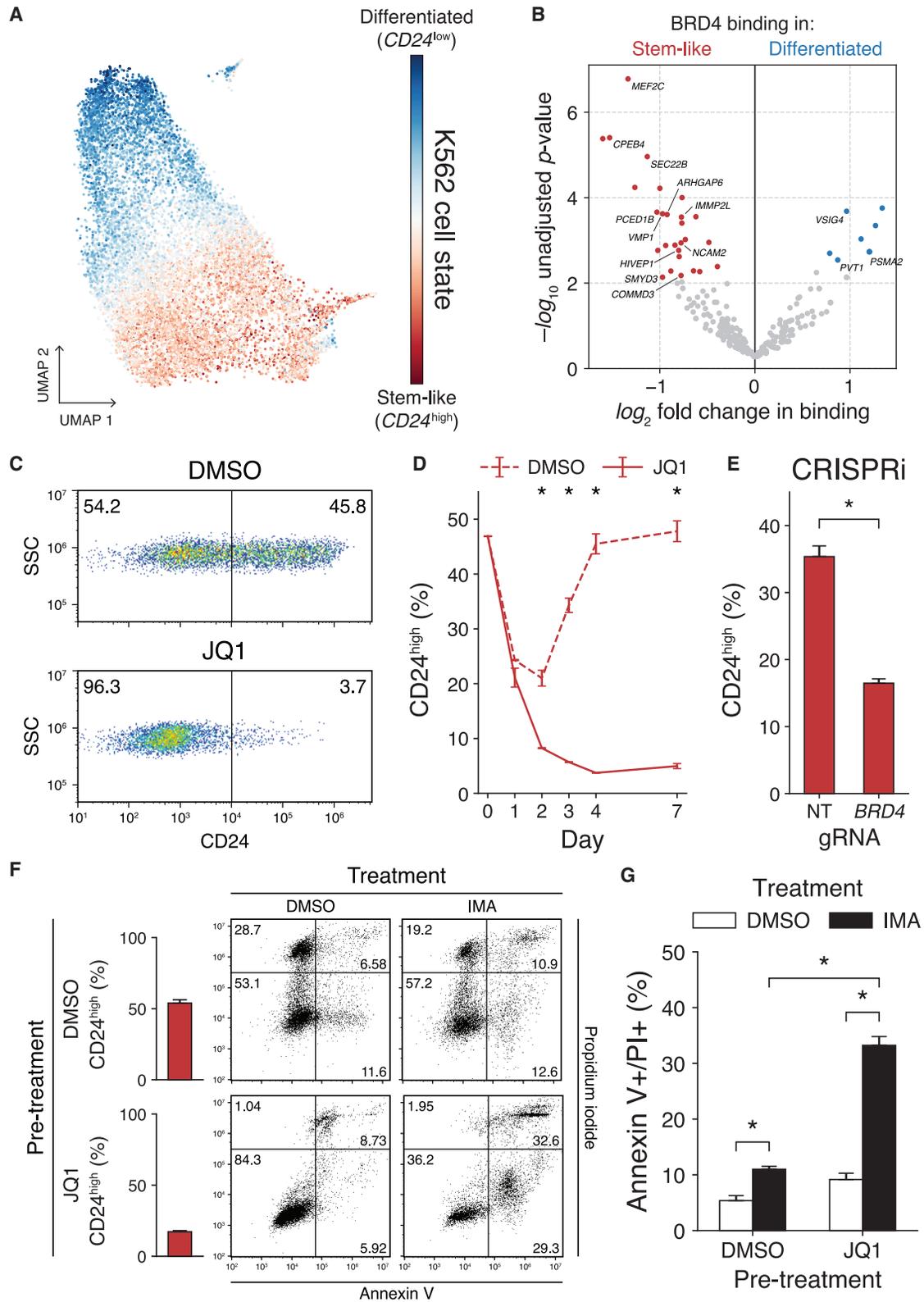


**Figure 4. scCC Works with a Variety of Transcription Factors and Cell Lines**

(A–D) scCC with SP1–HyPB in HCT-116 cells reveal SP1 binding sites. (A) Browser view of a peak from SP1 scCC. (B) Mean SP1 ChIP-seq signal at scCC SP1 peaks. (C) Heatmap of SP1 ChIP-seq signal across all scCC SP1 peaks. (D) Core SP1 motif elicited from SP1 scCC peaks. (E–H) Same as (A)–(D) but in K562 cells. (I–L) scCC with FOXA2–HyPB in HepG2 cells reveal FOXA2 binding sites. (I) Browser view of a peak from FOXA2 scCC. (J) Mean FOXA2 ChIP-seq signal at scCC FOXA2 peaks. (K) Heatmap of FOXA2 ChIP-seq signal across all scCC FOXA2 peaks. (L) Core FOXA2 motif elicited from FOXA2 scCC peaks. (M–P) scCC with BAP1–HyPB in OCM-1A cells reveal BAP1 binding sites. (M) Browser view of a peak from BAP1 scCC. (N) Mean bulk BAP1 calling cards signal at scCC BAP1 peaks. (O) Heatmap of bulk BAP1 calling cards signal across all scCC BAP1 peaks. (P) YY1 motif elicited from BAP1 scCC peaks. See also [Figure S4](#). FC, fold change.

sharply delineated disease states (Hnisz et al., 2013; Whyte et al., 2013), recent studies have shown that even closely related subpopulations of the same cell type can show subtle changes in BRD4 enrichment and enhancer utilization (Knoechel et al., 2014; Rathert et al., 2015). Recently, K562 cultures have been shown to be mixtures of a stem-like state characterized by high levels of the surface marker CD24, and a more differentiated, erythroleukemic state marked by low CD24 expression, with individual cells dynamically oscillating between these two extremes (Litzenburger et al., 2017). As we had profiled BRD4 binding in K562 cells with scCC, we wondered whether we could see evidence of these two states and, if so, whether there was differential utilization of BRD4 between them.

We first scored cells based on a principal-component analysis (PCA) of gene expression (Figures S5A and S5B), which revealed a gradient of identities along a stem-like-to-differentiated cell-state axis (Figure 5A). We then separated cells into  $CD24^{\text{high}}$  and  $CD24^{\text{low}}$  clusters (Figures S5C and S5D) and asked whether we could detect any differences in BRD4 binding between them in our scCC data. Indeed, we found multiple peaks that showed significant differential binding (Figure 5B). We corroborated these hits by comparing our peak calls to bulk BRD4 and H3K27ac ChIP-seq data, as well as to RNA pol II ChIA-PET data, which connects putative enhancers to actively transcribed genes (Fullwood et al., 2009). We highlight two genes that showed both differential binding and expression: *VMP1*, bound



**Figure 5. scCC Uncovers Bromodomain-Dependent Cell-State Dynamics in K562 Cells**

(A) Gradient of cell states from scRNA-seq analysis of K562 cells.

(B) Differential BRD4 binding analysis of undirected HyPBBase peaks in K562 cells.

(legend continued on next page)

more in the  $CD24^{\text{high}}$  stem-like cells, and *PVT1*, bound more in the differentiated,  $CD24^{\text{low}}$  cells (Figures S5F and S5G). *VMP1* overexpression is sufficient to induce autophagy (Ropolo et al., 2007), which is important for hematopoietic stem cell function (Folkerts et al., 2019; Ho et al., 2017) and may be one pathway recruited during these dynamic state transitions. *PVT1* can act as both a tumor-suppressor and oncogene, in both instances acting on the *MYC* locus (Cho et al., 2018).

We next investigated whether the observed differences in BRD4 binding might be causally responsible for establishing these two cell states. Downregulating BRD4 has been shown to influence cell identity across a range of cell types (Di Micco et al., 2014; Kfoury et al., 2017; Najafova et al., 2017). Thus, we hypothesized that BRD4 inhibition would change the distribution of cells in the stem-like and differentiated states. Moreover, due to the asymmetric nature of significant hits (Figure 5B), there is a subset of peaks specific to the  $CD24^{\text{high}}$  state that are not shared by the  $CD24^{\text{low}}$  state, suggesting that there may be a gene regulatory network that is recruited as cells transit from the differentiated to stem-like state and lost as they return. Hence, not only should the distribution of  $CD24^{\text{high}}/CD24^{\text{low}}$  cells change upon BRD4 perturbation but the stem-like  $CD24^{\text{high}}$  population should be more susceptible to such an intervention.

We tested this hypothesis by treating cells with JQ1, a small-molecule bromodomain inhibitor often used to disrupt BRD4 binding and alter target gene expression (Delmore et al., 2011; Garcia-Carpizo et al., 2018; Lovén et al., 2013; Sdelci et al., 2019). JQ1 treatment shifted the population from one containing equal proportions of  $CD24^{\text{high}}/CD24^{\text{low}}$  cells to one composed of almost exclusively  $CD24^{\text{low}}$  cells (>95%, Figure 5C). This conversion took place rapidly over the first 2 days, plateaued by day 4, and remained stable 1 week after treatment. In contrast, the control cells remained evenly split between the two states at this time point (Figure 5D; two-way ANOVA  $p < 0.01$ ). JQ1 was not selectively cytotoxic to  $CD24^{\text{high}}$  cells as there were no significant differences in the levels of annexin V, an early marker of apoptosis, between  $CD24^{\text{high}}$  and  $CD24^{\text{low}}$  cells, regardless of whether they had been exposed to JQ1 or DMSO (Figure S6A; three-way ANOVA  $p = 0.84$ ). Additionally, we examined whether *CD24* is a direct target of BRD4, which would imply that the loss of *CD24* staining was an unremarkable consequence of JQ1 treatment. We did not find evidence of BRD4 binding sites, either by ChIP-seq or calling cards, or of elevated H3K27 acetylation in the vicinity of *CD24* (Figure S6B). We also compared the relative changes in mRNA levels of *MYC*, a known BRD4 target (Knoechel et al., 2014; Lovén et al., 2013; Rathert et al., 2015; Zuber et al., 2011), to that of *CD24* during the first 24 h of JQ1 exposure. While *MYC* levels fell within the first 3 h of exposure, transcript levels of *CD24* decreased most precipitously between 3 and

9 h after JQ1 induction (Figure S6C). This delayed response suggests a regulatory cascade rather than direct transcriptional control. Thus, JQ1 treatment does not trivially downregulate a cell-surface marker but rather likely perturbs gene regulatory networks that include *CD24*.

While JQ1 shows greatest affinity for BRD4, it does have some promiscuity toward other bromodomains, including the orthologs BRD2 and BRD3 (Filippakopoulos et al., 2010). To address whether off-target effects were responsible for the shift in cell states, we downregulated *BRD4* expression with CRISPR interference (CRISPRi). We confirmed that our *BRD4* guide RNA (gRNA) specifically reduced expression of *BRD4* and not *BRD2* nor *BRD3* (Figure S6D; Welch's t test  $p < 0.05$ ). As with JQ1, we observed a significant decrease in the proportion of  $CD24^{\text{high}}$  cells with the *BRD4* gRNA compared to the non-targeting (NT) gRNA (Figure 5E; Welch's t test  $p < 0.01$ ), though not to the same levels as JQ1. This suggests that, while BRD4 is necessary for the observed cell-state dynamics between  $CD24^{\text{high}}$  and  $CD24^{\text{low}}$  cells, it is likely that other bromodomains also play a role.

$CD24^{\text{high}}/CD24^{\text{low}}$  cells have been previously shown to have different chemosensitivities, with the latter population showing more apoptosis when exposed to imatinib (Litzenburger et al., 2017). Therefore, we asked whether BRD4 inhibition increases imatinib sensitivity in K562 cells. If so, it would imply that the observed state shift functionally alters K562 cells as opposed to simply modulating a cell-surface marker. We first pretreated K562 cells with either DMSO or JQ1 and then challenged each pretreatment group with either DMSO or imatinib and stained for apoptosis. In the DMSO pretreatment group, the percentage of  $CD24^{\text{high}}$  cells rose to 54% on average, while for JQ1-pretreated cells the mean was 17% (Figure 5F). When imatinib was added, a substantially greater fraction of JQ1-pretreated cells underwent apoptosis relative to DMSO-pretreated cells (Figures 5F and 5G; two-way ANOVA  $p < 0.01$ ). Thus, JQ1 sensitizes K562 cells to imatinib. Furthermore, *BRD4* CRISPRi partially phenocopied this sensitization (Figures S6E and S6F; Tukey's honestly significant difference  $p = 0.68$ ). This phenomenon is likely dosage dependent: in our experiments, CRISPRi reduced *BRD4* mRNA levels by less than 50% (Figure S6D), whereas the JQ1 concentration we used is expected to almost completely abolish BRD4 activity (Filippakopoulos et al., 2010). Thus, while a mild knockdown can reduce *CD24* expression, greater inhibition may be necessary to induce imatinib sensitivity. Nevertheless, these results establish that BRD4 inhibition functionally and phenotypically shifts the underlying cell state of K562 cells.

Finally, we examined whether the JQ1-induced K562 cell-state shift was a non-specific response to generic drug

(C) Representative distributions of  $CD24^{\text{high}}$  and  $CD24^{\text{low}}$  cells after either 96 h of DMSO (top) or JQ1 (bottom) treatment.

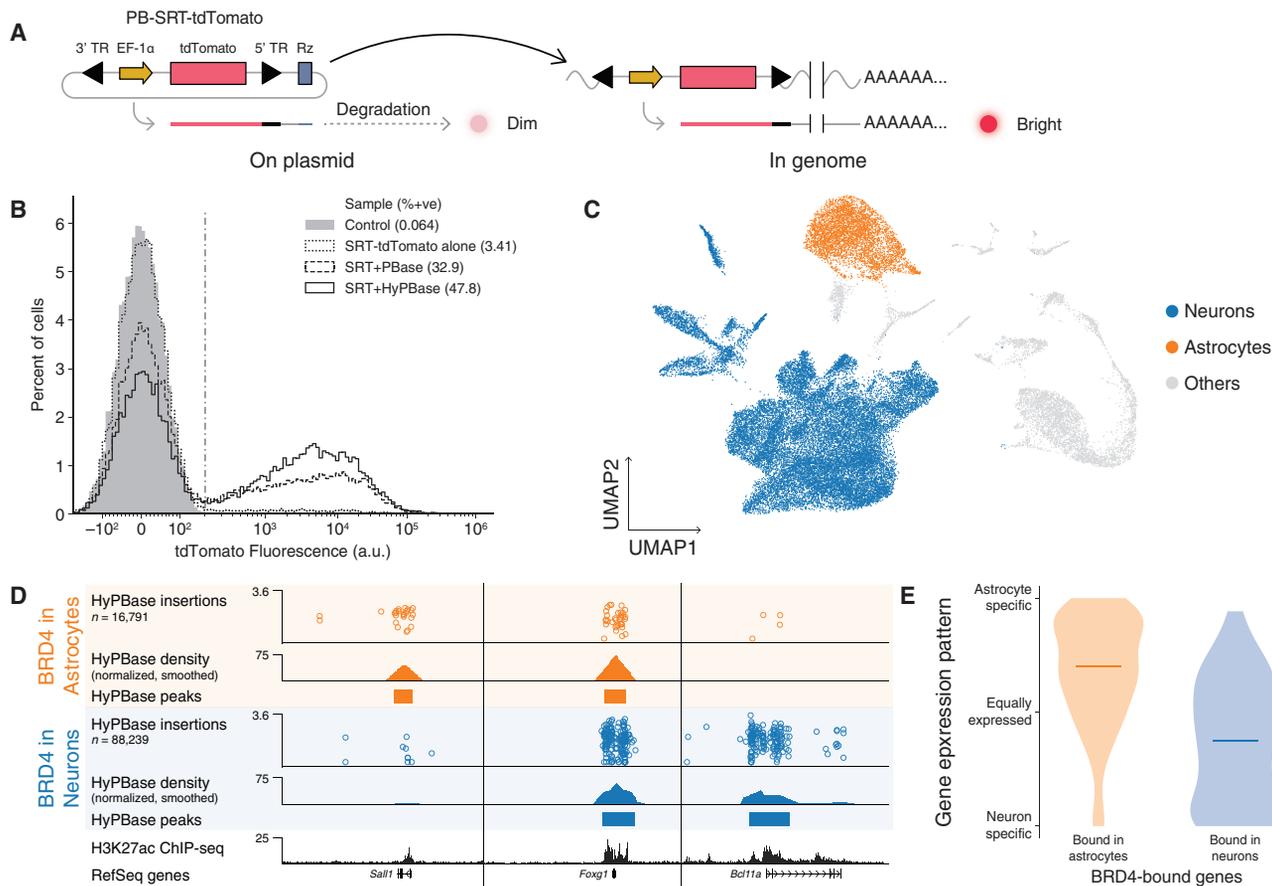
(D) Proportion of  $CD24^{\text{high}}$  cells over a 7-day time course of JQ1 treatment (three-way ANOVA  $p < 0.01$ ).

(E) Proportion of  $CD24^{\text{high}}$  cells after BRD4 CRISPRi (Welch's t test  $p < 0.01$ ).

(F) Representative plots of annexin V and PI staining in K562 cells pretreated with either DMSO or JQ1 (250 nM) and subsequently treated for 48 h with either DMSO or imatinib (1  $\mu\text{M}$ ).

(G) Quantification of (F) (two-way ANOVA  $p < 0.01$ ).

See also Figures S5 and S6. Bars represent means; error bars denote standard deviations. Experiments were performed in triplicate. DMSO, dimethyl sulfoxide; SSC, side scatter; CRISPRi, CRISPR interference; NT, non-targeting; gRNA, guide RNA; IMA, imatinib; PI, propidium iodide.



**Figure 6. scCC Deconvolves BRD4-Bound Loci in the Mouse Cortex**

(A) Schematic of PB-SRT-tdTomato.  
 (B) Distribution of fluorescence intensity in K562 cells transfected with PB-SRT-tdTomato with and without *piggyBac* transposase.  
 (C) Neuron and astrocyte clusters from scRNA-seq analysis of mouse cortex libraries transduced with AAV-HyPBBase and AAV-PB-SRT-tdTomato.  
 (D) Browser view of scCC HyPBBase peaks in astrocytes and neurons alongside whole-cortex H3K27ac ChIP-seq.  
 (E) Expression specificity distributions of genes overlapping astrocyte or neuron peaks; horizontal lines indicate medians of the distributions. See also Figure S7.  
 TR, terminal repeat; Rz, ribozyme.

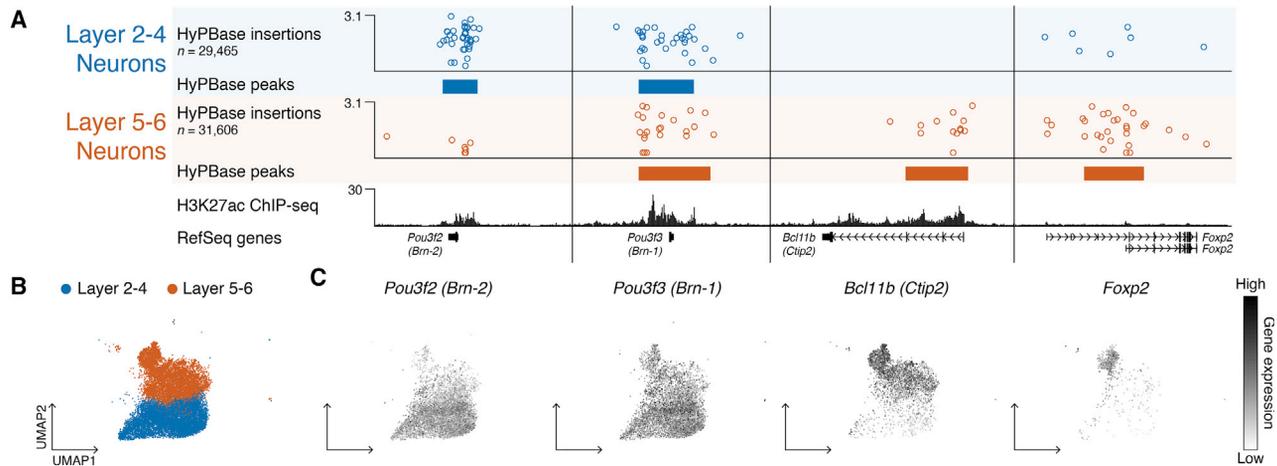
treatment. We treated K562 cultures with a panel of cell-cycle inhibitors, another class of commonly used antineoplastic agents. We first confirmed that all drugs altered the proportions of cells in either G1 or G2/M phase (Figure S6G). Cultures remained under drug treatment until 5 days had elapsed, at which point we measured CD24 levels and stained for apoptosis (Figure S6H). JQ1 caused the greatest reduction in CD24<sup>high</sup> cells (one-way ANOVA  $p < 0.01$ ) and induced significantly less apoptosis than lovastatin, its closest competitor (one-way ANOVA  $p < 0.01$ ). Thus, JQ1's effect on cell state appears to be mediated by a unique mechanism of action that is not readily replicated by cell-cycle perturbation.

### scCC Deconvolves Cell-Type-Specific BRD4 Binding Sites in the Mouse Cortex

To establish broad utility for scCC, we sought to record TF binding *in vivo*. Since *in vivo* models preclude puromycin selection, we designed an SRT carrying a fluorescent reporter (Figure 6A) and tested this reagent in cell culture. When this element was

transfected without transposase, merely 3.4% of cells registered as positive, likely due to the action of the self-cleaving ribozyme downstream of the transposon. However, when the construct was co-transfected with *piggyBac*, this figure rose as high as 48%, a 16-fold increase in signal (Figure 6B). Thus, this new construct, PB-SRT-tdTomato, allows us to collect cells carrying calling card insertions by fluorescence activated cell sorting (FACS).

Next, we delivered PB-SRT-tdTomato and HyPBBase to the postnatal mouse cortex by performing adeno-associated viral (AAV) transduction (Cammack et al., 2020) in P0–2 pups and then generating scRNA-seq and scCC libraries between P14 and P28. Most transduced cells were neurons or astrocytes (Tables S2 and S3; Figure 6C; Methods S1), which is consistent with the known tropism of AAV9 (Cammack et al., 2020; Schuster et al., 2014). Therefore, we analyzed insertions in neurons (excluding neuroblasts and doublets) and astrocytes to determine whether scCC could recover biological differences between cell types *in vivo*. After calling peaks, we



**Figure 7. scCC Unmixes BRD4 Binding in Cortical Excitatory Neurons and Identifies Known Layer Markers**

(A) Browser view of scCC HyPBase peaks in upper (layer 2–4) or lower (layer 5–6) cortical excitatory neurons alongside whole-cortex H3K27ac ChIP-seq. (B) Layer 2–4 and layer 5–6 cortical excitatory neurons highlighted among the scRNA-seq clusters. (C) Single-cell gene expression patterns of the four genes from (A).

identified astrocyte-specific, neuron-specific, and shared BRD4 binding sites (Figure 6D). Since BRD4 ChIP-seq has not yet been reported for the mouse brain, we compared our peak calls to a recent cortical H3K27ac ChIP-seq dataset (Stroud et al., 2017), and, while this dataset is a mélange of all cell types in the brain, scCC peaks in both astrocytes and neurons showed statistically significant enrichment of H3K27ac signal (Figures S7A and S7C; KS test  $p < 10^{-9}$  in each case). Moreover, genes near astrocyte peaks were more likely to be specifically expressed in astrocytes and vice versa for genes near neuron peaks (Figure 4E; STAR Methods). Furthermore, Gene Ontology enrichment analysis (Mi et al., 2017) on the set of genes near astrocyte peaks included terms like “gliogenesis,” and “glial cell differentiation,” as well as copper metabolism (Figure S7B), a known function of astrocytes (Scheiber and Dringen, 2013), while the set of genes near neuronal peaks was enriched for terms related to synapse assembly, axonal guidance, and neuron development (Figure S7D). We conclude that scCC can accurately identify cell-type-specific BRD4 binding sites *in vivo*.

Last, we asked whether scCC *in vivo* could discriminate BRD4 binding between closely related cell types, much as we had shown *in vitro* with K562 cells. From our scRNA-seq data, we identified upper- and lower-layer cortical excitatory neurons (Figure 7B; Table S3; Methods S1) and compared HyPBase scCC data between them to identify shared and specific BRD4-bound loci. As a positive control, we found a shared BRD4 binding site at the *Pou3f3* (*Brn-1*) locus (Figure 7A), which was broadly expressed in both populations (Figure 7C) and has been used to label layers 2–5 of the postnatal cortex (Molyneaux et al., 2007; Pucilowska et al., 2012). Differential binding analysis showed specific BRD4 enrichment at *Pou3f2* (*Brn-2*) in upper-layer neurons, which is more restricted to layers 2–4 than *Pou3f3* (Fan et al., 2008; Molyneaux et al., 2007), while lower cortical neurons showed BRD4 binding at *Bcl11b* (*Ctip2*) and *Foxp2*, common markers of layer 5 and

layer 6 neurons, respectively (Figure 7A; one-tailed Poisson  $p < 10^{-9}$  in each instance) (Molyneaux et al., 2007; Rašin et al., 2007). The expression patterns of these genes mirrored BRD4’s binding specificity, with *Pou3f2*’s expression mostly contained in the layer 2–4 cluster and the expression of *Bcl11b* and *Foxp2* restricted to the layer 5–6 neuron population (Figure 7C). Thus, scCC can identify differentially bound loci between very similar cell types *in vivo*.

## DISCUSSION

scCC enables simultaneous characterization of gene expression and TF binding in heterogeneous systems. The method is robust and flexible: we have demonstrated that it can map multiple kinds of DNA binding proteins—from sequence-specific TFs like SP1 and FOXA2, to indirect, chromatin-associated factors like BRD4 and BAP1—in a variety of *in vitro* systems and *in vivo* in the mouse cortex. Furthermore, our finding that cell-state transitions in K562 cells are mediated by bromodomain proteins including BRD4 demonstrates how scCC can lead to new hypotheses about transcriptional regulation in dynamic systems. Our approach fills a recognized void in the field (Shapiro et al., 2013; Shema et al., 2019) and is readily compatible with high-throughput droplet microfluidic platforms such as the 10x Chromium. We anticipate this technique will empower researchers to study TF binding in a variety of challenging *ex vivo* and *in situ* models.

The defining feature of scCC is the SRT. While here we have reported *piggyBac* and *Sleeping Beauty* SRTs (Methods S1), the self-reporting paradigm may be broadly generalizable. Expanding the palette of SRT systems could yield further insight into chromatin dynamics (Yoshida et al., 2017). Moreover, SRTs may enable multiplexed studies of TF binding, either through the simultaneous expression of many TFs, each tagged to a different transposase, or through the use of multiple bar-coded TF-*piggyBac* fusions expressed polyclonally in culture.

Since SRTs can be widely dispersed through the genome, full-length sequencing of self-reporting transcripts may find new PASs (Methods S1). Finally, SRTs could lead to new single-cell transposon-based assays. For example, just as CRISPR/Cas9 has been combined with scRNA-seq to assess the transcriptional effects of many single gene perturbations in parallel (Datlinger et al., 2017; Dixit et al., 2016), SRTs could enable massively multiplexed transposon mutagenesis screens to be read out by scRNA-seq.

One concern with calling cards is the potential for insertional mutagenesis of target genes leading to cell death and, consequently, false negatives. Previous work in diploid yeast found that calling cards are deposited into the promoters of essential and non-essential genes at comparable frequencies (Wang et al., 2011). Since mammalian genomes have much larger intergenic regions than yeast, human and mice genomes are likely also able to tolerate calling card transpositions. Long-term follow-up of mice transduced intracranially with AAV calling cards showed no significant tissue pathology, behavioral deficits, developmental defects, or metabolic dysregulation (Cammack et al., 2020). This suggests calling cards imposes, at most, a small mutagenic burden, though more studies are needed to verify this.

Another potential drawback of calling cards is that exogenous expression of a TF at supraphysiological levels may lead to ectopic binding and, consequently, false positives. We note that over 90% of our peaks from scCC of SP1 in HCT-116 cells and FOXA2 in HepG2 cells were within 1,000 bp of a ChIP-seq peak from the respective TF. This suggests that calling card peaks reflect endogenous binding, though this behavior may vary by factor. Overexpression might also alter the transcriptome of transfected cells. Comparing gene expression levels between cells treated with TF-*piggyBac* and the undirected *piggyBac* control cells can determine whether there is transcriptional perturbation and to what extent. Tagging the endogenous TF locus with *piggyBac* ensures native expression levels and would alleviate both concerns.

The relatively few insertions recovered on a per-cell basis inflates the number of cells that must be analyzed. We recommend processing enough cells to obtain at least 15,000 insertions to analyze BRD4-bound SEs with undirected *piggyBac*, and at least 30,000 insertions for both constructs in TF-directed experiments. This should achieve moderate sensitivities (~50%; Methods S1) that can be increased by collecting more insertions. The scant data recovered on a per-cell level likely stem from limited transposase activity—up to 15–30 insertions per cell for PBase (Kettlun et al., 2011; Saridey et al., 2009; Wang et al., 2008; Wilson et al., 2007) and potentially up to 100 for HyPBase (Kalhor et al., 2018; Yusa et al., 2011)—and the low capture rate of mRNA transcripts in droplet scRNA-seq (Hwang et al., 2018). This sparsity precludes certain kinds of analyses, such as multimodal data integration. Moreover, *piggyBac*'s strict preference for TTAAT tetramers also contributes to broader peaks with lower spatial resolution. While we overcame the latter constraint by focusing on peak centers and narrow peaks, peak width is inversely correlated with the number of insertions analyzed; as such, improving recovery of SRTs from single cells should be prioritized. Some of these gains may come organically as the

transcript capture rates of scRNA-seq technologies improve. Since the per-cell costs for scRNA-seq are falling exponentially (Svensson et al., 2018), combining scCC with sample multiplexing strategies like cell hashing (Stoeckius et al., 2018) or combinatorial barcoding (Rosenberg et al., 2018) may be an attractive approach to increase sensitivity.

Finally, calling card insertions, being integrated into the genome and preserved through mitosis, could serve as a molecular memory for recording TF binding events. The use of an inducible transposase (Qi et al., 2017) would enable the recording and identification of temporally restricted TF binding sites. This would help uncover the stepwise order of events underlying the regulation of specific genes and inform cell-fate decision making. More generally, transposon insertions could serve as barcodes of developmental lineage. Single transposition events have been used to delineate relationships during hematopoiesis (Rodriguez-Fraticelli et al., 2018; Sun et al., 2014). Multiplexing several SRTs across every cell in an organism could code lineage in a cumulative and combinatorially diverse fashion, generating high-resolution cellular phylogenies.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - DNA- versus RNA-based recovery
  - *In vitro* bulk calling card experiments
  - Isolation and RT of bulk RNA
  - Amplifying self-reporting transcripts from RNA
  - Generation of bulk RNA calling card libraries
  - Colony formation assay
  - *In vitro* single cell calling card experiments
  - Single cell RNA-seq library preparation
  - Single cell calling cards library preparation
  - Staining protocols for K562 cells
  - JQ1 treatment of K562 cells
  - BRD4 CRISPRi of K562 cells
  - Imatinib treatments of K562 cells
  - Cell cycle perturbation of K562 cells
  - SRT-tdTomato fluorescence validation
  - *In vivo* scCC experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Interpreting calling card tracks
  - Sequencing and analysis: bulk DNA CC libraries
  - Sequencing and analysis: bulk RNA CC libraries
  - Sequencing and analysis: scRNA-seq libraries
  - Sequencing and analysis: scCC libraries
  - Peak calling on calling card data
  - TF binding analysis

- BRD4 sensitivity, specificity, and precision
- Downsampling and replication analysis
- Analysis of external datasets
- Cell state analyses of K562: scRNA-seq and scCC
- Analysis of K562 experiments
- *In vivo* scCC analysis and validation
- **ADDITIONAL RESOURCES**

#### SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2020.06.037>.

#### ACKNOWLEDGMENTS

We thank Jessica Hoisington-Lopez and MariaLynn Crosby from the DNA Sequencing Innovation Lab at The Edison Family Center for Genome Sciences and Systems Biology for their sequencing expertise. Additional sequencing was performed by the Genome Technology Access Center (GTAC) in the Department of Genetics at Washington University School of Medicine. Flow cytometry was conducted and analyzed at Siteman Flow Cytometry Core at the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital in St. Louis, MO. Both GTAC and the Flow Cytometry Core are supported by NCI Cancer Center Support Grant P30 CA91842 to the Siteman Cancer Center. GTAC is further supported 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 work was supported by the Hope Center Viral Vectors Core and a P30 Neuroscience Blueprint Interdisciplinary Center Core award to Washington University (P30 NS057105). This work was also supported by NIH grants R21 HG009750 (R.D.M. and S.A.M.), R01 GM126112 (S.A.M.), U01 MH109133 (J.D.D. and R.D.M.), and RF1 MH117070 (J.D.D. and R.D.M.) and a grant from the Children's Discovery Institute (#MC-II-2016-533, R.D.M.). S.A.M. was supported by an Allen Distinguished Investigator Award (through the Paul G. Allen Frontiers Group), a Vallee Scholar Award, and a Sloan Research Fellowship. A.M. was supported by NIH grants T32 GM007200, T32 HG000045, and F30 HG009986, A.J.C. by NIH T32 GM008151; M.J.V. by NIH F32 NS105363; M.A.L. by NIH T32 HL125241; and T.L. by NIH T32 GM007067. We thank Donald Conrad and Ben Humphreys for their advice and constructive feedback during this project, and Shirley McKinney for laboratory support. Last, we thank the three anonymous reviewers for their feedback and suggestions. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NIH.

#### AUTHOR CONTRIBUTIONS

A.M., M.N.W., Z.Q., and R.D.M. developed the self-reporting transposon (SRT) technology. A.M. and M.N.W. created and optimized the molecular workflow for recovering SRTs from bulk RNA-seq libraries, with contributions from Z.Q. A.M. developed and optimized the protocol for recovering SRTs from single-cell RNA-seq libraries. S.A.M. provided guidance and assistance with single-cell experiments. Z.Q., T.L., and C.G. cloned SRT constructs, while M.A.L. generated lentivirus. A.M., M.N.W., J.D.D., and R.D.M. designed the experiments. A.M., M.N.W., X.C., and J.H. performed the *in vitro* experiments. M.N.W., A.J.C., and M.J.V. performed the *in vivo* experiments. A.M., X.C., and J.H. generated sequencing libraries. A.M., J.D.D., and R.D.M. analyzed the data. A.M. and R.D.M. wrote the manuscript in consultation with all authors.

#### DECLARATION OF INTERESTS

A.M., M.N.W., Z.Q., and R.D.M. have filed patent applications related to this work.

Received: February 21, 2019

Revised: April 14, 2020

Accepted: June 23, 2020

Published: July 24, 2020

#### SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Chen et al. (2017); Chiu et al. (2018); Ivics et al. (1997); Picelli et al. (2014); Sabari et al. (2018); Yen et al. (2004); Zheng et al. (2016).

#### REFERENCES

- Ai, S., Xiong, H., Li, C.C., Luo, Y., Shi, Q., Liu, Y., Yu, X., Li, C., and He, A. (2019). Profiling chromatin states using single-cell ChIP-seq. *Nat. Cell Biol.* *21*, 1164–1172.
- Angermueller, C., Clark, S.J., Lee, H.J., Macaulay, I.C., Teng, M.J., Hu, T.X., Krueger, F., Smallwood, S., Ponting, C.P., Voet, T., et al. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* *13*, 229–232.
- Avey, D., Sankararaman, S., Yim, A.K.Y., Barve, R., Milbrandt, J., and Mitra, R.D. (2018). Single-Cell RNA-Seq Uncovers a Robust Transcriptional Response to Morphine by Glia. *Cell Rep.* *24*, 3619–3629.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* *57*, 289–300.
- Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. *Nature* *371*, 435–438.
- Brooks, E.E., Gray, N.S., Joly, A., Kerwar, S.S., Lum, R., Mackman, R.L., Norman, T.C., Rosete, J., Rowe, M., Schow, S.R., et al. (1997). CVT-313, a specific and potent inhibitor of CDK2 that prevents neointimal proliferation. *J. Biol. Chem.* *272*, 29207–29211.
- Brunner, A.L., Johnson, D.S., Kim, S.W., Valouev, A., Reddy, T.E., Neff, N.F., Anton, E., Medina, C., Nguyen, L., Chiao, E., et al. (2009). Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res.* *19*, 1044–1056.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* *10*, 1213–1218.
- Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* *523*, 486–490.
- Cadiñanos, J., and Bradley, A. (2007). Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res.* *35*, e87–e87.
- Cammack, A.J., Moudgil, A., Chen, J., Vasek, M.J., Shabsovich, M., McCullough, K., Yen, A., Lagunas, T., Maloney, S.E., He, J., et al. (2020). A viral toolkit for recording transcription factor–DNA interactions in live mouse tissues. *Proc. Natl. Acad. Sci. USA* *117*, 10003–10014.
- Campagne, A., Lee, M.-K., Zielinski, D., Michaud, A., Le Corre, S., Dingli, F., Chen, H., Shahidian, L.Z., Vassilev, I., Servant, N., et al. (2019). BAP1 complex promotes transcription by opposing PRC1-mediated H2A ubiquitylation. *Nat. Commun.* *10*, 348.
- Campbell, J.N., Macosko, E.Z., Fenselau, H., Pers, T.H., Lyubetskaya, A., Tenen, D., Goldman, M., Verstegen, A.M.J., Resch, J.M., McCarroll, S.A., et al. (2017). A molecular census of arcuate hypothalamus and median eminence cell types. *Nat. Neurosci.* *20*, 484–496.
- Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* *357*, 661–667.
- Cao, J., Cusanovich, D.A., Ramani, V., Aghamirzaie, D., Pliner, H.A., Hill, A.J., Daza, R.M., McFaline-Figueroa, J.L., Packer, J.S., Christiansen, L., et al.

- (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 361, 1380–1385.
- Carbone, M., Yang, H., Pass, H.I., Krausz, T., Testa, J.R., and Gaudino, G. (2013). BAP1 and cancer. *Nat. Rev. Cancer* 13, 153–159.
- Carter, B., Ku, W.L., Kang, J.Y., Hu, G., Perrie, J., Tang, Q., and Zhao, K. (2019). Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). *Nat. Commun.* 10, 3747.
- Castillo-Hair, S.M., Sexton, J.T., Landry, B.P., Olson, E.J., Igoshin, O.A., and Tabor, J.J. (2016). FlowCite: A User-Friendly, Open Source Software Tool for Automatically Converting Flow Cytometry Data from Arbitrary to Calibrated Units. *ACS Synth. Biol.* 5, 774–780.
- Chen, W., Jia, Q., Song, Y., Fu, H., Wei, G., and Ni, T. (2017). Alternative Polyadenylation: Methods, Findings, and Impacts. *Genomics Proteomics Bioinformatics* 15, 287–300.
- Chiu, A.C., Suzuki, H.I., Wu, X., Mahat, D.B., Kriz, A.J., and Sharp, P.A. (2018). Transcriptional Pause Sites Delineate Stable Nucleosome-Associated Premature Polyadenylation Suppressed by U1 snRNP. *Mol. Cell* 69, 648–663.
- Cho, S.W., Xu, J., Sun, R., Mumbach, M.R., Carter, A.C., Chen, Y.G., Yost, K.E., Kim, J., He, J., Nevins, S.A., et al. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* 173, 1398–1412.
- Clark, S.J., Argelaguet, R., Kapourani, C.-A., Stubbs, T.M., Lee, H.J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J.C., et al. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* 9, 781.
- Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* 14, 297–301.
- Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51, 987–1000.
- Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., Jacobs, H.M., Kastriitis, E., Gilpatrick, T., Paranal, R.M., Qi, J., et al. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* 146, 904–917.
- Dey, A., Seshasayee, D., Noubade, R., French, D.M., Liu, J., Chaurushiya, M.S., Kirkpatrick, D.S., Pham, V.C., Lill, J.R., Bakalarski, C.E., et al. (2012). Loss of the tumor suppressor BAP1 causes myeloid transformation. *Science* 337, 1541–1546.
- Dey, S.S., Kester, L., Spanjaard, B., Bienko, M., and van Oudenaarden, A. (2015). Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* 33, 285–289.
- Di Micco, R., Fontanals-Cirera, B., Low, V., Ntziachristos, P., Yuen, S.K., Lovell, C.D., Dolgalev, I., Yonekubo, Y., Zhang, G., Rusinova, E., et al. (2014). Control of embryonic stem cell identity by BRD4-dependent transcriptional elongation of super-enhancer-associated pluripotency genes. *Cell Rep.* 9, 234–247.
- Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* 122, 473–483.
- Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167, 1853–1866.e17.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
- Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.
- Fan, X., Kim, H.-J., Bouton, D., Warner, M., and Gustafsson, J.-Å. (2008). Expression of liver X receptor  $\beta$  is essential for formation of superficial cortical layers and migration of later-born neurons. *Proc. Natl. Acad. Sci. USA* 105, 13445–13450.
- Filippakopoulos, P., Qi, J., Picaud, S., Shen, Y., Smith, W.B., Fedorov, O., Morse, E.M., Keates, T., Hickman, T.T., Felletar, I., et al. (2010). Selective inhibition of BET bromodomains. *Nature* 468, 1067–1073.
- Fincher, C.T., Wurtzel, O., de Hoog, T., Kravarik, K.M., and Reddien, P.W. (2018). Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* 360, eaaq1736–e1757.
- Fogarty, N.M.E., McCarthy, A., Sniijders, K.E., Powell, B.E., Kubikova, N., Blakeley, P., Lea, R., Elder, K., Wamaitha, S.E., Kim, D., et al. (2017). Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* 550, 67–73.
- Folkerts, H., Wierenga, A.T., van den Heuvel, F.A., Woldhuis, R.R., Kluit, D.S., Jaques, J., Schuringa, J.J., and Vellenga, E. (2019). Elevated VMP1 expression in acute myeloid leukemia amplifies autophagy and is protective against venetoclax-induced apoptosis. *Cell Death Dis.* 10, 421.
- Fournier, M., Bourriquen, G., Lamaze, F.C., Côté, M.C., Fournier, É., Joly-Beauparlant, C., Caron, V., Gobeil, S., Droit, A., and Bilodeau, S. (2016). FOXA and master transcription factors recruit Mediator and Cohesin to the core transcriptional regulatory circuitry of cancer cells. *Sci. Rep.* 6, 34962.
- Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 354, 769–773.
- Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor- $\alpha$ -bound human chromatin interactome. *Nature* 462, 58–64.
- García-Carpizo, V., Ruiz-Llorente, S., Sarmentero, J., Graña-Castro, O., Pisanò, D.G., and Barrero, M.J. (2018). CREBBP/EP300 bromodomains are critical to sustain the GATA1/MYC regulatory axis in proliferation. *Epigenetics Chromatin* 11, 30.
- Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* 176, 377–390.
- Gogol-Döring, A., Ammar, I., Gupta, S., Bunse, M., Miskey, C., Chen, W., Uckert, W., Schulz, T.F., Izsvák, Z., and Ivics, Z. (2016). 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.
- Gonen, N., Futtner, C.R., Wood, S., Garcia-Moreno, S.A., Salamone, I.M., Samson, S.C., Sekido, R., Poulat, F., Maatouk, D.M., and Lovell-Badge, R. (2018). Sex reversal following deletion of a single distal enhancer of *Sox9*. *Science* 360, 1469–1473.
- Greil, F., Moorman, C., and van Steensel, B. (2006). DamID: Mapping of In Vivo Protein–Genome Interactions Using Tethered DNA Adenine Methyltransferase. *Methods Enzymol.* 410, 342–359.
- Grosselin, K., Durand, A., Marsolier, J., Poitou, A., Marangoni, E., Nemati, F., Dahmani, A., Lameiras, S., Reyat, F., Frenoy, O., et al. (2019). High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* 51, 1060–1066.
- Gupta, S., Stamatoyannopoulos, J.A., Bailey, T.L., and Noble, W.S. (2007). Quantifying similarity between motifs. *Genome Biol.* 8, R24.
- Gurdon, J.B. (2016). Cell Fate Determination by Transcription Factors. *Curr. Top. Dev. Biol.* 116, 445–454.
- Haffer, B.P., Surzenko, N., Beier, K.T., Punzo, C., Trimarchi, J.M., Kong, J.H., and Cepko, C.L. (2012). Transcription factor Olig2 defines subpopulations of retinal progenitor cells biased toward specific cell fates. *Proc. Natl. Acad. Sci. USA* 109, 7882–7887.
- Hainer, S.J., Bošković, A., McCannell, K.N., Rando, O.J., and Fazio, T.G. (2019). Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* 177, 1319–1329.
- Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the Mouse Cell Atlas by Microfluidic-Seq. *Cell* 172, 1091–1107.

- Harada, A., Maehara, K., Handa, T., Arimura, Y., Nogami, J., Hayashi-Takanaka, Y., Shirahige, K., Kurumizaka, H., Kimura, H., and Ohkawa, Y. (2019). A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat. Cell Biol.* **21**, 287–296.
- Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947.
- Ho, T.T., Warr, M.R., Adelman, E.R., Lansinger, O.M., Flach, J., Verovskaya, E.V., Figueroa, M.E., and Passequé, E. (2017). Autophagy maintains the metabolism and function of young and old stem cells. *Nature* **543**, 205–210.
- Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* **50**, 346–363.
- Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95.
- Hwang, B., Lee, J.H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 96.
- Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvák, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501–510.
- Jackman, J., and O'Connor, P.M. (2001). Methods for Synchronizing Cells at Specific Stages of the Cell Cycle. *Curr. Protoc.* Published online May 1, 2001. <https://doi.org/10.1002/0471143030.cb0803s00>.
- Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502.
- Kalhor, R., Kalhor, K., Mejia, L., Leeper, K., Graveline, A., Mali, P., and Church, G.M. (2018). Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804.
- Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The *Drosophila* embryo at single-cell transcriptome resolution. *Science* **358**, 194–199.
- Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930.
- Kettlun, C., Galvan, D.L., George, A.L., Jr., Kaja, A., and Wilson, M.H. (2011). Manipulating piggyBac transposon chromosomal integration site selection in human cells. *Mol. Ther.* **19**, 1636–1644.
- Kfoury, N., Qi, Z., Wilkinson, M., Broestl, L., Berrett, K., Moudgil, A., Sankararaman, S., Chen, X., Gertz, J., Mitra, R., et al. (2017). Brd4-bound enhancers drive critical sex differences in glioblastoma. *bioRxiv*. <https://doi.org/10.1101/199059>.
- Kind, J., Pagie, L., Ortazokoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A., and van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178–192.
- Kind, J., Pagie, L., de Vries, S.S., Nahidiazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015). Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134–147.
- Knoechel, B., Roderick, J.E., Williamson, K.E., Zhu, J., Lohr, J.G., Cotton, M.J., Gillespie, S.M., Fernandez, D., Ku, M., Wang, H., et al. (2014). An epigenetic mechanism of resistance to targeted therapy in T cell acute lymphoblastic leukemia. *Nat. Genet.* **46**, 364–370.
- Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J., Tissières, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A., et al. (2016). Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**, 633–642.
- Lalli, M.A., Avey, D., Dougherty, J.D., Milbrandt, J., and Mitra, R.D. (2019). High-throughput single-cell functional elucidation of neurodevelopmental disease-associated genes reveals convergent mechanisms altering neuronal differentiation. *bioRxiv*. <https://doi.org/10.1101/862680>.
- Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* **32**, 42–56.
- Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237–1251.
- Lee, C.S., Friedman, J.R., Fulmer, J.T., and Kaestner, K.H. (2005). The initiation of liver development is dependent on Foxa transcription factors. *Nature* **435**, 944–947.
- Litzenburger, U.M., Buenrostro, J.D., Wu, B., Shen, Y., Sheffield, N.C., Kathirika, A., Greenleaf, W.J., and Chang, H.Y. (2017). Single-cell epigenomic variability reveals functional cancer heterogeneity. *Genome Biol.* **18**, 15.
- Liu, X., Huang, J., Chen, T., Wang, Y., Xin, S., Li, J., Pei, G., and Kang, J. (2008). Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells. *Cell Res.* **18**, 1177–1189.
- Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334.
- Macaulay, I.C., Haerty, W., Kumar, P., Li, Y.L., Hu, T.X., Teng, M.J., Goolam, M., Saurat, N., Coupland, P., Shirley, L.M., et al. (2015). G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519–522.
- Machanic, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697.
- Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes Dev.* **8**, 2282–2292.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12.
- Martins, A.L., Walavalkar, N.M., Anderson, W.D., Zang, C., and Guertin, M.J. (2018). Universal correction of enzymatic sequence bias reveals molecular signatures of protein/DNA interactions. *Nucleic Acids Res.* **46**, e9.
- Matatall, K.A., Agapova, O.A., Onken, M.D., Worley, L.A., Bowcock, A.M., and Harbour, J.W. (2013). BAP1 deficiency causes loss of melanocytic cell identity in uveal melanoma. *BMC Cancer* **13**, 371.
- Mátés, L., Chuah, M.K.L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., et al. (2009). Molecular evolution of a novel hyperactive *Sleeping Beauty* transposase enables robust stable gene transfer in vertebrates. *Nat. Genet.* **41**, 753–761.
- McClelland, M.L., Mesh, K., Lorenzana, E., Chopra, V.S., Segal, E., Watanabe, C., Haley, B., Mayba, O., Yaylaoglu, M., Gnad, F., and Firestein, R. (2016). CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer. *J. Clin. Invest.* **126**, 639–652.
- Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* **45** (D1), D183–D189.
- Mizuguchi, R., Sugimori, M., Takebayashi, H., Kosako, H., Nagao, M., Yoshida, S., Nabeshima, Y., Shimamura, K., and Nakafuku, M. (2001). Combinatorial roles of olig2 and neurogenin2 in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons. *Neuron* **31**, 757–771.
- Molyneux, B.J., Arlotta, P., Menezes, J.R.L., and Macklis, J.D. (2007). Neuronal subtype specification in the cerebral cortex. *Nat. Rev. Neurosci.* **8**, 427–437.
- Najafova, Z., Tirado-Magallanes, R., Subramaniam, M., Hossan, T., Schmidt, G., Nagarajan, S., Baumgart, S.J., Mishra, V.K., Bedi, U., Hesse, E., et al. (2017). BRD4 localization to lineage-specific enhancers is associated with a distinct transcription factor repertoire. *Nucleic Acids Res.* **45**, 127–141.
- Oliphant, T.E. (2015). *Guide to NumPy* (Continuum Press).

- Peterson, V.M., Zhang, K.X., Kumar, N., Wong, J., Li, L., Wilson, D.C., Moore, R., McClanahan, T.K., Sadekova, S., and Klappenbach, J.A. (2017). Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939.
- Philipsen, S., and Suske, G. (1999). A tale of three fingers: the family of mammalian Sp/XKLF transcription factors. *Nucleic Acids Res.* **27**, 2991–3000.
- Picelli, S., Björklund, A.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040.
- Ponnaluri, V.K.C., Zhang, G., Estève, P.-O., Spracklin, G., Sian, S., Xu, S.Y., Benoukrif, T., and Pradhan, S. (2017). NicE-seq: high resolution open chromatin profiling. *Genome Biol.* **18**, 122.
- Pott, S., and Lieb, J.D. (2015). What are super-enhancers? *Nat. Genet.* **47**, 8–12.
- Pucilowska, J., Puzerey, P.A., Karlo, J.C., Galán, R.F., and Landreth, G.E. (2012). Disrupted ERK signaling during cortical development leads to abnormal progenitor proliferation, neuronal and network excitability and behavior, modeling human neuro-cardio-facial-cutaneous and related syndromes. *J. Neurosci.* **32**, 8663–8677.
- Qi, Z., Wilkinson, M.N., Chen, X., Sankararaman, S., Mayhew, D., and Mitra, R.D. (2017). An optimized, broadly applicable piggyBac transposon induction system. *Nucleic Acids Res.* **45**, e55.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
- Raff, T., van der Giet, M., Endemann, D., Wiederholt, T., and Paul, M. (1997). Design and testing of  $\beta$ -actin primers for RT-PCR that do not co-amplify processed pseudogenes. *Biotechniques* **23**, 456–460.
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44** (W1), W160–W165.
- Rašin, M.-R., Gazula, V.-R., Breunig, J.J., Kwan, K.Y., Johnson, M.B., Liu-Chen, S., Li, H.-S., Jan, L.Y., Jan, Y.-N., Rakic, P., and Sestan, N. (2007). Numb and NumbL are required for maintenance of cadherin-based adhesion and polarity of neural progenitors. *Nat. Neurosci.* **10**, 819–827.
- Rathert, P., Roth, M., Neumann, T., Muedter, F., Roe, J.-S., Muhar, M., Deswal, S., Cerny-Reiterer, S., Peter, B., Jude, J., et al. (2015). Transcriptional plasticity promotes primary and acquired resistance to BET inhibition. *Nature* **525**, 543–547.
- Robitaille, T.P., Tollerud, E.J., Greenfield, P., Droettboom, M., Bray, E., Aldcroft, T., Davis, M., Ginsburg, A., Price-Whelan, A.M., Kerzendorf, W.E., et al. (2013). Astropy: A community Python package for astronomy. *Astron. Astrophys.* **558**, A33.
- Rodriguez-Fraticelli, A.E., Wolock, S.L., Weinreb, C.S., Panero, R., Patel, S.H., Jankovic, M., Sun, J., Calogero, R.A., Klein, A.M., and Camargo, F.D. (2018). Clonal analysis of lineage fate in native haematopoiesis. *Nature* **553**, 212–216.
- Rooijers, K., Markodimitrakaki, C.M., Rang, F.J., de Vries, S.S., Chialastri, A., de Luca, K.L., Mooijman, D., Dey, S.S., and Kind, J. (2019). Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nat. Biotechnol.* **37**, 766–772.
- Ropolo, A., Grasso, D., Pardo, R., Sacchetti, M.L., Archange, C., Lo Re, A., Seux, M., Nowak, J., Gonzalez, C.D., Iovanna, J.L., and Vaccaro, M.I. (2007). The pancreatitis-induced vacuole membrane protein 1 triggers autophagy in mammalian cells. *J. Biol. Chem.* **282**, 37124–37133.
- Rosenberg, A.B., Roco, C.M., Muscat, R.A., Kuchina, A., Sample, P., Yao, Z., Grayback, L.T., Peeler, D.J., Mukherjee, S., Chen, W., et al. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182.
- Rotem, A., Ram, O., Shores, N., Sperling, R.A., Goren, A., Weitz, D.A., and Bernstein, B.E. (2015). Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat. Biotechnol.* **33**, 1165–1172.
- Rouillard, A.D., Gunderson, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., and Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)* **2016**, baw100.
- Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958.
- Sanson, K.R., Hanna, R.E., Hegde, M., Donovan, K.F., Strand, C., Sullender, M.E., Vaimberg, E.W., Goodale, A., Root, D.E., Piccioni, F., and Doench, J.G. (2018). Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 5416.
- Saridey, S.K., Liu, L., Doherty, J.E., Kaja, A., Galvan, D.L., Fletcher, B.S., and Wilson, M.H. (2009). PiggyBac transposon-based inducible gene expression in vivo after somatic cell gene transfer. *Mol. Ther.* **17**, 2115–2120.
- Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., et al. (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015–1030.
- Saxena, A., Wagatsuma, A., Noro, Y., Kuji, T., Asaka-Oba, A., Watahiki, A., Gurnot, C., Fagiolini, M., Hensch, T.K., and Carninci, P. (2012). Trehalose-enhanced isolation of neuronal sub-types from adult mouse brain. *Bio-techniques* **52**, 381–385.
- Scargle, J.D., Norris, J.P., Jackson, B., and Chiang, J. (2013). Studies in astronomical time series analysis. VI. Bayesian block representation. *Astrophys. J.* **764**, 167.
- Scheiber, I.F., and Dringen, R. (2013). Astrocyte functions in the copper homeostasis of the brain. *Neurochem. Int.* **62**, 556–565.
- Schuster, D.J., Dykstra, J.A., Riedl, M.S., Kitto, K.F., Belur, L.R., Mclvor, R.S., Eide, R.P., Fairbanks, C.A., and Vulchanova, L. (2014). Biodistribution of adeno-associated virus serotype 9 (AAV9) vector after intrathecal and intravenous delivery in mouse. *Front. Neuroanat.* **8**, 42.
- Sdelci, S., Rendeiro, A.F., Rathert, P., You, W., Lin, J.G., Ringler, A., Hofstätter, G., Moll, H.P., Gürtl, B., Farlik, M., et al. (2019). MTHFD1 interaction with BRD4 links folate metabolism to transcriptional regulation. *Nat. Genet.* **51**, 990–998.
- Seabold, S., and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference, Vol. 57 (SciPy Society)*, p. 61.
- Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630.
- Shema, E., Bernstein, B.E., and Buenrostro, J.D. (2019). Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nat. Genet.* **51**, 19–25.
- Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C., Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T., et al. (2016). ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44** (D1), D726–D732.
- Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868.
- Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., 3rd, Smibert, P., and Satija, R. (2018). Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224.
- Stroud, H., Su, S.C., Hrvatin, S., Greben, A.W., Renthall, W., Boxer, L.D., Nagy, M.A., Hochbaum, D.R., Kinde, B., Gabel, H.W., and Greenberg, M.E. (2017). Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States. *Cell* **171**, 1151–1164.
- Sun, J., Ramos, A., Chapman, B., Johnnidis, J.B., Le, L., Ho, Y.-J., Klein, A., Hofmann, O., and Camargo, F.D. (2014). Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327.
- Svensson, V., Vento-Tormo, R., and Teichmann, S.A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604.

- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.
- Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garen, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature* 563, 72–78.
- The Astropy Collaboration, Price-Whelan, A.M., Sipőcz, B.M., Günther, H.M., Lim, P.L., Crawford, S.M., Conseil, S., Shupe, D.L., Craig, M.W., Dencheva, N., et al. (2018). The Astropy Project: Building an inclusive, open-science project and status of the v2.0 core package. *Astron. J.* 156, 123.
- Vassilev, L.T., Tovar, C., Chen, S., Knezevic, D., Zhao, X., Sun, H., Heimbrook, D.C., and Chen, L. (2006). Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1. *Proc. Natl. Acad. Sci. USA* 103, 10660–10665.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al.; SciPy 1.0 Contributors (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272.
- Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat. Protoc.* 2, 1467–1478.
- Wang, H., Johnston, M., and Mitra, R.D. (2007). Calling cards for DNA-binding proteins. *Genome Res.* 17, 1202–1209.
- Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., and Liu, P. (2008). Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. USA* 105, 9290–9295.
- Wang, H., Mayhew, D., Chen, X., Johnston, M., and Mitra, R.D. (2011). Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins. *Genome Res.* 21, 748–755.
- Wang, H., Mayhew, D., Chen, X., Johnston, M., and Mitra, R.D. (2012a). “Calling cards” for DNA-binding proteins in mammalian cells. *Genetics* 190, 941–949.
- Wang, X., Spandidos, A., Wang, H., and Seed, B. (2012b). PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* 40, D1144–D1149.
- Wang, Q., Xiong, H., Ai, S., Yu, X., Liu, Y., Zhang, J., and He, A. (2019). Co-BATCH for High-Throughput Single-Cell Epigenomic Profiling. *Mol. Cell* 76, 206–216.e7.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319.
- Wilson, M.H., Coates, C.J., and George, A.L., Jr. (2007). PiggyBac transposon-mediated gene transfer in human cells. *Mol. Ther.* 15, 139–145.
- Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15.
- Wu, S.C.-Y., Meir, Y.-J.J., Coates, C.J., Handler, A.M., Pelczar, P., Moisyadi, S., and Kaminski, J.M. (2006). piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc. Natl. Acad. Sci. USA* 103, 15008–15013.
- Xie, S., Duan, J., Li, B., Zhou, P., and Hon, G.C. (2017). Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol. Cell* 66, 285–299.e5.
- Yen, L., Svendsen, J., Lee, J.-S., Gray, J.T., Magnier, M., Baba, T., D’Amato, R.J., and Mulligan, R.C. (2004). Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* 431, 471–476.
- Yen, M., Qi, Z., Chen, X., Cooper, J.A., Mitra, R.D., and Onken, M.D. (2018). Transposase mapping identifies the genomic targets of BAP1 in uveal melanoma. *BMC Med. Genomics* 11, 97.
- Yoshida, J., Akagi, K., Misawa, R., Kokubu, C., Takeda, J., and Horie, K. (2017). Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Sci. Rep.* 7, 43613.
- Yu, H., Mashtalir, N., Daou, S., Hammond-Martel, I., Ross, J., Sui, G., Hart, G.W., Rauscher, F.J., 3rd, Drobetsky, E., Milot, E., et al. (2010). The ubiquitin carboxyl hydrolase BAP1 forms a ternary complex with YY1 and HCF-1 and is a critical regulator of gene expression. *Mol. Cell Biol.* 30, 5071–5085.
- Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci. USA* 108, 1531–1536.
- Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347, 1138–1142.
- Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E., La Manno, G., et al. (2018). Molecular Architecture of the Mouse Nervous System. *Cell* 174, 999–1014.e22.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.
- Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O’Keefe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* 34, 11929–11947.
- Zheng, D., Liu, X., and Tian, B. (2016). 3’READS+, a sensitive and accurate method for 3’ end sequencing of polyadenylated RNA. *RNA* 22, 1631–1639.
- Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049.
- Zhou, X., Maricque, B., Xie, M., Li, D., Sundaram, V., Martin, E.A., Koebbe, B.C., Nielsen, C., Hirst, M., Farnham, P., et al. (2011). The Human Epigenome Browser at Washington University. *Nat. Methods* 8, 989–990.
- Zhu, X., Zuo, H., Maher, B.J., Serwanski, D.R., LoTurco, J.J., Lu, Q.R., and Nishiyama, A. (2012). Olig2-dependent developmental fate switch of NG2 cells. *Development* 139, 2299–2307.
- Zuber, J., Shi, J., Wang, E., Rappaport, A.R., Herrmann, H., Sison, E.A., Magoon, D., Qi, J., Blatt, K., Wunderlich, M., et al. (2011). RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* 478, 524–528.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Brilliant Violet 421 anti-human CD24 Antibody (clone ML5)	BioLegend	Cat# 311121; RRID: AB_10915556
Brilliant Violet 421 Mouse IgG2a, $\kappa$ Isotype Ctrl Antibody (clone MOPC-173)	BioLegend	Cat# 400259; RRID: AB_10895919
APC anti-human CD24 Antibody (clone ML5)	BioLegend	Cat# 311117; RRID: AB_1877150
APC Rat IgG2a, $\kappa$ Isotype Ctrl (clone RTK2758)	BioLegend	Cat# 400511; RRID: AB_2814702
<b>Bacterial and Virus Strains</b>		
AAV9-PB-SRT-tdTomato	Joseph D. Dougherty (Cammack et al., 2020)	N/A
AAV9-HyPBase	Joseph D. Dougherty (Cammack et al., 2020)	N/A
Lenti-dCas9-KRAB	This study	N/A
Lenti-BRD4-CRISPRi	This study	N/A
Lenti-NT-CRISPRi	This study	N/A
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
DMEM	GIBCO	Cat# 11965-084
Antibiotic-Antimycotic (100X)	GIBCO	Cat# 15240-062
FBS	Peak Serum	Cat# PS-FB3
RPMI 1640 Medium	GIBCO	Cat# 11875-085
Lipofectamine 3000 Transfection Reagent	Invitrogen	Cat# L3000015
Trypsin-EDTA solution	Sigma-Aldrich	Cat# T4049
DPBS, no calcium, no magnesium	GIBCO	Cat# 14190-136
RNAprotect Cell Reagent	QIAGEN	Cat# 76526
2-Mercaptoethanol	GIBCO	Cat# 21985-023
RNase-Free DNase Set	QIAGEN	Cat# 79254
Maxima H Minus Reverse Transcriptase	Thermo Scientific	Cat# EP0752
Advantage® UltraPure PCR Deoxynucleotide Mix	Takara Bio	Cat# 639125
RNaseOUT Recombinant Ribonuclease Inhibitor	Invitrogen	Cat# 10777019
TransIT®-LT1 Transfection Reagent	Mirus	Cat# MIR2304
RNase H	New England BioLabs	Cat# M0297S
HiFi HotStart ReadyMix (2X)	Kapa Biosystems	Cat# KK2601
AMPure XP beads	Beckman Coulter	Cat# A63880
Puromycin dihydrochloride	Sigma-Aldrich	Cat# P8833
Crystal violet	Sigma-Aldrich	Cat# C0775
Methanol	Fisher Scientific	Cat# A452-4
Formaldehyde	Fisher Scientific	Cat# BP531-500
High Sensitivity D1000 Reagents	Agilent	Cat# 5067-5585
Ficoll PM400 (Dry Powder)	GE Healthcare	Cat# 17030010
NxGen® RNase Inhibitor	Lucigen	Cat# 30281-1
Dynabeads MyOne Silane	Life Technologies	Cat# 37002D
IDTE pH 8.0 (1X TE Solution)	IDT	Cat# 11-05-01-13
High Sensitivity D5000 Reagents	Agilent	Cat# 5067-5593
NEBuffer 2	New England BioLabs	Cat# B7002S
Buffer EB	QIAGEN	Cat# 19086
Hibernate-A Medium	GIBCO	Cat# A1247501

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
D-(+)-Trehalose dihydrate	Sigma-Aldrich	Cat# T9531
B-27 Supplement (50X), serum free	GIBCO	Cat# 17504044
0.5M EDTA, pH 8.0	Corning	Cat# 46-034-CI
Papain, Lyophilized	Worthington Biochemical	Cat# LS003118
Deoxyribonuclease I, Filtered	Worthington Biochemical	Cat# LS002060
Trypsin Inhibitor, Ovomuroid	Worthington Biochemical	Cat# LS003087
Bovine Serum Albumin	Sigma-Aldrich	Cat# A9418
OptiPrep Density Gradient Medium	Sigma-Aldrich	Cat# D1556
HBSS (10X)	GIBCO	Cat# 14185052
Magnesium chloride	Sigma-Aldrich	Cat# M4880
Magnesium sulfate	Sigma-Aldrich	Cat# M2643
Calcium chloride dihydrate	Sigma-Aldrich	Cat# C7902
D-(+)-Glucose	Sigma-Aldrich	Cat# G7021
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich	Cat# D2650
Cell Staining Buffer	BioLegend	Cat# 420201
Annexin V Binding Buffer	BioLegend	Cat# 422201
SuperScript VIL0 cDNA Synthesis Kit	Invitrogen	Cat# 11754250
PowerUp SYBR Green Master Mix	Applied Biosystems	Cat# 25742
(+)-JQ1	Selleck Chemicals	Cat# S7110
Propidium iodide (PI)	Invitrogen	Cat# P3566
Hoechst 33342	Thermo Scientific	Cat# 62249
Blasticidin S HCl	GIBCO	Cat# A1113903
Lenti-X Concentrator	Takara Bio	Cat# 631232
Lipofectamine 2000 Transfection Reagent	Invitrogen	Cat# 11668030
Polybrene Infection / Transfection Reagent	Sigma-Aldrich	Cat# TR-1003
Esp3l	New England BioLabs	Cat# R0734S
T4 DNA Ligase	New England BioLabs	Cat# M0202S
IMDM	GIBCO	Cat# 12440046
Penicillin-Streptomycin (10,000 U/mL)	GIBCO	Cat# 15140122
Imatinib mesylate	Sigma-Aldrich	Cat# SML1027
Lovastatin	Sigma-Aldrich	Cat# M2147
Nocodazole	Sigma-Aldrich	Cat# M1404
CVT-313	Sigma-Aldrich	Cat# 238803
RO-3306	Sigma-Aldrich	Cat# SML0569
Annexin V-FITC	BioLegend	Cat# 640905
<b>Critical Commercial Assays</b>		
Neon Transfection System 100 $\mu$ L Kit	Invitrogen	Cat# MPK10025
RNeasy Plus Mini Kit	QIAGEN	Cat# 74134
Qubit RNA HS Assay Kit	Invitrogen	Cat# Q32852
Qubit dsDNA HS Assay Kit	Invitrogen	Cat# Q32851
Nextera XT DNA Library Preparation Kit	Illumina	Cat# FC-131-1024
High Sensitivity D1000 ScreenTape	Agilent	Cat# 5067-5584
Chromium Single Cell 3' Library & Gel Bead Kit v2	10x Genomics	Cat# PN-120267
High Sensitivity D5000 ScreenTape	Agilent	Cat# 5067-5592
Nextera Mate Pair Library Prep Kit	Illumina	Cat# FC-132-1001
<b>Deposited Data</b>		
K562 CpG islands	Richard Myers	GEO: GSM1014203
HCT-116 SP1 ChIP-seq	Richard Myers	ENCODE: ENCF000PCT

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HCT-116 CTCF ChIP-seq	Richard Myers	ENCODE: ENCFF000OZC
HCT-116 ChIP-seq input control (SP1, CTCF)	Richard Myers	ENCODE: ENCFF000PBO
HCT-116 BRD4 ChIP-seq	Ron Firestein	SRA: SRR2481799
HCT-116 ChIP-seq input control (BRD4)	Ron Firestein	SRA: SRR2481800
HCT-116 H3K27ac ChIP-seq	Bradley Bernstein	ENCODE: ENCFF082JPN, ENCFF176BXC
HCT-116 H3K4me1 ChIP-seq	Bradley Bernstein	ENCODE: ENCFF088BWP, ENCFF804MJI
HCT-116 H3K4me2 ChIP-seq	Bradley Bernstein	ENCODE: ENCFF936MMN, ENCFF937OOL
HCT-116 H3K4me3 ChIP-seq	Bradley Bernstein	ENCODE: ENCFF183OZI, ENCFF659FPR
HCT-116 H3K9me2 ChIP-seq	Bradley Bernstein	ENCODE: ENCFF760OZN, ENCFF565FDP
HCT-116 H3K9me3 ChIP-seq	Bradley Bernstein	ENCODE: ENCFF578MDZ, ENCFF033XOG
HCT-116 H3K27me3 ChIP-seq	Bradley Bernstein	ENCODE: ENCFF281SBT, ENCFF124GII
HCT-116 H3K36me3 ChIP-seq	Bradley Bernstein	ENCODE: ENCFF850EAH, ENCFF312RKB
HCT-116 H3K79me2 ChIP-seq	Bradley Bernstein	ENCODE: ENCFF865KPW, ENCFF947YPU
HCT-116 H4K20me1 ChIP-seq	Bradley Bernstein	ENCODE: ENCFF070JDY, ENCFF334HHB
HCT-116 ChIP-seq input control (H3K27ac, H3K4me1, H3K4me2, H3K4me3, H3K9me2, H3K9me3, H3K27me3, H3K36me3, H3K79me2, H4K20me1)	Bradley Bernstein	ENCODE: ENCFF048ZQO, ENCFF827YXC
HCT-116 H3K9ac ChIP-seq	Bradley Bernstein	ENCODE: ENCFF408RRT
HCT-116 ChIP-seq input control (H3K9ac)	Bradley Bernstein	ENCODE: ENCFF413RQG
K562 BRD4 ChIP-seq	Bradley Bernstein	ENCODE: ENCFF335PHG
K562 H3K27ac ChIP-seq	Bradley Bernstein	ENCODE: ENCFF000BXH
K562 ChIP-seq input control (BRD4, H3K27ac)	Bradley Bernstein	ENCODE: ENCFF000BWK
K562 SP1 ChIP-seq	Michael Snyder	ENCODE: ENCFF002DPL, ENCFF002EGC
K562 ChIP-seq input control (SP1)	Michael Snyder	ENCODE: ENCFF002EGI, ENCFF002EGA
HepG2 FOXA2 ChIP-seq	Richard Myers	ENCODE: ENCFF000PIX
HepG2 ChIP-seq input control (FOXA2)	Richard Myers	ENCODE: ENCFF000POV
OCM-1A HyPBase DNA calling cards	Michael Onken	<a href="https://doi.org/10.1186/s12920-018-0424-0">https://doi.org/10.1186/s12920-018-0424-0</a>
OCM-1A BAP1-HyPBase DNA calling cards	Michael Onken	<a href="https://doi.org/10.1186/s12920-018-0424-0">https://doi.org/10.1186/s12920-018-0424-0</a>
OCM-1A RNA-seq (BAP1 and control shRNA)	Michael Onken	GEO: GSE110193
Mouse cortex H3K27ac ChIP-seq	Michael Greenberg	SRA: SRR6129714
Mouse cortex ChIP-seq input control (H3K27ac)	Michael Greenberg	SRA: SRR6129695
K562 RNA Pol II ChIA-PET	Yijun Ruan	ENCODE: ENCFF000KYH
HCT-116 DNase-seq	John Stamatoyannopoulos	ENCODE: ENCFF001DCK
HCT-116 ATAC-seq	Sriharsa Pradhan	SRA: SRR5453778
HCT-116 ATAC-seq control	Michael Guertin	GEO: GSE92674
HCT-116 CpG islands	Richard Myers	GEO: GSM1014209
Sequencing data and processed output	This study	GEO: GSE148448
<b>Experimental Models: Cell Lines</b>		
Neuro-2a (N2a)	ATCC	Cat# CCL-131
K-562	ATCC	Cat# CCL-243
Hep G2	ATCC	Cat# HB-8065
OCM-1A	Michael Onken ( <a href="#">Yen et al., 2018</a> )	N/A
HCT 116	ATCC	Cat# CCL-247
293T/17 [HEK293T/17]	ATCC	Cat# CRL-11268
<b>Experimental Models: Organisms/Strains</b>		
Mouse: C57BL/6J	Joseph D. Dougherty ( <a href="#">Cammack et al., 2020</a> )	N/A

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
Primers and oligonucleotides	This study, see <a href="#">Table S4</a>	N/A
Recombinant DNA		
pRM1024: PBase	This study	N/A
pRM1114: HyPBase	This study	N/A
pRM1023: SP1-PBase	This study	N/A
pRM1677: SP1-HyPBase	This study	N/A
pRM1882: FOXA2-HyPBase	This study	N/A
pRM1863: BAP1-HyPBase	This study	N/A
pRM1304: PB-SRT-Puro	This study	RRID: Addgene_154884
pRM1535: PB-SRT-tdTomato	This study	RRID: Addgene_154885
pCMV(CAT)T7-SB100	Zsuzsanna Izsvák	RRID: Addgene_34879
pRM1665: SP1-SB100X	This study	RRID: Addgene_154887
pRM1668: SB-SRT-Puro	This study	RRID: Addgene_154888
pRM1217: AAV-HyPBase	Joseph D. Dougherty ( <a href="#">Cammack et al., 2020</a> )	N/A
pRM1648: AAV-PB-SRT-tdTomato	Joseph D. Dougherty ( <a href="#">Cammack et al., 2020</a> )	RRID: Addgene_154889
pUC19 Vector	New England BioLabs	Cat# N3041S
Lenti-dCas9-KRAB-blast	Gary Hon	RRID: Addgene_89567
sgOpti	Eric Lander & David Sabatini	RRID: Addgene_85681
pMD2.G	Didier Trono	RRID: Addgene_12259
psPAX2	Didier Trono	RRID: Addgene_12260
pRM1889: BRD4 CRISPRi plasmid	This study	RRID: Addgene_154890
pRM1890: Non-targeting CRISPRi plasmid	Robi D. Mitra ( <a href="#">Lalli et al., 2019</a> )	RRID: Addgene_154891
Software and Algorithms		
cutadapt 1.16	<a href="#">Martin, 2011</a>	RRID: SCR_011841
NovoAlign 3	Novocraft Technologies	RRID: SCR_014818
Cell Ranger 2.1.0	10x Genomics	RRID: SCR_017344
scanpy 1.3.7	<a href="#">Wolf et al., 2018</a>	RRID: SCR_018139
Drop-seq tools 1.11	<a href="#">Macosko et al., 2015</a>	RRID: SCR_018142
astropy 3.2.1	<a href="#">Robitaille et al., 2013</a>	RRID: SCR_018148
WashU Human Epigenome Browser 46	<a href="#">Zhou et al., 2011</a>	RRID: SCR_006208
MEME-ChIP 4.11.2	<a href="#">Machanic and Bailey, 2011</a>	RRID: SCR_001783
Tomtom 5.1.0	<a href="#">Gupta et al., 2007</a>	RRID: SCR_001783
MACS 1.4.1	<a href="#">Zhang et al., 2008</a>	RRID: SCR_013291
BEDTools 2.27.1	<a href="#">Quinlan and Hall, 2010</a>	RRID: SCR_006646
NumPy 1.17.2	<a href="#">Oliphant, 2015</a>	RRID: SCR_008633
SciPy 1.4.1	<a href="#">Virtanen et al., 2020</a>	RRID: SCR_008058
statsmodels 0.10.1	<a href="#">Seabold and Perktold, 2010</a>	RRID: SCR_016074
matplotlib 3.0.3	<a href="#">Hunter, 2007</a>	RRID: SCR_008624
deeptools 3.0.1	<a href="#">Ramírez et al., 2016</a>	RRID: SCR_016366
ChromHMM 1.15	<a href="#">Ernst et al., 2011</a>	RRID: SCR_018141
liftOver	<a href="#">Hinrichs et al., 2006</a>	RRID: SCR_018160
FlowCal 1.2.0	<a href="#">Castillo-Hair et al., 2016</a>	RRID: SCR_018140
PANTHER 14.0	<a href="#">Mi et al., 2017</a>	RRID: SCR_004869

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ROSE 0.1	Whyte et al., 2013 & Lovén et al., 2013	RRID: SCR_017390
FlowJo Software for Mac Version 10	Becton, Dickson and Company	RRID: SCR_008520
Multcomp 1.4-12	Hothorn et al., 2008	RRID: SCR_018255
Custom calling card code	This study	<a href="https://github.com/arnavm/calling_cards">https://github.com/arnavm/calling_cards</a>
<b>Other</b>		
Qubit® 3.0 Fluorometer	Thermo Fisher	Cat# Q33216
4200 TapeStation System	Agilent	Cat# G2991AA
E220 Focused-ultrasonicator	Covaris	N/A
MasterCycler Pro PCR System	Eppendorf	Cat# 950030010
Attune NxT Flow Cytometer	Thermo Fisher	N/A
CytoFLEX S	Beckman-Coulter	Cat# B75442
QuantStudio	Applied Biosystems	Cat# A28567
Protocol: Mammalian Calling Cards Quick Start Guide	This study	<a href="https://doi.org/10.17504/protocols.io.xurfnv6">https://doi.org/10.17504/protocols.io.xurfnv6</a>
Protocol: Bulk Calling Cards Library Preparation	This study	<a href="https://doi.org/10.17504/protocols.io.xwhfpb6">https://doi.org/10.17504/protocols.io.xwhfpb6</a>
Protocol: Single Cell Calling Cards Library Preparation	This study	<a href="https://doi.org/10.17504/protocols.io.xwifpce">https://doi.org/10.17504/protocols.io.xwifpce</a>
Protocol: Processing Bulk Calling Card Sequencing Data	This study	<a href="https://doi.org/10.17504/protocols.io.xwjfpcn">https://doi.org/10.17504/protocols.io.xwjfpcn</a>
Protocol: Processing Single Cell Calling Card Sequencing Data	This study	<a href="https://doi.org/10.17504/protocols.io.4phgvj6">https://doi.org/10.17504/protocols.io.4phgvj6</a>
Protocol: Calling Peaks on <i>piggyBac</i> Calling Card Data	This study	<a href="https://doi.org/10.17504/protocols.io.bb9xir7n">https://doi.org/10.17504/protocols.io.bb9xir7n</a>
Protocol: Visualizing Calling Card Data on the WashU Epigenome Browser	This study	<a href="https://doi.org/10.17504/protocols.io.bca8ishw">https://doi.org/10.17504/protocols.io.bca8ishw</a>

**RESOURCE AVAILABILITY**

**Lead Contact**

Further information and requests for resources and reagents should be directed to, and will be fulfilled by, the Lead Contact, Robi D. Mitra ([rmitra@wustl.edu](mailto:rmitra@wustl.edu)).

**Materials Availability**

Plasmids generated in this study have been deposited to Addgene, where possible, and are available to the community. Plasmids encoding the *piggyBac* transposase are not available through Addgene due to licensing restrictions. These plasmids are available upon request to the Lead Contact.

**Data and Code Availability**

Data generated in this study have been submitted to the Gene Expression Omnibus (GEO) with accession number GSE148448. All code used to analyze the data is available online at [https://github.com/arnavm/calling\\_cards](https://github.com/arnavm/calling_cards).

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

HCT-116, N2a, HEK293T, and HepG2 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% of either penicillin-streptomycin or antibiotic-antimycotic. K562 (unless otherwise indicated) and OCM-1A cells were grown under the same conditions as described above, replacing DMEM with RPMI 1640 Medium. Cells were grown at 37°C with 5% carbon dioxide (CO<sub>2</sub>). Media was replenished every 2 days. HepG2 cells were a gift from the Genome Engineering iPSC Center (GEIC) at Washington University in St. Louis School of Medicine. OCM-1A cells were a gift from Dr. Michael Onken. For the CD24<sup>high</sup>/CD24<sup>low</sup> cell state analyses, K562 cells were grown in IMDM containing 10% v/v FBS and 1% penicillin-streptomycin at 37°C with 5% CO<sub>2</sub>. Frozen aliquots were thawed and passaged every 48 hours until they reached a maximum

concentration of 800,000 cells/ml. For experiments, cells were seeded at mid-log phase concentrations, around 400,000 cells/ml. At this point, ratio of CD24<sup>high</sup>/CD24<sup>low</sup> cells was approximately 1:1, as determined by flow cytometry.

All mouse experiments were done following procedures described in (Cammack et al., 2020). In brief, we cloned the PB-SRT-tdTomato and HyPBase constructs into AAV vectors. The Hope Center Viral Vectors Core at Washington University in St. Louis packaged each construct in AAV9 capsids. Titers for each virus ranged between  $1.1 \times 10^{13}$  and  $2.2 \times 10^{13}$  viral genomes/ml. We mixed equal volumes of each virus and performed intracranial cortical injections of the mixture into newborn wild-type C57BL/6J pups (P0-2). As a gating control, we injected one litter-matched animal with AAV9-PB-SRT-tdTomato only. After 2 to 4 weeks, we sacrificed mice and dissected the cortex (8 libraries) or hippocampus (1 library). The sex of mice was not taken into consideration. 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.

## METHOD DETAILS

### DNA- versus RNA-based recovery

Approximately 500,000 HCT-116 cells were plated in a single well of a 6-well plate. Cells were transfected with 2.5  $\mu$ g of the SP1-PBase plasmid and 2.5  $\mu$ g of the PB-SRT-Puro plasmid using Lipofectamine 3000 following manufacturer's instructions. After 24 hours, cells were split and plated 1:10 in each of three 10 cm dishes. Puromycin was then added to a final concentration of 2  $\mu$ g/ml and colonies were grown under selection for two weeks. We obtained approximately 2,300 colonies. All cells were pooled together and split into two populations. One half was subjected to DNA extraction, self-ligation, and inverse PCR, as described previously (Wang et al., 2012a), with the following modification: digestion with MspI was not performed as the SRT construct contained an second MspI cut site near the terminal repeat. The other half of cells underwent RNA extraction and SRT library preparation (see below).

### In vitro bulk calling card experiments

We cotransfected 10-12 replicates of HCT-116 cells with 5  $\mu$ g of PB-SRT-Puro plasmid and 5  $\mu$ g PBase plasmid via Neon electroporation. Each replicate contained  $2 \times 10^6$  cells. As a negative control, we transfected one replicate of HCT-116 cells with 5  $\mu$ g PB-SRT-Puro plasmid only. We used the following settings—pulse voltage: 1,530 V; pulse width: 20 ms; pulse number: 1. We used the same experimental setup for experiments with PB-SRT-Puro and each of SP1-PBase, HyPBase, and SP1-HyPBase plasmids, as well as with SB-SRT-Puro and SB100X (the latter a gift from Dr. Zsuzsanna Izsák; Mátés et al., 2009) plasmids. After transfection, each replicate was plated into a 10 cm dish. For the OCM-1A library, we transfected 1.25  $\mu$ g of PB-SRT-Puro and 1.25  $\mu$ g of either HyPBase or BAP1-HyPBase (the latter a gift from Dr. Michael Onken; Yen et al., 2018) using the *TransIT-LT1* transfection reagent following manufacturer's protocol for 6-well plates. Puromycin was added after 24 hours to a final concentration of 2  $\mu$ g/ml. Cells were grown under selection for one week, by which time almost all negative control transfectants were dead. After 7 days, we dissociated each replicate with trypsin-EDTA and created single cell suspensions in phosphate-buffered saline (PBS). Aliquots of each replicate were cryopreserved in cell culture media (see above) supplemented with 5% DMSO. The remaining cells were pelleted by centrifugation at 300g for 5 minutes. Cell pellets were either processed immediately or kept at  $-80^\circ\text{C}$  in RNAProtect Cell Reagent.

### Isolation and RT of bulk RNA

Total RNA was isolated from each replicate using the RNEasy Plus Mini Kit following manufacturer's instructions. Briefly, cell pellets were resuspended in 600  $\mu$ l of Buffer RLT Plus with 1% 2-mercaptoethanol. Cells were homogenized by vortexing. DNA was removed by running lysate through gDNA Eliminator spin columns, while RNA was bound by passing the flow-through over RNEasy spin columns. An on-column treatment with DNaseI was also performed. After washing, RNA was eluted in 40  $\mu$ l RNase-free H<sub>2</sub>O. RNA was quantitated using the Qubit RNA HS Assay Kit.

We performed first strand synthesis on each replicate with Maxima H Minus Reverse Transcriptase. We mixed 2  $\mu$ g of total RNA with 1  $\mu$ l 10 mM dNTPs and 1  $\mu$ l of 50  $\mu$ M SMART\_dT18VN primer (for a complete list of oligonucleotides, see Table S4), brought the total volume up to 14  $\mu$ l, and incubated it at  $65^\circ\text{C}$  for 5 minutes. After transferring to ice and letting rest for 1 minute, we added 4  $\mu$ l 5X Maxima RT Buffer, 1  $\mu$ l RNaseOUT, and 1  $\mu$ l of 1:1 Maxima H Minus Reverse Transcriptase diluted in 1x RT Buffer (100 U). The solution was mixed by pipetting and incubated at  $50^\circ\text{C}$  for 1 hour followed by heat inactivation at  $85^\circ\text{C}$  for 10 minutes. Finally, we digested with 1  $\mu$ l RNaseH at  $37^\circ\text{C}$  for 30 minutes. cDNA was stored at  $-20^\circ\text{C}$ .

### Amplifying self-reporting transcripts from RNA

The PCR conditions for amplifying self-reporting transcripts (i.e., transcripts derived from self-reporting transposons) involved mixing 1  $\mu$ l cDNA template with 12.5  $\mu$ l Kapa HiFi HotStart ReadyMix, 0.5  $\mu$ l 25  $\mu$ M SMART primer, and either 1  $\mu$ l of 25  $\mu$ M SRT\_PAC\_F1 primer (in the case of puromycin selection) or 0.5  $\mu$ l of 25  $\mu$ M SRT\_tdTomato\_F1 primer (in the case of tdTomato screening). The mixture was brought up to 25  $\mu$ l with ddH<sub>2</sub>O. Thermocycling parameters were as follows:  $95^\circ\text{C}$  for 3 minutes; 20 cycles of:  $98^\circ\text{C}$  for 20 s— $65^\circ\text{C}$  for 30 s— $72^\circ\text{C}$  for 5 minutes;  $72^\circ\text{C}$  for 10 minutes; hold at  $4^\circ\text{C}$  forever. As a control, cDNA quality can be assessed with exon-spanning primers for  $\beta$ -actin [see Table S4 for examples of human primers (Raff et al., 1997)] under the same thermocycling settings.

PCR products were purified using AMPure XP beads. 12  $\mu$ l of resuspended beads were added to the 25  $\mu$ l PCR product and mixed homogeneously by pipetting. After a 5-minute incubation at room temperature, the solution was placed on a magnetic rack for 2 minutes. The supernatant was aspirated and discarded. The pellet was washed twice with 200  $\mu$ l of 70% ethanol (incubated for 30 s each time), discarding the supernatant each time. The pellet was left to dry at room temperature for 2 minutes. To elute, we added 20  $\mu$ l ddH<sub>2</sub>O to the pellet, resuspended by pipetting, incubated at room temperature for 2 minutes, and placed on a magnetic rack for one minute. Once clear, the solution was transferred to a clean 1.5 mL tube. DNA concentration was measured on the Qubit 3.0 Fluorometer using the dsDNA High Sensitivity Assay Kit.

### Generation of bulk RNA calling card libraries

Calling card libraries from bulk RNA were generated using the Nextera XT DNA Library Preparation Kit. One nanogram of PCR product was resuspended in 5  $\mu$ l ddH<sub>2</sub>O. To this mixture we added 10  $\mu$ l Tagment DNA (TD) Buffer and 5  $\mu$ l Amplicon Tagment Mix (ATM). After pipetting to mix, we incubated the solution in a thermocycler preheated to 55°C. The tagmentation reaction was halted by adding 5  $\mu$ l Neutralization Tagment (NT) Buffer and was kept at room temperature for 5 minutes. The final PCR was set up by adding 15  $\mu$ l Nextera PCR Mix (NPM), 8  $\mu$ l ddH<sub>2</sub>O, 1  $\mu$ l of 10  $\mu$ M transposon primer (e.g., OM-PB-NNN) and 1  $\mu$ l Nextera N7 indexed primer. The transposon primer anneals to the end of the transposon terminal repeat—*piggyBac*, in the case of OM-PB primers, or *Sleeping Beauty*, in the case of OM-SB primers—and contains a 3 base pair barcode sequence. Every N7 primer contains a unique index sequence that is demultiplexed by the sequencer. Each replicate was assigned a unique combination of barcoded transposon primer and indexed N7 primer, enabling precise identification of each library's sequencing reads.

The final PCR was run under the following conditions: 95°C for 30 s; 13 cycles of: 95°C for 10 s—50°C for 30 s—72°C for 30 s; 72°C for 5 minutes; hold at 4°C forever. After PCR, the final library was purified using 30  $\mu$ l (0.6x) AMPure XP beads, as described above. The library was eluted in 11  $\mu$ l ddH<sub>2</sub>O and quantitated on an Agilent TapeStation 4200 System using the High Sensitivity D1000 ScreenTape.

### Colony formation assay

For the *piggyBac* transfectants, we electroporated 500,000 HCT-116 cells with 750 ng of PB-SRT-Puro and 750 ng of either HyPBase or SP1-HyPBase plasmid using the aforementioned Neon settings. For the SRT-only conditions, cells received 750 ng of PB-SRT-Puro. We used the same design for the *Sleeping Beauty* transfectants, replacing the transposases with either SB100X or SP1-SB100X and using SB-SRT-Puro as the transposon. Each set of 500,000 cells were plated in a single well of a 6-well plate and allowed to recover overnight. We then added puromycin to a final concentration of 2  $\mu$ g/ml. We cultured cells under selection for one week. Colonies were visualized using a solution comprising 1X PBS, 1% formaldehyde, 1% methanol, and 0.05% w/v crystal violet. After aspirating medium, we covered cells with this solution, staining plates for 20 minutes washing under cold water and air drying.

### In vitro single cell calling card experiments

All cell lines (HCT-116, K562, N2a, HepG2, and OCM-1A) were cultured as described above. HCT-116 cells were transfected using Neon electroporation with the aforementioned settings. K562 cells were electroporated with the following settings—pulse voltage: 1,450 V; pulse width: 10 ms; pulse number: 3. N2a cells were electroporated with the following settings—pulse voltage: 1,050 V; pulse width: 30 ms; pulse number: 2. HepG2 cells were electroporated with the following settings—pulse voltage: 1,200 V; pulse width: 50 ms; pulse number: 1. Each replicate for electroporation was comprised of  $2 \times 10^6$  cells. All cells were allowed to recover for 24 hours before undergoing puromycin selection. A negative control replicate, transfected only with PB-SRT-Puro, was treated identically in parallel. Replicates were harvested once the negative control cells had died. For the species mixing experiment, we transfected one replicate each of HCT-116 and N2a cells with 5  $\mu$ g PB-SRT-Puro and 5  $\mu$ g HyPBase. For the cell line mixing experiment, we transfected four replicates each of HCT-116 and K562 cells with 5  $\mu$ g PB-SRT-Puro and 5  $\mu$ g HyPBase. In all cases, cells were cultured independently and mixed immediately prior to generating single cell emulsions. For single cell calling cards analysis of SP1 binding in HCT-116 and K562 cells, we transfected four replicates each with 5  $\mu$ g PB-SRT-Puro and 5  $\mu$ g SP1-HyPBase. These libraries were not mixed. We used the demultiplexed data from the cell line mixing experiment with HyPBase as controls. For single cell calling cards analysis of FOXA2 binding in HepG2 cells, we transfected six replicates each with 5  $\mu$ g PB-SRT-Puro; three of these replicates were co-transfected with 5  $\mu$ g HyPBase, while the other three were co-transfected with 5  $\mu$ g FOXA2-HyPBase. We used the mouse ortholog of FOXA2, which has 97% primary sequence identity with human FOXA2. For single cell calling cards analysis of BAP1 binding in OCM-1A cells, we lipofected (as described above) six replicates each with 1.25  $\mu$ g PB-SRT-Puro; three of these replicates were co-transfected with 1.25  $\mu$ g HyPBase, while the other three were co-transfected with 1.25  $\mu$ g BAP1-HyPBase.

### Single cell RNA-seq library preparation

Single cell RNA-seq libraries were prepared using 10x Genomics' Chromium Single Cell 3' Library and Gel Bead Kit. Each replicate was targeted for recovery of 6,000 cells. Library preparation followed a modified version of the manufacturer's protocol. We prepared the Single Cell Master Mix without RT Primer, replacing it with an equivalent volume of Low TE Buffer. Gel-in-emulsion (GEM) generation and GEM-RT incubation proceeded as instructed. At the end of Post GEM-RT cleanup, we added 36.5  $\mu$ l Elution Solution I and transferred 36  $\mu$ l of the eluted sample to a new tube (instead of 35.5  $\mu$ l and 35  $\mu$ l, respectively). The eluate was split into two 18  $\mu$ l

aliquots and kept at  $-20^{\circ}\text{C}$  until ready for further processing. One fraction was kept for single cell calling cards library preparation (see next section), while the other half was further processed into a single cell RNA-seq library.

We then added the RT Primer sequence to the products in the scRNA-seq aliquot. We created an RT master mix by adding 20  $\mu\text{l}$  of Maxima 5X RT Buffer, 20  $\mu\text{l}$  of 20% w/v Ficoll PM-400, 10  $\mu\text{l}$  of 10 mM dNTPs, 2.5  $\mu\text{l}$  RNase Inhibitor and 2.5  $\mu\text{l}$  of 100  $\mu\text{M}$  10x\_TSO. To this solution we added 18  $\mu\text{l}$  of the first RT product and 22  $\mu\text{l}$  of ddH<sub>2</sub>O. Finally, we added 5  $\mu\text{l}$  Maxima H Minus Reverse Transcriptase, mixed by flicking, and centrifuged briefly. This reaction was incubated at  $25^{\circ}\text{C}$  for 30 minutes followed by  $50^{\circ}\text{C}$  for 90 minutes and heat inactivated at  $85^{\circ}\text{C}$  for 5 minutes.

The solution was purified using DynaBeads MyOne Silane following 10x Genomics' instructions, beginning at "Post GEM-RT Cleanup – Silane DynaBeads" step D. The remainder of the single cell RNA-seq protocol, including purification, amplification, fragmentation, and final library amplification, followed manufacturer's instructions.

### Single cell calling cards library preparation

To amplify self-reporting transcripts from single cell RNA-seq libraries, we took 9  $\mu\text{l}$  of RT product (the other half was kept in reserve) and added it to 25  $\mu\text{l}$  Kapa HiFi HotStart ReadyMix and 15  $\mu\text{l}$  ddH<sub>2</sub>O. We then prepared a PCR primer cocktail comprising 5  $\mu\text{l}$  of 100  $\mu\text{M}$  Bio\_Illumina\_Seq1\_scCC\_10X\_3xPT primer, 5  $\mu\text{l}$  of 100  $\mu\text{M}$  Bio\_Long\_PB\_LTR\_3xPT, and 10  $\mu\text{l}$  of 10 mM Tris-HCl, 0.1 mM EDTA buffer. One  $\mu\text{l}$  of this cocktail was added to the PCR mixture and placed in a thermocycler. Thermocycling settings were as follows:  $98^{\circ}\text{C}$  for 3 minutes; 20-22 cycles of  $98^{\circ}\text{C}$  for 20 s– $67^{\circ}\text{C}$  for 30 s– $72^{\circ}\text{C}$  for 5 minutes;  $72^{\circ}\text{C}$  for 10 minutes;  $4^{\circ}\text{C}$  forever. PCR purification was performed with 30  $\mu\text{l}$  AMPure XP beads (0.6x ratio) as described previously. The resulting library was quantitated on an Agilent TapeStation 4200 System using the High Sensitivity D5000 ScreenTape.

Single cell calling card library preparation was performed using the Nextera Mate Pair Sample Prep Kit with modifications to the manufacturer's protocol. The library was circularized by bringing 300 fmol (approximately 200 ng) of DNA up to a final volume of 268  $\mu\text{l}$  with ddH<sub>2</sub>O, then adding 30  $\mu\text{l}$  Circularization Buffer 10x and 2  $\mu\text{l}$  Circularization Ligase (final concentration: 1 nM). This reaction was incubated overnight (12-16 hours) at  $30^{\circ}\text{C}$ . After removal of linear DNA (following manufacturer's instructions), we sheared the library on a Covaris E220 Focused-ultrasonicator with the following settings—peak power intensity: 200; duty factor: 20%; cycles per burst: 200; time: 40 s; temperature:  $6^{\circ}\text{C}$ .

The library preparation was performed per manufacturer's instructions until adaptor ligation. We designed custom adapters (Table S4) so that the standard Illumina sequencing primers would not interfere with our library. Adapters were prepared by combining 4.5  $\mu\text{l}$  of 100  $\mu\text{M}$  scCC\_P5\_adapter, 4.5  $\mu\text{l}$  of 100  $\mu\text{M}$  scCC\_P7\_adapter, and 1  $\mu\text{l}$  of NEBuffer 2, then heating in a thermocycler at  $95^{\circ}\text{C}$  for 5 minutes, then holding at  $70^{\circ}\text{C}$  for 15 minutes, then ramping down at 1% until it reached  $25^{\circ}\text{C}$ , holding at that temperature for 5 minutes, before keeping at  $4^{\circ}\text{C}$  forever. One microliter of this custom adaptor mix was used in place of the manufacturer's recommended DNA Adaptor Index. The ligation product was cleaned per manufacturer's instructions. For the final PCR, the master mix was created by combining 20  $\mu\text{l}$  Enhanced PCR Mix with 28  $\mu\text{l}$  of ddH<sub>2</sub>O and 1  $\mu\text{l}$  each of 25  $\mu\text{M}$  scCC\_P5\_primer and 25  $\mu\text{M}$  scCC\_P7\_primer. This was then added to the streptavidin bead-bound DNA and amplified under the following conditions:  $98^{\circ}\text{C}$  for 30 s; 15 cycles of:  $98^{\circ}\text{C}$  for 10 s– $60^{\circ}\text{C}$  for 30 s– $72^{\circ}\text{C}$  for 2 minutes;  $72^{\circ}\text{C}$  for 5 minutes;  $4^{\circ}\text{C}$  forever. All of the PCR supernatant was transferred to a new tube and purified with 35  $\mu\text{l}$  (0.7x) AMPure XP beads following manufacturer's instructions. The final library was eluted in 25  $\mu\text{l}$  Elution Buffer and quantitated on an Agilent TapeStation 4200 System using the High Sensitivity D1000 ScreenTape.

### Staining protocols for K562 cells

CD24 surface protein was quantified using monoclonal human antibodies. Cells were spun down at 300g for 3 minutes and washed twice with 1 mL of Cell Staining Buffer. The cell pellet was then resuspended in 50  $\mu\text{l}$  of Cell Staining Buffer containing 0.2  $\mu\text{g}$  of either CD24-APC or CD24-BV421. The tube was rotated at  $4^{\circ}\text{C}$  in the dark for 30 minutes. After, cells were washed twice (as before) and finally resuspended in 200  $\mu\text{l}$  of Cell Staining Buffer. Cells were excited with 450/45 and 660/20 lasers (wavelength/filter bandwidth, both in nm). For concomitant analysis of DNA content, we used CD24-APC. Cells were incubated with 10  $\mu\text{g}/\text{ml}$  Hoechst 33342 in 5 mL of growth medium for 30 minutes prior to the staining protocol. For simultaneous assessment of apoptosis, cells were stained with CD24-BV421. After the final wash, instead of resuspending in 200  $\mu\text{l}$  of Cell Staining Buffer, cells were washed twice with Annexin V Staining Buffer. Cells were then incubated in 50  $\mu\text{l}$  Annexin V Staining Buffer containing 0.2  $\mu\text{g}$  Annexin V-FITC and 100  $\mu\text{g}/\text{ml}$  propidium iodide (PI). The reaction was incubated for 15 minutes at room temperature in the dark. Afterward, we added 150  $\mu\text{l}$  of Annexin V Staining Buffer and proceeded to flow cytometry. All samples were measured on a Beckman-Coulter CytoFLEX S flow cytometer. Cells were excited with 450/45, 525/40, and 610/20 lasers. We collected 10,000 events per sample. The resulting data were processed with FlowJo Software for Mac Version 10.

### JQ1 treatment of K562 cells

For the longitudinal treatment of K562 cells with JQ1, we seeded cells at log phase growth and treated them with growth medium containing DMSO ( $\sim 0.4\%$  final concentration) or 250 nM JQ1 (dissolved in DMSO). Medium was replaced every 48 hours without splitting. On days 1, 2, 3, 4, and 7, cells were split in half: one half was stained for CD24 and DNA content, while the other half was stained for CD24 and apoptosis (both described above). Experiments were performed with three biological replicates.

For qRT-PCR, we cultured K562 cells in either DMSO or 250 nM JQ1, in triplicate, and collected cells at 0, 3, 6, 9, 12, and 24 hours of treatment. Cells were pelleted, resuspended in 300  $\mu$ l of RNA CellProtect, and stored at  $-80^{\circ}\text{C}$ . When we were ready to extract RNA, we thawed cells, prepared samples using QIAGEN RNEasy Plus Mini Kit, and quantitated with the Qubit RNA High Sensitivity kit. We reverse transcribed 500 ng of RNA with the SuperScript VILO cDNA Synthesis Kit in a 20  $\mu$ l reaction, with the following thermocycling parameters:  $25^{\circ}\text{C}$  for 10 minutes;  $42^{\circ}\text{C}$  for 2 hours;  $85^{\circ}\text{C}$  for 5 minutes. We then performed PCR with 2  $\mu$ l of the RT product as template, 1  $\mu$ l each of forward and reverse primer (10  $\mu\text{M}$ ), 6  $\mu$ l ddH<sub>2</sub>O, and 10  $\mu$ l PowerUp SYBR Green Master Mix. We ran the PCR on an ABI QuantStudio 3 with the following settings: 2 minutes at  $50^{\circ}\text{C}$ , then 2 minutes at  $95^{\circ}\text{C}$  (hot start); 45 cycles of  $95^{\circ}\text{C}$  for 15 s followed by  $60^{\circ}\text{C}$  for 1 minute. We generated melt curves after each PCR and all samples yielded a single peak. Gene-specific primers were obtained from PrimerBank (Wang et al., 2012b). Data were normalized to the levels of  $\beta$ -actin.

### BRD4 CRISPRi of K562 cells

For CRISPRi, we first made lentivirus expressing dCas9-KRAB (Fulco et al., 2016; Xie et al., 2017) from Addgene plasmid #89567, a gift from Gary Hon, packaged in HEK293T cells along with pMD2.G (Addgene plasmid #12259) and psPAX2 (Addgene plasmid #12260), both gifts from Didier Trono. We cloned a *BRD4* guide RNA, selected from the Dolcetto collection (Sanson et al., 2018), into the sgOpti plasmid (Addgene plasmid #85681, a gift from Eric Lander & David Sabatini) using Golden Gate assembly with Esp3I. We used an in-house pipeline to design a non-targeting gRNA sequence, which was cloned into CROP-seq-opti (Lalli et al., 2019). Plasmids were transfected into HEK293T cells using Lipofectamine 2000. Media was collected after 24 and 48 hours, and subsequently concentrated using Lenti-X Concentrator. Viral titers were functionally assayed on HEK293T cells using the appropriate antibiotic (blasticidin or puromycin).

Next, we generated a polyclonal pool of dCas9-KRAB-expressing K562 cells. We seeded each well of a 6-well plate with 200,000 cells each containing 2 mL of growth media supplemented with 4  $\mu\text{g}/\text{ml}$  polybrene and 1,000,000 infectious lentiviral particles for an estimated multiplicity of infection (MOI) of 5. Plates were centrifuged at 2,000g for 30 minutes and returned to the incubator. After 48 hours, cells were split to mid-log phase concentration ( $\sim$ 400,000 cells/ml) and selected on blasticidin (10  $\mu\text{g}/\text{ml}$ ) for 48 hours. We made frozen stocks from these cells.

For the knockdown experiments, cells were thawed and allowed to recover for 4 days. We confirmed that the proportions of CD24<sup>high</sup>/CD24<sup>low</sup> was approximately equal at this point. We then seeded 200,000 cells into each well of a 6-well plate. Three wells received the *BRD4* gRNA lentivirus, while the other three received the non-targeting gRNA lentivirus, at MOI 2.5. We followed the same transduction protocol described above. After 48 hours of incubation, puromycin was added to the medium at a final concentration of 2  $\mu\text{g}/\text{ml}$ . After a further 48 hours, cells were passaged 1:1 into 10 cm dishes containing 10 mL of growth medium. The surviving cells were allowed to expand for a further 5 days before being stained for CD24 (nine days after gRNA transduction).

The *BRD4* gRNA was validated by performing qRT-PCR on RNA samples from treated cells with primers for either *BRD2*, *BRD3*, or *BRD4*, as described above.

### Imatinib treatments of K562 cells

Cells were challenged with imatinib either after JQ1 treatment or BRD4 CRISPRi. For the former, we plated 200,000 cells each well of a 6-well plate with 2 mL of growth medium. Half of the wells received DMSO while the other half received 250 nM JQ1. Cells were incubated for 5 days, with fresh media changes on days 1, 2, and 3. On day 5, a portion of each well was stained for CD24. The remaining cells in each well were split between two new wells. One well continued to receive medium supplemented with DMSO, while the other was treated with medium containing imatinib mesylate at a concentration of 1  $\mu\text{M}$ . After 48 hours, every well was stained for CD24 as well as annexin V and propidium iodide, for apoptotic activity. Cells undergoing BRD4 or non-targeted CRISPRi were split in two and treated with either DMSO or imatinib (1  $\mu\text{M}$ ) as described and in triplicate. The resulting data were processed with FlowJo. We set gates such that we could exclude debris but that we would capture both live and dying cells. This gate was used to calculate levels of annexin V and PI.

### Cell cycle perturbation of K562 cells

We perturbed the cell cycle with lovastatin and nocodazole, two drugs classically used to synchronize cells in culture (Jackman and O'Connor, 2001), as well as the cyclin-dependent kinase inhibitors CVT-313 (Brooks et al., 1997) and RO-3306 (Vassilev et al., 2006). All drugs were dissolved in DMSO except nocodazole, which was dissolved in ethanol. We treated 200,000 cells per well in 6-well plates with either DMSO, ethanol ( $\sim$ 0.4% final concentration), 250 nM JQ1, 12  $\mu\text{M}$  lovastatin, 40 ng/ $\mu$ l nocodazole (in ethanol), 2  $\mu\text{M}$  CVT-313, or 4.5  $\mu\text{M}$  RO-3306. Media was refreshed every 48 hours. After 36 hours of treatment, we stained for CD24 levels and nuclear DNA content. We gated for live, single cells using the forward scatter (FSC) and side scatter channels (SSC). Univariate cell cycle analysis was performed with FlowJo. We confirmed that all drugs perturbed cell cycle by altering the proportions of cells in either G1 or G2/M phase (Figure S6G). CVT-313 caused a significant increase in G1 arrest cells (one-way ANOVA  $p < 0.05$ ) and both nocodazole and RO-3306 caused significant G2 arrest (one-way ANOVA  $p < 0.01$ ). While lovastatin has been reported to arrest cells in G1, in our hands it caused a significant decrease in G1 phase K562 cells (one-way ANOVA  $p < 0.01$ ). Cultures remained under drug treatment until five days had elapsed, at which point we measured CD24 levels and stained for apoptosis (Figure S6H). As before, we set gates to exclude debris to quantitate annexin V and PI, and measured CD24 in live cells gated on FSC and SSC. The G2 inhibitors, in particular, had very few cells in the FSC/SSC gate (typically below 5%).

### SRT-tdTomato fluorescence validation

To test the fluorescence properties of the SRT-tdTomato construct, we transfected K562 cells as previously described with either 1  $\mu\text{g}$  of pUC19 plasmid; 0.5  $\mu\text{g}$  of PB-SRT-tdTomato plasmid and 0.5  $\mu\text{g}$  pUC19; 0.5  $\mu\text{g}$  of PB-SRT-tdTomato and 0.5  $\mu\text{g}$  pBase plasmid; and 0.5  $\mu\text{g}$  of PB-SRT-tdTomato and 0.5  $\mu\text{g}$  HyPBBase plasmid. Cells were allowed to expand for 8 days, after which fluorescence activity was assayed on an Attune NxT Flow Cytometer with an excitation wavelength of 561 nm. Flow cytometry data were visualized using FlowCal (Castillo-Hair et al., 2016). We also performed bulk RNA calling cards on HEK293T cells transfected with SRT-tdTomato with or without HyPBBase plasmid. While these cells were not sorted based on fluorescence activity, the SRT library from cells transfected with both SRT and transposase were more complex and contained many more insertions than the library from cells receiving SRT alone (Methods S1).

### In vivo scCC experiments

We separately packaged the PB-SRT-tdTomato and HyPBBase constructs in AAV9 viral particles (Cammack et al., 2020) and delivered mixtures of both viruses to the developing mouse cortex via intracranial injections at P1. After 2-4 weeks, we dissected the cortex, dissociated it to a single cell suspension, performed FACS to isolate tdTomato-positive cells, and generated both scRNA-seq and scCC libraries.

Mouse cortical tissues were dissociated to single suspensions following a modification of previously published methods (Avey et al., 2018; Saxena et al., 2012). We incubated samples in a papain solution containing Hibernate-A with 5% v/v trehalose, 1x B-27 Supplement, 0.7 mM EDTA, 70  $\mu\text{M}$  2-mercaptoethanol, and 2.8 mg/ml papain. After incubation at 37°C, cells were treated with DNaseI, triturated through increasingly narrow fire-polished pipettes, and passed through a 40-micron filter prewetted with resuspension solution: Hibernate-A containing 5% v/v trehalose, 0.5% Ovomucoid Trypsin Inhibitor, 0.5% Bovine Serum Albumin (BSA), 33  $\mu\text{g}/\text{ml}$  DNaseI (Worthington), and 1x B-27 Supplement. The filter was washed with 6 mL of resuspension solution. The resulting suspension was centrifuged for 4 minutes at 250 g. The supernatant was discarded. The pellet was then resuspended in 2 mL of resuspension solution and resuspended by gentle pipetting.

We eliminated subcellular debris using gradient centrifugation. We first prepared a working solution of 30% w/v OptiPrep Density Gradient Medium mixed with an equal volume of 1x Hank's Balanced Salt Solution (HBSS) with 0.5% BSA. We then prepared solutions of densities 1.057, 1.043, 1.036, and 1.029 g/ml using by combining the working solution with resuspension solution at ratios of 0.33:0.67, 0.23:0.77, 0.18:0.82, and 0.13:0.87, respectively. We layered 1 mL aliquots of each solution in a 15 mL conical tube beginning with the densest solution on the bottom. The cell suspension was added last to the tube and centrifuged for 20 minutes at 800g at 12°C. The top layer was then aspirated and purified cells were isolated from the remaining layers. These cells were then resuspended in FACS buffer: 1x HBSS, 2 mM  $\text{MgCl}_2$ , 2 mM  $\text{MgSO}_4$ , 1.25 mM  $\text{CaCl}_2$ , 1 mM D-glucose, 0.02% BSA, and 5% v/v trehalose. Cells were centrifuged for 4 minutes at 250 g, the supernatant was discarded, and the pellet was resuspended in FACS buffer by gentle pipetting.

Cells were then sorted based on fluorescence activity. As a gating control, we analyzed cells from cortices injected with AAV9-PB-SRT-tdTomato only. We then collected cells from brains transfected with AAV9-PB-SRT-tdTomato and AAV9-HyPBBase whose fluorescence values exceeded the gate. After sorting, cells were centrifuged for 3 minutes at 250 g. The supernatant was discarded and cells were resuspended in FACS buffer at a concentration appropriate for 10x Chromium 3' scRNA-seq library preparation.

### QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed in Python 3.7.3 using SciPy (Virtanen et al., 2020) and statsmodels (Seabold and Perktold, 2010) as well as R 3.5.3 using the multcomp package (Hothorn et al., 2008). Flow cytometry figures were created with FlowJo. All other figures were created with Python using matplotlib (Hunter, 2007). Statistical details for individual experiments have been provided in the main text, figure legends, and Method Details. In general, we used 10-12 replicates for bulk RNA calling cards experiments; at least three separate libraries for single cell calling cards experiments; and three biological replicates for the K562 cell state experiments.

### Interpreting calling card tracks

Calling card tracks depict recovered transposons as discrete data points. Each circle in the track is an independent transposition event whose genomic coordinate is along the x axis. The y axis is the number of reads supporting each insertion on a  $\log_{10}$  scale. The total, genome-wide library size is shown at left ( $n$ ). To better compare transpositions across libraries with different numbers of insertions, we also plotted the normalized local insertion rate as a density track.

### Sequencing and analysis: bulk DNA CC libraries

DNA calling card libraries were sequenced on the Illumina HiSeq 2500 platform. To increase the complexity of the library, PhiX was added at a final loading concentration of 50%. Reads were demultiplexed by the 3 base-pair barcode TAG followed by the end of the transposon terminal repeat, culminating with the *piggyBac* insertion site motif TTAA. Reads that had exact matches to these sequences were hard trimmed using cutadapt (Martin, 2011) with the following settings: `-g ^TAGTTTACGCAGACTATCTTTCTAGGGTTAA`—minimum-length 1—discard-untrimmed -e 0—no-indels. Reads passing this filter were then trimmed of vector sequence along read 2 using

cutadapt with the following settings: -g “^ATCACTTAAGCCGGTAC” –minimum-length 1 –discard-untrimmed -e 0 –no-indels. The remaining reads were aligned to the human genome (build hg38) with NovoAlign and the following settings: -n 40 -o SAM -o SoftClip. Aligned reads were validated by confirming that they mapped adjacent to the insertion site motif. Successful reads were then converted to calling card format (.ccf.; see [http://wiki.wubrowse.org/Calling\\_card](http://wiki.wubrowse.org/Calling_card)) using custom programs (available at [https://github.com/arnavm/calling\\_cards](https://github.com/arnavm/calling_cards)) and visualized on the WashU Epigenome Browser v46 (Zhou et al., 2011) (<http://epigenomegateway.wustl.edu/legacy/>).

### Sequencing and analysis: bulk RNA CC libraries

Multiple calling card libraries were pooled together for sequencing on the Illumina HiSeq 2500 platform with 50% phiX. Reads were demultiplexed by the N7 index sequences added during the final PCR. Read 1 began with the 3 base-pair barcode followed by the end of the transposon terminal repeat, culminating with the insertion site motif (TTAA in the case of *piggyBac*; TA in the case of *Sleeping Beauty*) before entering the genome. *piggyBac* reads were checked for exact matches to the barcode, transposon sequence, and insertion site at the beginning of reads before being hard trimmed using cutadapt with the following settings: -g “^NNNGCGTCAATTTTACGAGACTATCTTTCTAGGGTTAA” –minimum-length 1 –discard-untrimmed -e 0 –no-indels, where NNN is replaced with the primer barcode. *Sleeping Beauty* libraries were trimmed with the following settings: -g “^NNNTAAGTGTATGTAACTTCCGACTTCAACTGTA” –minimum-length 1 –discard-untrimmed -e 0 –no-indels. Reads passing this filter were then trimmed of any trailing Nextera adaptor sequence, again using cutadapt and the following settings: -a “CTGTCTCTTATACACATCTCCGAGCCCACGAGACTNNNNNNNNNTCTCGTATGCCGTCTTCTGCTTG” –minimum-length 1. The remaining reads were aligned to the human genome (build hg38) with NovoAlign and the following settings: -n 40 -o SAM -o SoftClip. Aligned reads were validated by confirming that they mapped adjacent to the insertion site motif. Successful reads were then converted to calling card format (.ccf.) and visualized on the WashU Epigenome Browser v46 (Zhou et al., 2011) (<http://epigenomegateway.wustl.edu/legacy/>).

### Sequencing and analysis: scRNA-seq libraries

scRNA-seq libraries were sequenced on either Illumina HiSeq 2500 or NovaSeq machines. Reads were analyzed using 10x Genomics’ Cell Ranger with the following settings: –expect-cells = 6000 –chemistry = SC3Pv2 –localcores = 16 –localmem = 30. The digital gene expression matrices from 10x were then further processed with scanpy (Wolf et al., 2018) for identification of highly variable genes, batch correction, dimensionality reduction, and Louvain clustering. Processed scRNA-seq datasets were stored as .loom files (<http://loompy.org>). We cross-referenced gene expression data with published datasets (Rosenberg et al., 2018; Rouillard et al., 2016; Saunders et al., 2018; Tasic et al., 2018; Zeisel et al., 2018) to assign cell types. The species mixing analysis was performed using Drop-seq\_tools (Macosko et al., 2015).

### Sequencing and analysis: scCC libraries

scCC libraries were sequenced on Illumina NextSeq 500 machines (v2 Reagent Cartridges) with 50% PhiX. We used the standard Illumina primers for read 1 and index 2 (BP10 and BP14, respectively), and custom primers for read 2 and index 1 (Table S4). Read 1 sequenced the cell barcode and unique molecular index of each self-reporting transcript. Read 2 began with GGTTAA (end of the *piggyBac* terminal repeat and insertion site motif) before continuing into the genome. Reads containing this exact hexamer were trimmed using cutadapt with the following settings: -g “^GGTTAA” –minimum-length 1 –discard-untrimmed -e 0 –no-indels. Reads passing this filter were then trimmed of any trailing P7 adaptor sequence, again using cutadapt and the following settings: -a “AGAGACTGGCAAGTACACGTCGCACTCACCATGANNNNNNNNATCTCGTATGCCGTCTTCTGCTTG” –minimum-length 1. Reads passing these filters were aligned using 10x Genomics’ cellranger with the following settings: –expect-cells = 6000 –nosecondary –chemistry = SC3Pv2 –localcores = 16 –localmem = 30. This workflow also managed barcode validation and collapsing of UMIs. Aligned reads were validated by verifying that they mapped adjacent to TTAA tetramers. Reads were then converted to calling card format (.ccf.). Finally, to minimize the presence of intermolecular artifacts, we required that each insertion must have been tagged by at least two different UMIs. We used the set of validated cell barcodes from each scRNA-seq library to demultiplex library-specific barcoded insertions from the scCC data. This approach requires no shared cell barcodes between individual scCC (and scRNA-seq) libraries. As a result, we excluded insertions from non-unique cell barcodes, which represented a very small number of total cells lost (< 1% per multiplexed library). More details on these steps are also provided in the associated protocols. For the species mixing experiment, cells were classified as either human or mouse if at least 80% of self-reporting transcripts in that cell mapped to the human or mouse genome, respectively, and as a multiplet. The estimated multiplet rate was calculated by doubling the observed percentage of human-mouse multiplet, to account for human-human and mouse-mouse doublets.

### Peak calling on calling card data

We called peaks in calling card data using Bayesian blocks (Scargle et al., 2013), a noise-tolerant algorithm for segmenting discrete, one-dimensional data, using the astropy implementation (Robitaille et al., 2013; The Astropy Collaboration et al., 2018). Bayesian blocks segments the genome into non-overlapping blocks where the density of calling card insertions is uniform. By comparing the segmentation against a background model, we were able to use Poisson statistics to assess whether a given block shows statistically significant enrichment for insertions. Let  $B = \{b_1, b_2, \dots, b_n\}$  represent the set of blocks found by performing Bayesian block

segmentation on all insertions from a TF-directed experiment (e.g., SP1-PBase). For each block  $b_i$ , let  $x_i$  be the number of insertions in that block in the TF-directed experiment. Similarly, let  $y_i$  be the number of insertions in that block in the undirected experiment (e.g., PBase) normalized to the total number of insertions found in the TF-directed experiment. Then, for each block we calculated the Poisson p value of observing at least  $x_i$  insertions assuming a Poisson distribution with expectation  $y_i$ :  $P(k \geq x_i | \lambda = y_i)$ . We accepted all blocks that were significant beyond a particular p value threshold.

For the analysis of TF-directed insertions, either in bulk or in single cells, we added a pseudocount of 1 to  $y_i$ , the number of insertions in block  $b_i$  in the undirected experiment. We selected all blocks whose p values were significant at a Benjamini-Hochberg false discovery rate of 5% (Benjamini and Hochberg, 1995). We polished peak calls by merging statistically significant blocks that were within 250 bases of each other and by aligning block edges to coincide with TTAAs.

To identify BRD4 binding sites from undirected *piggyBac* insertions, we segmented those insertions using Bayesian blocks. For each block  $b_i$ , we let  $x_i$  denote the number of undirected insertions in that block. We also calculated  $x_i'$ , the expected number of insertions in block  $b_i$  assuming *piggyBac* insertions were distributed uniformly across the genome. We did this by dividing the total number of TTAAs in the genome by the total number of undirected insertions, then multiplying this value by the number of TTAAs in block  $b_i$ . Then, for each block we calculated the Poisson p value  $P(k \geq x_i | \lambda = x_i')$ . We accepted all blocks that were significant beyond a particular p value threshold. Finally, we merged statistically-significant blocks that were within 12,500 bases of each other (Pott and Lieb, 2015; Whyte et al., 2013).

For the bulk PBase and HyPBase analysis, we used p value cutoffs of  $10^{-30}$  and  $10^{-62}$ , respectively. (We chose these stringent thresholds to better resolve super-enhancers, which is our primary focus here.) For both *in vitro* and *in vivo* single cell HyPBase analyses, we used a p value cutoff of  $10^{-9}$ . To identify the differentially-bound loci between  $CD24^{\text{high}}/CD24^{\text{low}}$  K562 cells, as well as between upper and lower cortical layer neurons (i.e., *Pou3f2/Brn-2*, *Bcl11b/Ctip2*, and *Foxp2*), we used the same framework as described above for TF-directed analysis but did reciprocal enrichment analyses, where one dataset was used as the “experiment” track and the other as the “control” track, and vice-versa. This results in two one-sided hypothesis tests. When analyzing differential binding between upper and lower cortical layer neurons, we used a p value cutoff of  $10^{-9}$ . For the  $CD24^{\text{high}}/CD24^{\text{low}}$  K562 analysis, we restricted our hypothesis testing to BRD4-bound peaks found in the cell line mixing experiment that had at least 20 insertions between both groups. For each peak, we normalized the number of insertions from each population by a library-specific scaling factor and calculated the fold change in binding as  $\log_2$  (Normalized  $CD24^{\text{high}}$  insertions/Normalized  $CD24^{\text{low}}$  insertions). We then took the smaller of the two p values and adjusted for multiple hypotheses at a Benjamini-Hochberg false discovery rate of 10%. This was plotted against the fold-change values to generate the volcano plot (colored circles indicate significant peaks after FDR correction). Data points were annotated when peaks overlapped or were near a single gene.

Density tracks were generated by taking the Bayesian blocks segmentation of each calling card dataset and, for each block, calculating the normalized number of insertions and dividing by the length of the block in kilobases (insertions per kilobase per million mapped insertions, or IPKM). This was plotted as a bedgraph file with smoothing applied in the WashU Epigenome Browser (25 pixel windows).

Custom code to facilitate these analyses is available online ([https://github.com/arnavm/calling\\_cards](https://github.com/arnavm/calling_cards)). Detailed instructions on how to analyze calling card data are provided in the linked protocols.

### TF binding analysis

We compared our TF-directed calling card peaks to publicly available ChIP-seq datasets. See below for more details on aligning and analyzing ChIP-seq data. We collated a list of unique transcription start sites (TSSs) by taking the 5'-most coordinates of RefSeq Curated genes in the hg38 build (UCSC Genome Browser). A list of CpG islands in HCT-116 and K562 cells and their methylation statuses were derived from previously-published Methyl-seq data (Brunner et al., 2009). We used the liftOver tool (Hinrichs et al., 2006) to convert coordinates from hg18 to hg38. We tested for enrichment in SP1-directed insertions at TSSs, CpG islands, and unmethylated CpG islands with the G test of independence. We used the same test when testing enrichment of BAP1-directed insertions at TSSs. For motif discovery, we restricted our analysis to peaks less than 5,000 bp in length. We then used MEME-ChIP (Machanic and Bailey, 2011) with a dinucleotide shuffled control and the following settings: -dna -nmeme 600 -seed 0 -ccut 250 -meme-mod zoops -meme-minw 4 -meme-nmotifs 10. Motifs were aligned on the web version of Tomtom (Gupta et al., 2007) querying the “Vertebrates (In vivo and in silico)” database. We cross-referenced BAP1 scCC binding sites with publicly available BAP1 shRNA data (Yen et al., 2018), focusing on genes that showed a significant change in gene expression (adjusted p value < 0.05).

### BRD4 sensitivity, specificity, and precision

We used a published BRD4 ChIP-seq dataset (McClelland et al., 2016) to identify BRD4-bound super-enhancers in HCT-116 cells, following previously-described methods (Lovén et al., 2013; Whyte et al., 2013). We first called peaks using MACS 1.4.1 (Zhang et al., 2008) at  $p < 10^{-9}$  (using the parameters -p 1e-9-keep-dup = “auto” -f BAM -g hs -w -S-space = 50), then fed this into ROSE. We discarded artifactual loci less than 2,000 bp in size, yielding a final list of 162 super-enhancers. To evaluate sensitivity, we used BEDtools (Quinlan and Hall, 2010) to ask what fraction of *piggyBac* peaks, at various p value thresholds, overlapped the set of BRD4-bound super-enhancers. To measure specificity, we created a list of regions predicted to be insignificantly enriched ( $p > 0.1$ ) for BRD4 ChIP-seq signal. We then sampled bases from this region such that the distribution of peak sizes was identical to that of the 162 super-enhancers. We sampled to 642x coverage, sufficient to cover each base with one peak, on average. We then asked

what fraction of our *piggyBac* peaks overlapped these negative peaks and subtracted that value from 1 to obtain specificity. Finally, we calculated precision, or positive predictive value, by dividing the total number of detected super-enhancer peaks by the sum of the super-enhancer peaks and the false positive peaks.

### Downsampling and replication analysis

When performing downsampling analyses on calling card insertions, we randomly sampled insertions without replacement and in proportion to the number of reads supporting each insertion. Peaks were called on the downsampled insertions at a range of p value cutoffs. Linear interpolation was performed using NumPy (Oliphant, 2015) and visualized using matplotlib (Hunter, 2007). Replication was assessed by splitting calling card insertions into two, approximately equal, files based on their barcode sequences. Each new file was treated as a single biological experiment. For each peak called from the joint set of all insertions, we plotted the number of normalized insertions (IPM) in one replicate on the x axis and the other replicate on y axis.

### Analysis of external datasets

For ChIP-seq, ATAC-seq, and DNase-seq data, we aligned raw reads using Novoalign with the following settings for single-end datasets: -o SAM -o SoftClip; while paired-end datasets were mapped with the additional flag -i PE 200-500. To calculate and visualize the fold enrichment in ChIP-seq signal at calling card peaks, we used deeptools (Ramírez et al., 2016). We tested for significant mean enrichment in BRD4 ChIP-seq signal at *piggyBac* peaks over randomly shuffled control peaks with the Kolmogorov-Smirnov test. Chromatin state analysis was performed using ChromHMM as previously described (Ernst et al., 2011). For each chromatin state, we plotted the mean and standard deviation of the rate of normalized insertions (IPKM). We called peaks on SP1 ChIP-seq, DNase- and ATAC-seq data using MACS 2 with the following settings: -q 0.05-keep-dup = "auto." For the analysis of "super-enhancers" from ATAC-seq data, we used control data derived from ATAC-seq on deproteinized human genomic DNA (Martins et al., 2018) and followed the same steps for calling super-enhancers from BRD4 ChIP-seq data (above). If necessary, files were converted to hg38 using liftOver (Hinrichs et al., 2006).

### Cell state analyses of K562: scRNA-seq and scCC

Cell state analysis was performed on batch-corrected K562 scRNA-seq data derived from the HyPBase cell line mixing experiment. Principal components analysis (PCA) of single cell gene expression (Figure S5A) revealed *CD24* as one of the top genes in PC1, while PC2 was enriched in hemoglobin genes, particularly the fetal-specific markers *HBE1* and *HBZ*. Furthermore, the expression of top PC1 and PC2 genes appear to be anticorrelated: cells that strongly expressed *CD24* are not likely to express *HBZ*, and vice-versa (Figure S5B), suggesting mutually exclusive states. We then scored cells based on the expression of *VIM*, *TMSB4X*, *HBG1*, and *HBG2*, revealing a gradient of cell states along a stem-like-to-differentiated axis (Figure 5A). We then modeled the distribution of this state score as a 3 component Gaussian mixture model, drawing cutoffs where adjacent Gaussian distributions intersected (Figure S5C). These cutoffs were then used to label cells as either stem-like (*CD24*<sup>high</sup>), differentiated (*CD24*<sup>low</sup>), or intermediate (Figure S5D). The expression levels of *CD24* and *HBZ*, which were not used to score cells, showed high specificity for the stem-like and differentiated clusters (Figure S5E). Differentially bound peaks were called as described above.

### Analysis of K562 experiments

We analyzed the JQ1 time course experiment using a two-way ANOVA with treatment and day as the independent variables and the percentage of *CD24*<sup>low</sup> cells as the dependent variable. For the analysis of annexin V levels in either JQ1- or DMSO-treated *CD24*<sup>high</sup> and *CD24*<sup>low</sup> cells, we used a three-way ANOVA with treatment, cell state, and day as independent variables. The imatinib experiments following either JQ1 or *BRD4* CRISPRi pretreatment were analyzed using a two-way ANOVA with pretreatment (JQ1/DMSO or NT/*BRD4* gRNA) and treatment as the independent variables. Multiple hypothesis correction was performed using Tukey's honestly significant difference. For the cell cycle inhibitor experiment, data were analyzed using a one-way ANOVA with Dunnett's post hoc test using either DMSO or EtOH (for RO-3306) as controls.

### In vivo scCC analysis and validation

Single cell RNA-seq and single cell calling card libraries were prepared, sequenced, and analyzed as described above. Cell types were assigned based on the expression of key marker genes and cross-referenced with recent cortical scRNA-seq datasets (Rosenberg et al., 2018; Saunders et al., 2018; Tasic et al., 2018; Zeisel et al., 2018). BRD4-bound peak calls were validated by comparing to a previously published cortical H3K27ac ChIP-seq dataset (Stroud et al., 2017). Read alignment and statistical analysis were performed as described above.

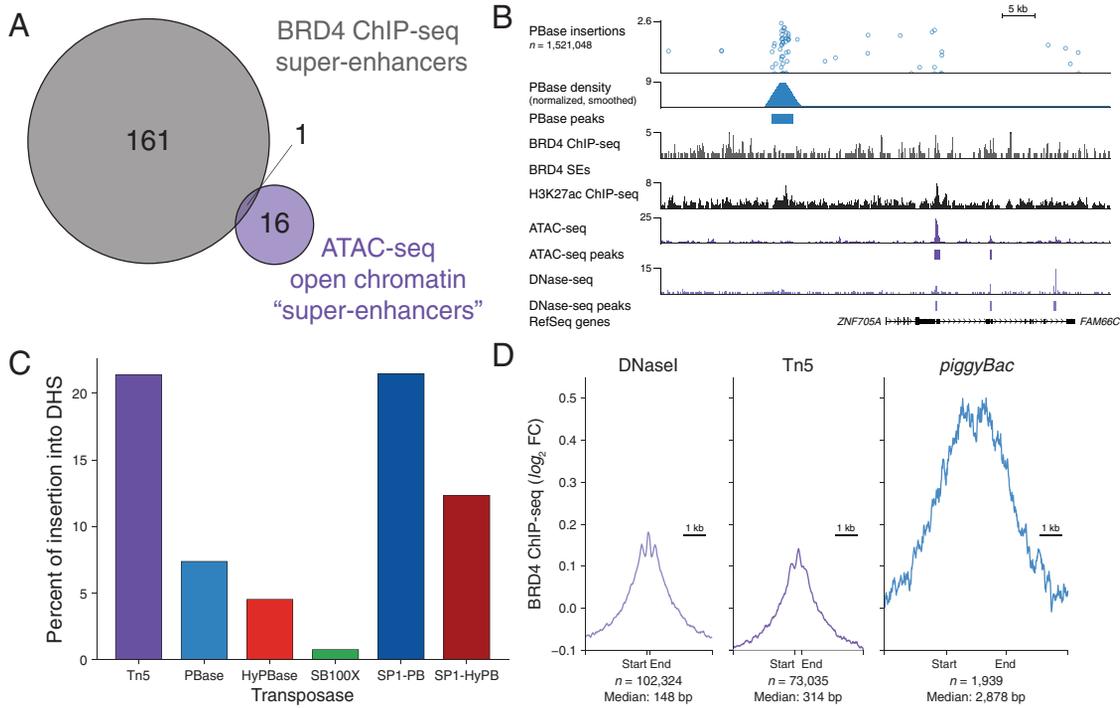
The specificity of BRD4-bound gene expression in astrocytes and neurons was analyzed by first identifying all genes within 10,000 bases of astrocyte and neuronal BRD4 peaks. Although assigning an enhancer to its target gene is a difficult problem, using the nearest gene is common practice (Gasparini et al., 2019). To control for sensitivity of gene detection, we downsampled the neuron insertions to the same number of astrocyte insertions, then called peaks and identified nearby genes in this subset. We used gene expression data from a bulk RNA-seq dataset (Zhang et al., 2014) to compute the specificity of gene expression between astrocytes and neurons. We first discarded genes whose expression was not measured, and then set the value for genes with 0.1 FPKM to zero (to better distinguish non-expressed genes from lowly-expressed genes). Finally, for each gene  $g_i$ , we calculated the specificity as

$\text{Astrocyte}_{\text{FPKM}}(g_i) / [\text{Astrocyte}_{\text{FPKM}}(g_i) + \text{Neuron}_{\text{FPKM}}(g_i)]$ . Thus, a value of 0 denotes a gene purely expressed in neurons, a value of 0.5 for a gene equally expressed in both cell types, and a value of 1 for a gene purely expressed in astrocytes. After accounting for differences in library size, we identified 383 genes near astrocyte peaks and 184 genes near neuron peaks, with 46 genes found in both datasets. We plotted the distributions of gene expression specificity for these gene sets. (Figure 4E). Gene Ontology analysis was performed on the same sets of genes using PANTHER (Mi et al., 2017) on the “GO biological process complete” database. Fisher’s exact test was used to compute p values, which were then subject to Bonferroni correction.

### ADDITIONAL RESOURCES

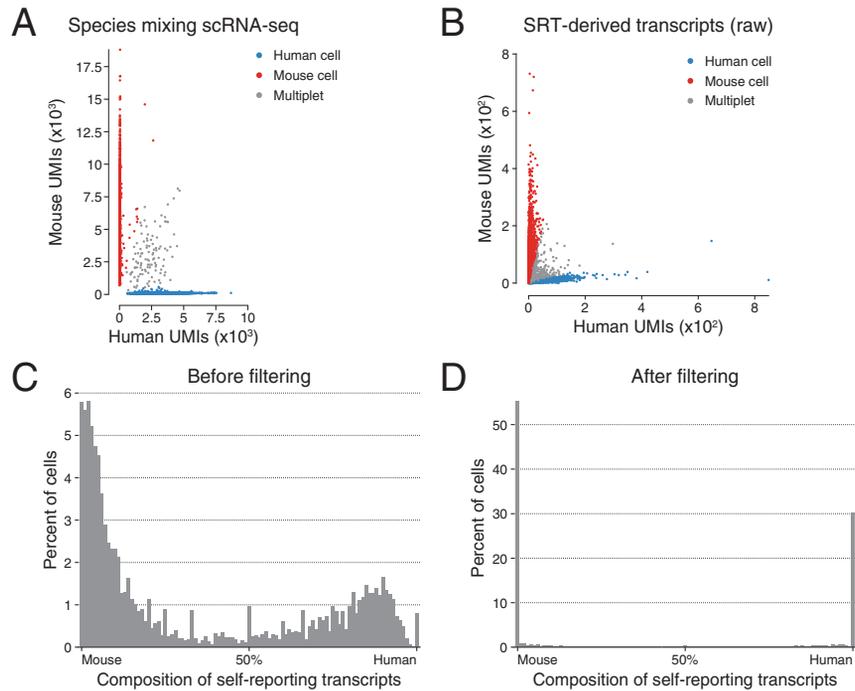
We have created a number of protocols describing how to perform all aspects of bulk and single cell calling cards, from molecular biology and sequencing through data analysis and visualization. While these are listed in the [Key Resources Table](#), we have also created a publicly accessible portal for easy access to all our workflows: <https://www.protocols.io/groups/calling-cards/>. Moving forward, this resource should contain the most up-to-date information.

# Supplemental Figures



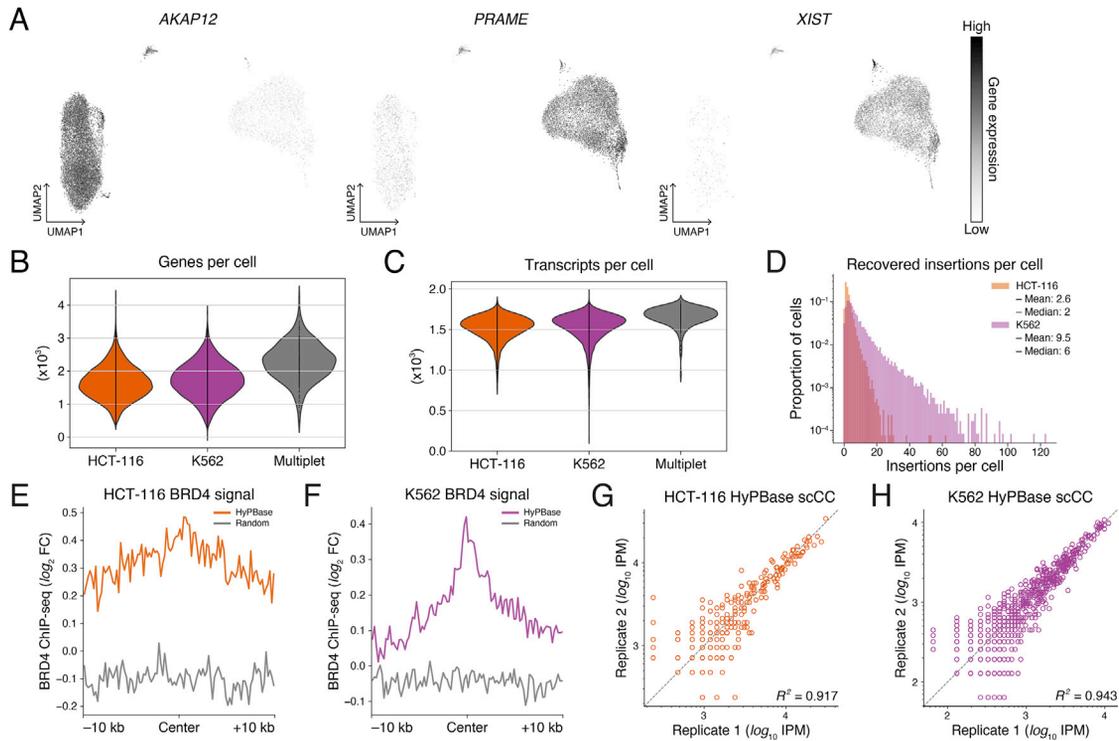
**Figure S1. BRD4 Calling Cards with Undirected *piggyBac* Are Not Equivalent to ATAC-Seq, Related to Figure 2**

(A) Overlap of BRD4 super-enhancers, as inferred from BRD4 ChIP-seq, and "super-enhancers" inferred from open chromatin ATAC-seq peaks in HCT-116 cells. (B) Browser view of a BRD4 calling card peak that is not detected by ATAC-seq nor DNase-seq. (C) Comparison of transposase predilections for accessible chromatin. (D) Comparison of peak sizes and BRD4 ChIP-seq enrichment as called by DNase-seq, ATAC-seq, and undirected *piggyBac* calling cards, respectively. Peaks are scaled to the median peak width (denoted by the start and end ticks) and are flanked by 3 kb in either direction. SE: super-enhancer; DHS: DNaseI hypersensitivity site; FC: fold change; kb: kilobase.



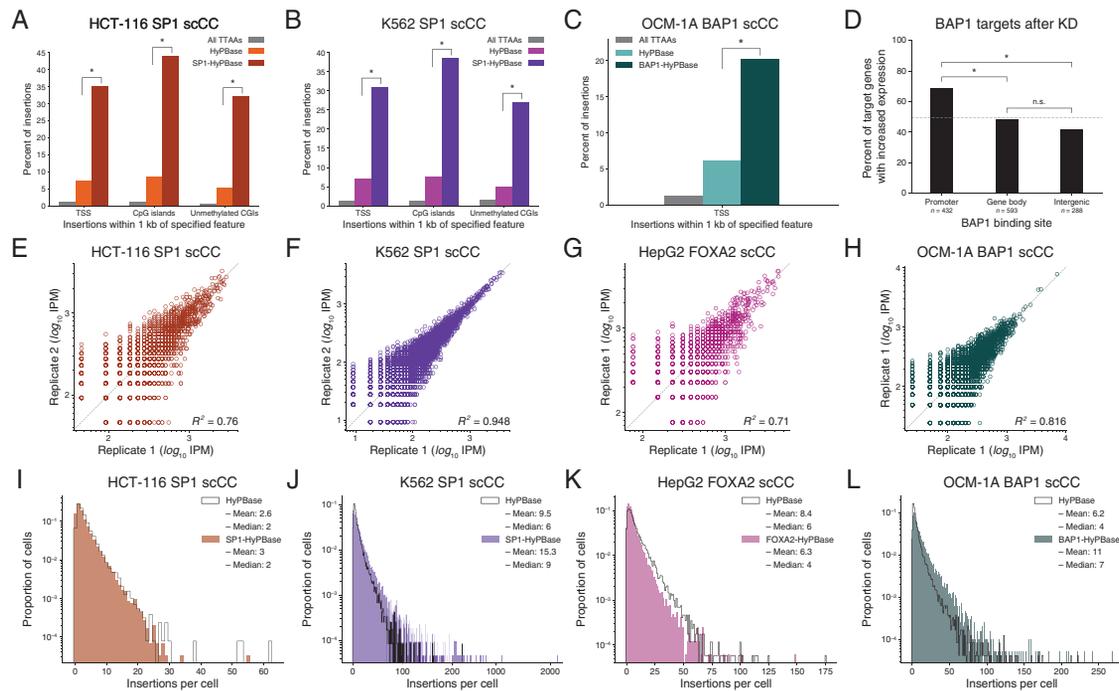
**Figure S2. Filtering Single-Cell SRTs Reduce Intermolecular Artifacts, Related to Figure 3**

(A) Barnyard plot from scRNA-seq of human HCT-116 and mouse N2a cells. (B) Barnyard plot from scCC of HCT-116 and N2a cells without filtering (estimated multiplet rate of 25.1%). (C) Distribution of cell barcode purity from unfiltered scCC data. The x axis is the proportion of transcripts mapping to the human or mouse genomes. (D) Distribution of species purity after filtering scCC data. UMI: unique molecular indexes.



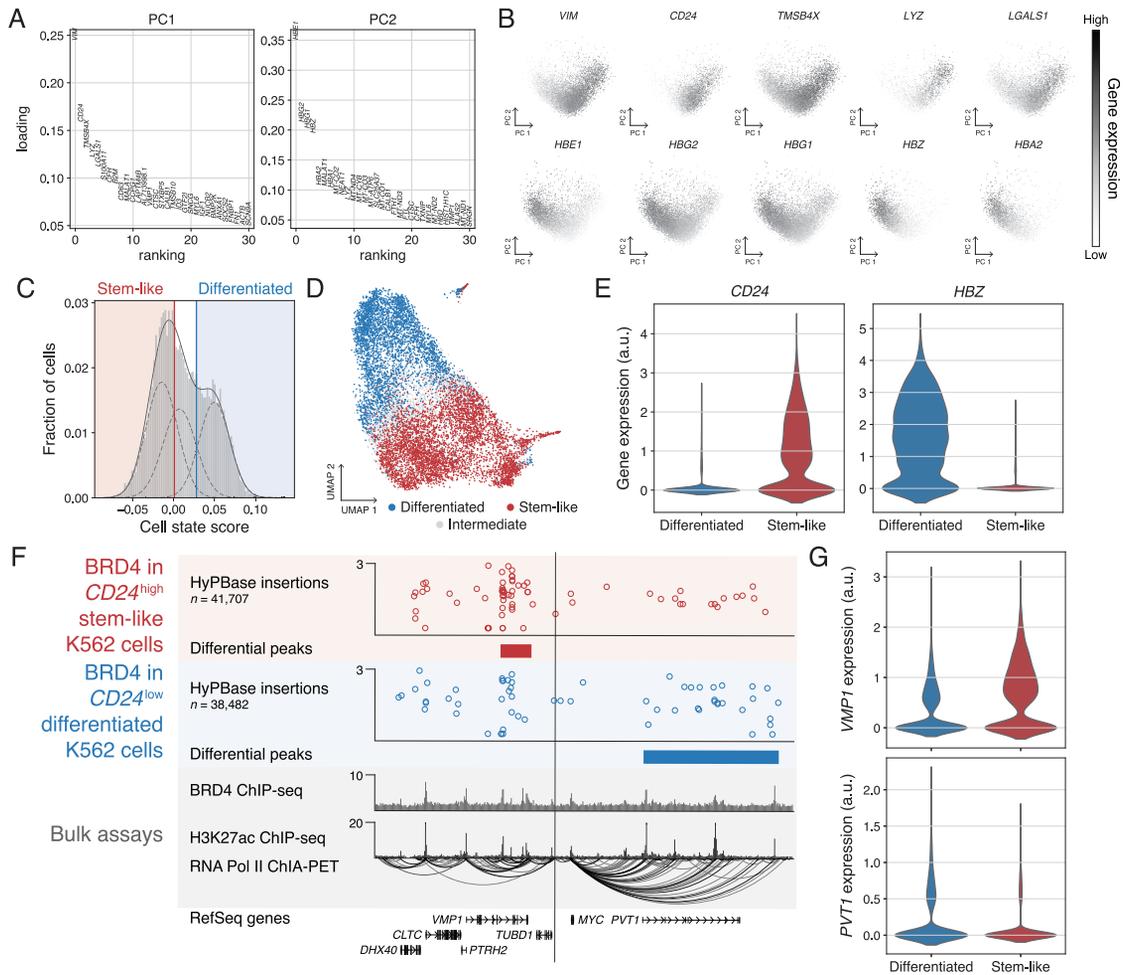
**Figure S3. Validation and Performance of Undirected *In Vitro* scCC, Related to Figure 3**

(A) Single cell expression levels of three marker genes in a mixed scRNA-seq library of human HCT-116 and K562 cells. (B) Distributions of genes per cell by cell type. (C) Distributions of transcripts per cell by cell type. (D) Distributions of HyPBase insertions recovered per cell in HCT-116 and K562 cells. (E-F) Mean BRD4 ChIP-seq signal at HyPBase peaks in HCT-116 and K562 cells, respectively, compared to randomly permuted peaks (KS test  $p < 10^{-9}$  in each case). (G-H) Reproducibility of normalized insertions deposited by HyPBase and recovered by scCC at BRD4 binding sites in HCT-116 and K562 cells, respectively. KS: Kolmogorov-Smirnov.



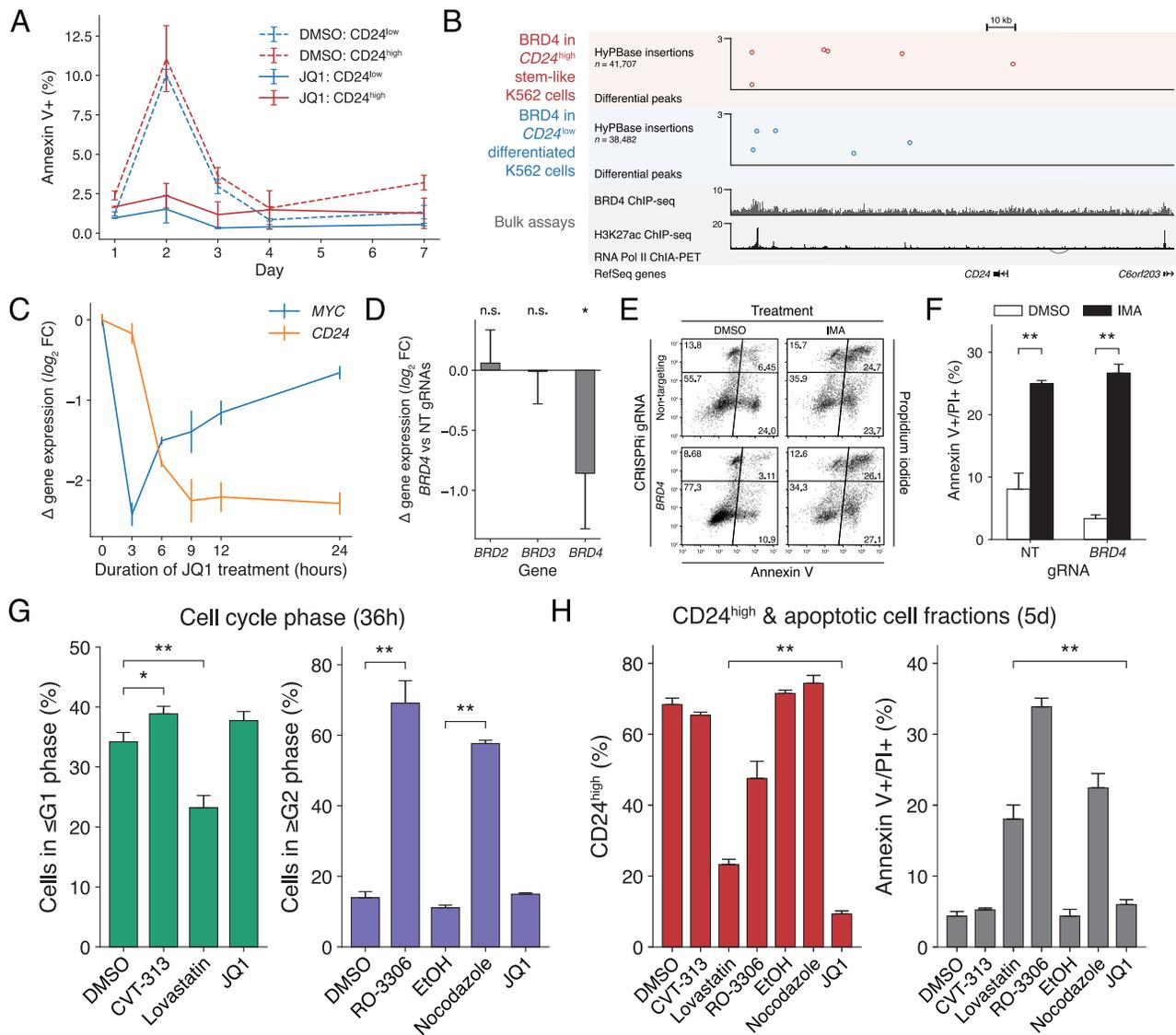
**Figure S4. Validation and Performance of TF-Directed *In Vitro* scCC, Related to Figure 4**

(A-B) Enrichment of SP1-HyPBase-directed insertions to TSSs, CGIs, and unmethylated CGIs in single HCT-116 and K562 cells, respectively (G test of independence  $p < 10^{-9}$ ). (C) Enrichment of BAP1-HyPBase-directed insertions TSSs in single OCM-1A cells (G test of independence  $p < 10^{-9}$ ). (D) Percent of BAP1 targets that increase expression upon BAP1 KD stratified by binding site (Fisher's exact test  $p < 10^{-9}$ ). The dashed gray line represents the overall fraction of genes that increased expression upon KD. (E-H) Reproducibility of normalized insertions deposited by either HyPBase or TF-HyPBase fusions and recovered by scCC at TF binding sites, for the respective TF-cell line pair. (I-L) The distribution of recovered insertions per cell by construct (HyPBase versus TF-HyPBase) and cell type. TF: transcription factor; TSS: transcription start site; CGI: CpG island; KD: knockdown; IPM: insertions per million mapped insertions; n.s.: not significant.



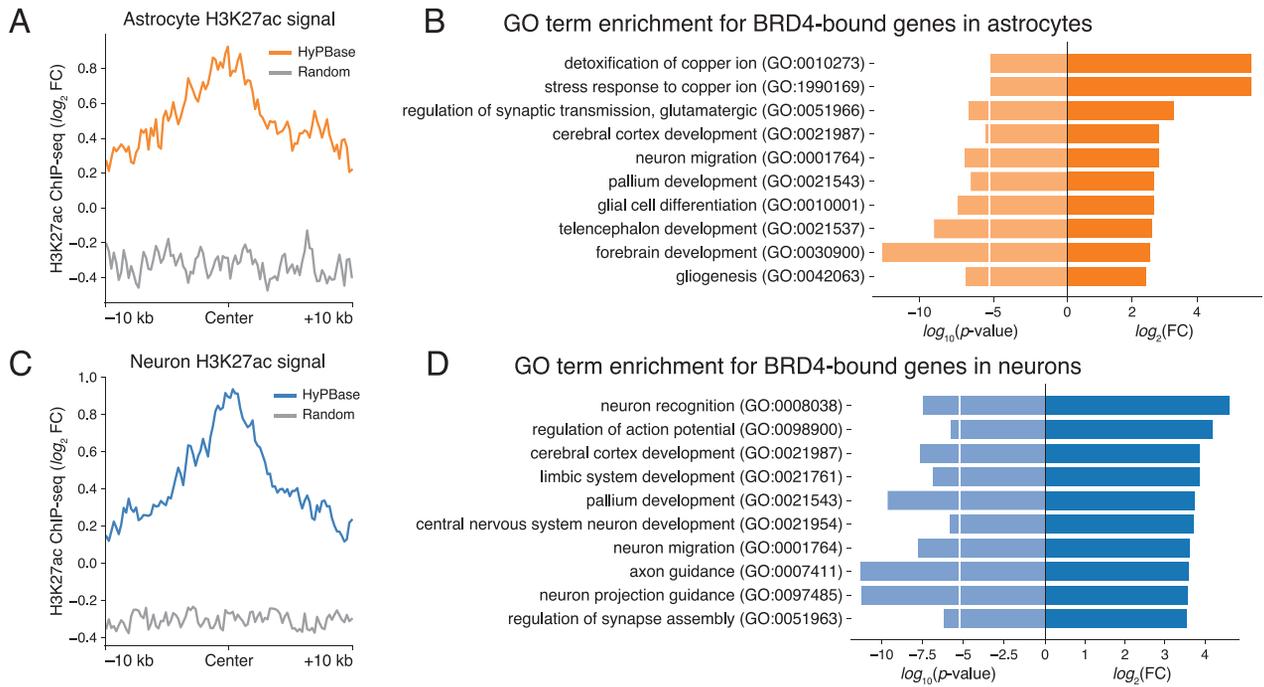
**Figure S5. Clustering of K562 Cells into Stem-like and Differentiated States, Related to Figure 5**

(A) Principal component analysis of K562 scRNA-seq data. (B) Relative expression levels of highest-ranking genes in PC1 (top) and PC2 (bottom). (C) Gaussian mixture modeling of a cell-state score to define stem-like and differentiated K562 clusters. (D) Visualization of assigned cell clusters in the UMAP projection. (E) Specific expression of *CD24* and *HBZ* in the stem-like and differentiated clusters, respectively. (F) Genome browser view of scCC in the stem-like and differentiated clusters alongside bulk BRD4 and H3K27ac ChIP-seq as well as RNA Pol II ChIA-PET. (G) Expression of *VMP1* and *PVT1* in the stem-like and differentiated clusters. PC: principal component.



**Figure S6. Validation of Bromodomain-Dependent K562 Cell States, Related to Figure 5**

(A) Annexin V staining in CD24<sup>high</sup> (red) or CD24<sup>low</sup> (blue) K562 cells treated with DMSO (dashed line) or JQ1 (solid line) over a seven-day time course. (B) Genome browser view of the *CD24* locus. (C) qRT-PCR for *MYC* and *CD24* expression levels in bulk K562 cells treated with JQ1 relative to DMSO-treated controls. (D) Expression changes in *BRD2*, *BRD3*, and *BRD4* in K562 cells transduced with dCas9-KRAB and *BRD4* CRISPRi gRNA (Welch's t test  $p < 0.05$ ). (E) Annexin V and PI co-staining in cells subjected to either non-targeting (top) or *BRD4* (bottom) CRISPRi followed by either DMSO (left) or imatinib (right) treatment. (F) Average percent of annexin V/PI double positive cells in either the non-targeted or *BRD4* CRISPRi replicates, stratified by either DMSO or imatinib exposure (two-way ANOVA  $p < 0.01$ ). (G) Percent of K562 cells in either G1 (left) or G2 (right) phase after 36 hours of drug treatment (one-way ANOVA with Dunnett's test  $*p < 0.05$ ,  $**p < 0.01$ ). (H) Percent of K562 cells in the CD24<sup>high</sup> state (left) after 5 days, and the percent of annexin V/PI double positive cells (right) at the same time point ( $p < 0.01$  in each instance, one-way ANOVA with Dunnett's test). Bars/points represent means; error bars denote standard deviations. Experiments were performed in triplicate. DMSO: dimethyl sulfoxide; n.s.: not significant; FC: fold change; SSC: side scatter; CRISPRi: CRISPR interference; NT: non-targeting; gRNA: guide RNA; IMA: imatinib; PI: propidium iodide.



**Figure S7. Validation of *In Vivo* BRD4 Binding in Astrocytes and Neurons, Related to Figure 6**

(A) Mean H3K27ac ChIP-seq signal at HyPBase peaks in astrocytes compared to randomly permuted peaks (KS test  $p < 10^{-9}$ ) (B) GO term enrichment analysis of genes near astrocytic BRD4 binding sites. (C) Mean H3K27ac ChIP-seq signal at HyPBase peaks in neurons compared to randomly permuted peaks (KS test  $p < 10^{-9}$ ). (D) GO term enrichment analysis of genes near neuronal BRD4 binding sites. (B and D) The white line indicates the Bonferroni-adjusted p value threshold at  $\alpha = 0.05$ . GO: Gene Ontology; KS: Kolmogorov-Smirnov; FC: fold change.