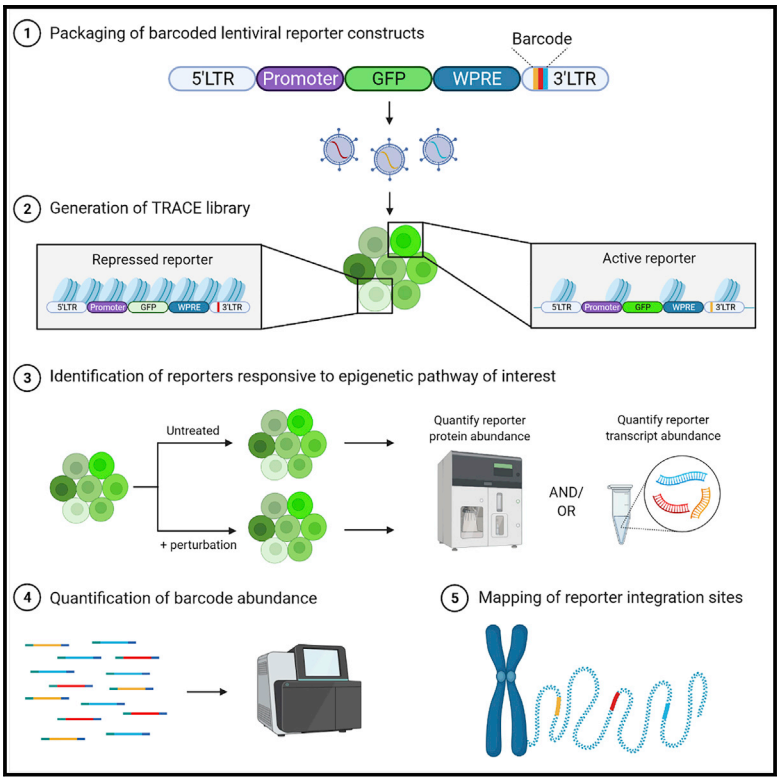
Molecular Cell

TRACE generates fluorescent human reporter cell lines to characterize epigenetic pathways

Graphical abstract



Highlights

- TRACE can identify fluorescent reporters for theoretically any epigenetic pathway
- Barcoded lentiviral GFP expression constructs report on their chromatin environment
- LEDGF chimeras redirect lentiviral integrants to desired chromatin states
- TRACE identifies reporters responsive to inhibition of LSD1 and loss of SUZ12

Authors

Iva A. Tchasovnikarova, Sharon K. Marr, Manashree Damle, Robert E. Kingston

Correspondence

it257@cam.ac.uk (I.A.T.), kingston@molbio.mgh.harvard.edu (R.E.K.)

In brief

Tchasovnikarova et al. present TRACE, a genome-wide technique that exploits barcoded lentiviral constructs randomly integrated across the genome to report on their chromatin environment. By profiling GFP expression from thousands of integrants in the presence and absence of a genetic or pharmacological perturbation, TRACE identifies phenotypic reporters of epigenetic pathways.

Tchasovnikarova et al., 2022, Molecular Cell *82*, 479–491
January 20, 2022
https://doi.org/10.1016/j.molcel.2021.11.035

CellPress

Molecular Cell

CellPress
OPEN ACCESS

Technology

TRACE generates fluorescent human reporter cell lines to characterize epigenetic pathways

Iva A. Tchasovnikarova,[1,2,3,4,*] Sharon K. Marr,[1,2] Manashree Damle,[1,2] and Robert E. Kingston[1,2,*]
[1]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA
[2]Department of Genetics, Harvard Medical School, Boston, MA 02114, USA
[3]The Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK
[4]Lead contact
*Correspondence: it257@cam.ac.uk (I.A.T.), kingston@molbio.mgh.harvard.edu (R.E.K.)
https://doi.org/10.1016/j.molcel.2021.11.035

SUMMARY

Genetically encoded biosensors are powerful tools to monitor cellular behavior, but the difficulty in generating appropriate reporters for chromatin factors hampers our ability to dissect epigenetic pathways. Here, we present TRACE (transgene reporters across chromatin environments), a high-throughput, genome-wide technique to generate fluorescent human reporter cell lines responsive to manipulation of epigenetic factors. By profiling GFP expression from a large pool of individually barcoded lentiviral integrants in the presence and absence of a perturbation, we identify reporters responsive to pharmacological inhibition of the histone lysine demethylase LSD1 and genetic ablation of the PRC2 subunit SUZ12. Furthermore, by manipulating the HIV-1 host factor LEDGF through targeted deletion or fusion to chromatin reader domains, we alter lentiviral integration site preferences, thus broadening the types of chromatin examined by TRACE. The phenotypic reporters generated through TRACE will allow the genetic interrogation of a broad range of epigenetic pathways, furthering our mechanistic understanding of chromatin biology.

INTRODUCTION

Genetically encoded biosensors are powerful tools to dissect the molecular mechanisms underlying complex cellular pathways. By producing a measurable luminescent or fluorescent signal upon detection of a particular biological stimulus, biosensors not only "report" on a biochemical process of interest (Mehta and Zhang, 2011) but also allow the mechanistic dissection of a cellular pathway through genetic screens. Indeed, the basis for our understanding of heterochromatin formation is largely derived from a series of classic forward genetic screens in *Drosophila melanogaster* designed to identify modifiers of the position-effect variegation (PEV) phenotype (Eissenberg et al., 1992; Rea et al., 2000; Reuter and Wolff, 1981; Tschiersch et al., 1994). With the advent of CRISPR/Cas9 technology and next-generation sequencing, analogous screens can now be performed in human cells. The power of this approach has already been demonstrated through the discovery of a number of vertebrate-specific epigenetic regulators, highlighting the benefits of performing such screens in mammalian systems (Fukuda et al., 2018; Liu et al., 2018; Robbez-Masson et al., 2018; Tchasovnikarova et al., 2015, 2017; Timms et al., 2016).

The broader application of these genetic methods, however, is stymied by the paucity of suitable reporters for epigenetic complexes in human cells. Existing methods to generate chromatin reporter lines involve either knockin of a fluorescent protein into an endogenous gene whose transcription is affected by a perturbation of interest or targeted knockin of an artificial expression cassette into a genomic locus thought to be regulated by the epigenetic pathway under investigation. However, both approaches have significant drawbacks. For the vast majority of endogenous genes, the magnitude of change in gene expression in response to a perturbation is likely to be relatively modest, and while a 2-fold change in transcription might be highly significant in a typical RNA sequencing (RNA-seq) experiment, it would not provide sufficient signal for a genetic screen. Conversely, expression constructs driven by exogenous promoters can exhibit far larger changes in response to a perturbation; however, there is no guarantee that upon targeting to a genomic location of interest (based on chromatin immunoprecipitation [ChIP]-seq profiles, for example), the reporter will be responsive to the desired perturbation. Thus, a broadly applicable method capable of generating chromatin reporter cell lines would unleash the power of mammalian genetics to study epigenetic processes in human cells.

Here, we present TRACE (Transgene Reporters Across Chromatin Environments), an unbiased, genome-wide method that allows for the identification of chromatin reporters for epigenetic pathways of interest. We showcase the utility of TRACE by generating fluorescent reporter lines that respond to pharmacological inhibition of the histone lysine demethylase LSD1. TRACE relies on the use of lentiviral vectors, which preferentially

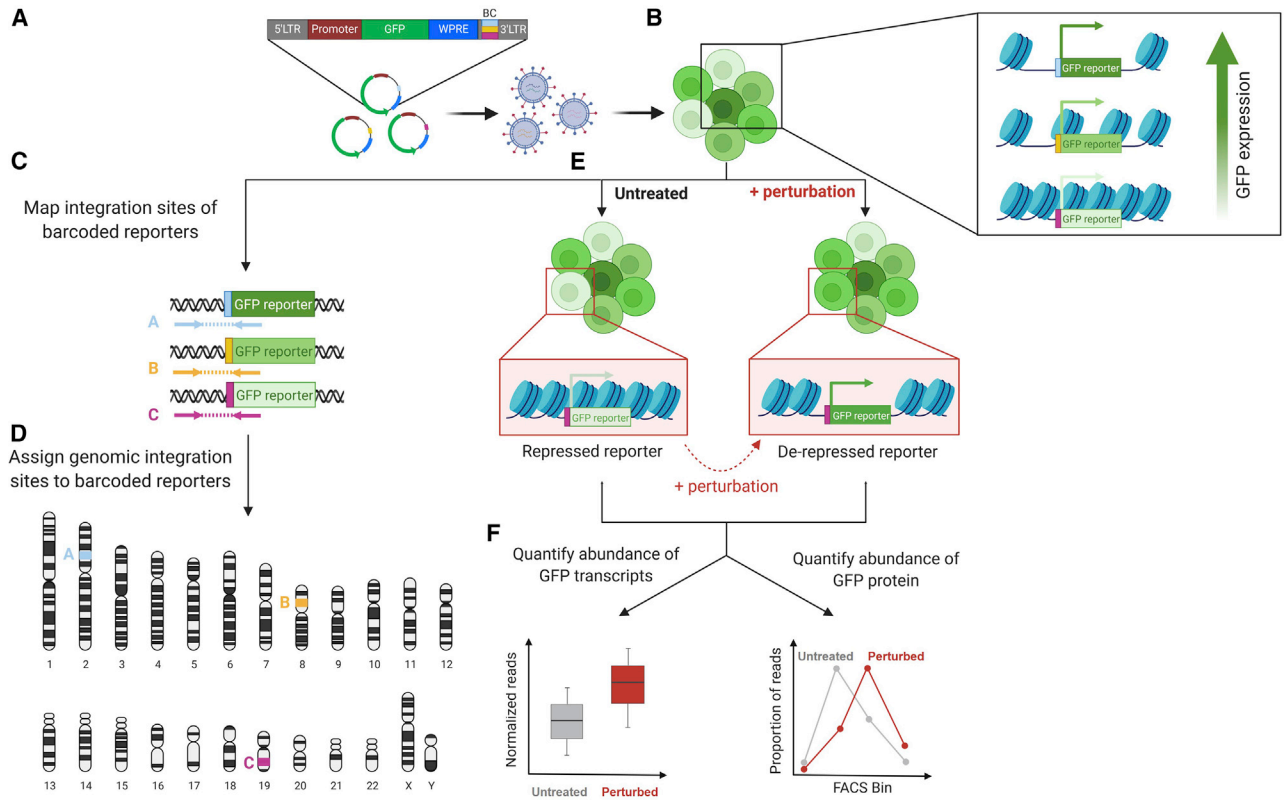


Figure 1. Overview of TRACE (Transgene Reporters Across Chromatin Environments)

(A) A barcoded GFP transgene library is introduced into cells through lentiviral expression.

(B) Upon integration into the host genome, the lentiviral transgenes become subject to chromatin position effects: permissive chromatin environments allow for strong GFP expression, while reporters in repressive chromatin environments are subject to silencing.

(C and D) To pair reporter barcodes to their respective genomic loci, lentiviral integration sites are mapped.

(E and F) Finally, by profiling transgene expression—at the transcript or protein level—in the presence and absence of a genetic or pharmacological perturbation, individual integrants located at specific sites in the genome that respond to the stimulus can be identified.

integrate into the bodies of actively transcribed genes; however, we demonstrate that ablation of the HIV-1 integrase-binding cellular factor LEDGF redirects lentiviral integrations away from gene bodies, thereby allowing TRACE to sample a wider range of chromatin environments. Furthermore, we create fusion proteins comprising the HIV-1 integrase-binding region of LEDGF and a range of chromatin reader domains to target lentiviral integrations to particular flavors of chromatin. We use one of these LEDGF chimeras, harboring a CBX7 chromodomain ($CBX7_{CD}$), to identify reporters responsive to loss of the PRC2 subunit suppressor of zeste 12 (SUZ12) and find that genomic loci harboring SUZ12-represed reporters are enriched not only in trimethylated lysine 27 of histone H3 (H3K27me3) but also in trimethylated lysine 9 of histone H3 (H3K9me3).

DESIGN

Inspired by the TRIP (Thousands of Reporters Integrated in Parallel) methodology (Akhtar et al., 2013), which demonstrated that genetically barcoded reporters can be used to measure the effect of different types of chromatin on gene expression in a highly multiplexed fashion, we designed a method capable of identifying transgene reporters that, by virtue of the nature of the chromatin environment in which they are integrated, exhibit altered expression in response to genetic or pharmacological perturbation of epigenetic pathways (Figure 1).

In a TRACE experiment, cultured cells are transduced at a low multiplicity of infection (MOI) with a lentiviral library of green fluorescent protein (GFP) expression vectors, such that each cell contains a single viral integration (Figure 1A). As the constructs are identical apart from a unique 24-nucleotide "barcode" cassette, any cell-to-cell variation in GFP expression levels reflects the influence of different chromatin environments at the site of integration (Figure 1B). The genomic integration sites of all lentiviral reporters are determined (Figure 1C), thus allowing barcodes to be assigned to their corresponding chromosomal location (Figure 1D). To identify responsive reporters, the library is partitioned in two: one-half undergoes a genetic or pharmacological perturbation to inhibit the pathway of interest, while the other serves as a control (Figure 1E). GFP expression—which can be measured at either the transcript level or the protein level (see below)—is then quantified through Illumina sequencing,

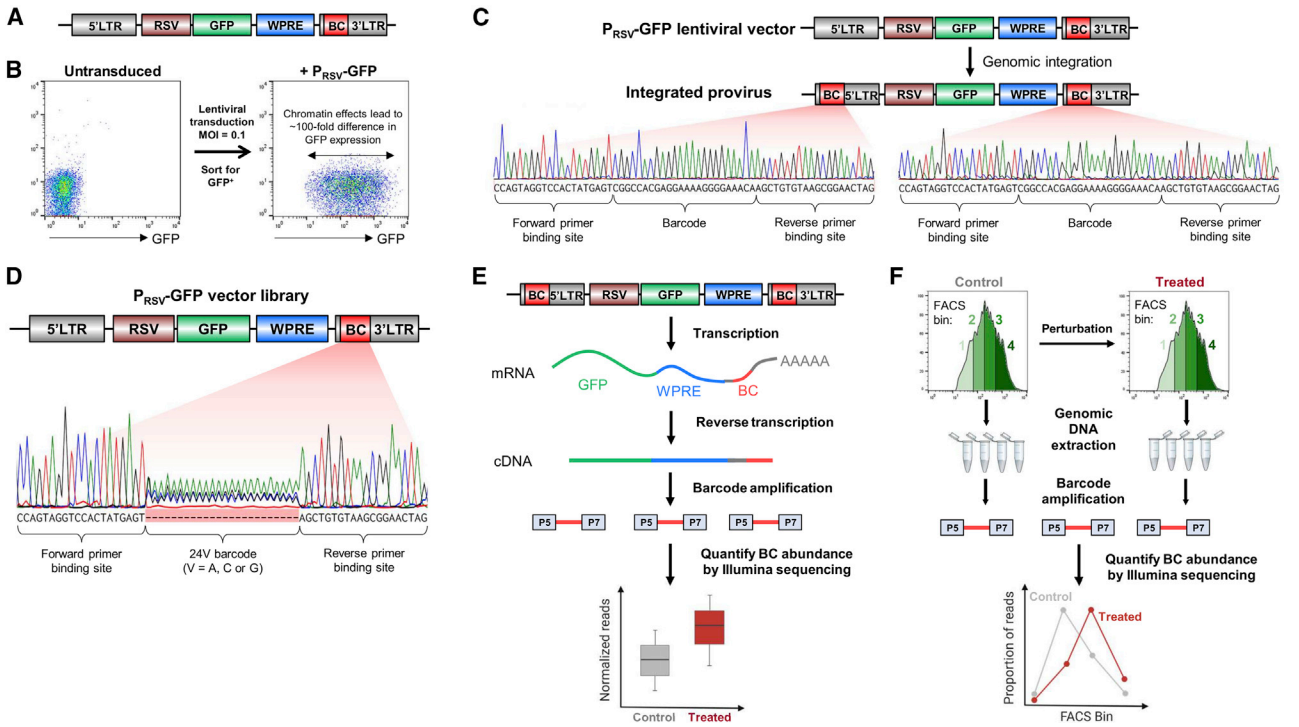


Figure 2. Implementation of TRACE

(A and B) Generation of GFP reporter library. (A) Schematic representation of the GFP reporter lentiviral vector. LTR, long terminal repeat; RSV, Rous sarcoma virus promoter; GFP, green fluorescent protein; WPRE, Woodchuck hepatitis virus post-transcriptional regulatory element; BC, barcode. (B) Transduction of KBM-7 cells with this vector results in a broad range of GFP expression.

(C and D) Barcoding of the TRACE vector. (C) Positioning the barcode cassette in the U3 region of the 3′ LTR results in duplication of the barcode upon integration into the genome. (D) Removal of thymine from the 24-nucleotide barcode prevents NlaIII cleavage within the barcode cassette during mapping of the lentiviral integration sites.

(E and F) Quantifying TRACE reporter expression by Illumina sequencing. (E) Amplification of the barcode cassette from cDNA allows the quantification of reporter transcript abundance, while (F) barcode amplification from genomic DNA following FACS partitioning allows for quantification of reporter expression at the protein level. In each case, reporter expression is assessed in both control and treated populations.

with the sequence of the barcode serving as a unique identifier and the abundance of the barcode sequence within the pool serving as a relative measure of expression (Figure 1F). By profiling reporter expression in the presence and absence of a genetic or pharmacological perturbation, individual integrants located at specific sites in the genome whose expression is altered by the stimulus can be identified. A successful TRACE experiment therefore defines a set of responsive genomic loci: CRISPR/Cas9-mediated knockin targeted to these sites subsequently generates individual fluorescent reporter clones, which can then then be subjected to a battery of genetic and biochemical assays to dissect the underlying biology.

The resolution of TRACE is dependent on the range of GFP expression across the population. The optimal reporter would drive very strong GFP expression in permissive chromatin environments, but would be efficiently suppressed when placed in repressive chromatin environments. Our previous work examining a panel of promoters driving GFP in a range of cell lines showed that transduction of a human chronic myeloid leukemia (CML) cell line, KBM-7, with a standard HIV-1-based lentiviral expression vector encoding GFP driven by the Rous sarcoma virus (RSV) long terminal repeat (LTR) promoter (P_{RSV}-GFP) (Figures 2A and 2B) resulted in the broadest range of GFP expression among the cell line/promoter combinations tested (Tchasovnikarova et al., 2015). Thus, all subsequent work developing TRACE was carried out using this combination.

To allow the unique identification of individual integrants, we inserted a barcode cassette into the U3 region of the 3′ LTR in the P_{RSV}-GFP vector (Figure 2C), upstream of the lentiviral polyadenylation signal; this ensures that the barcode is transcribed as part of the 3′ UTR of the reporter transcript, which is required if the barcode is to be used to readout reporter transcript abundance. Furthermore, the U3 region of the 3′ LTR is duplicated into the 5′ LTR upon integration, thus placing the barcode just 36 bp away from the adjacent genomic DNA. Minimizing the amount of invariant viral DNA between the barcode cassette and neighboring genomic DNA decreases the propensity for recombination during PCR amplification, which could otherwise uncouple barcodes from their corresponding genomic integration site. Finally, as our lentiviral mapping procedure relies on the digestion of genomic DNA with the restriction enzyme NlaIII (Timms et al., 2019), we removed one nucleotide (thymidine) from the barcode

(Figure 2D) to prevent occurrences of the NlaIII recognition site (CATG); this is important as cleavage at these sites would lead to the generation of aberrant truncated PCR products, which could be preferentially amplified and hence dominate the sequence reads resulting from the mapping procedure.

TRACE allows the relative expression of individual barcoded integrants to be measured at both the transcript level and the protein level. To quantify reporter transcripts, total RNA is extracted from the population of barcoded cells and reporter transcripts are converted to cDNA using a barcode cassette-specific primer for reverse transcription (Figure 2E). Barcodes can then be PCR amplified and transcript abundance quantified by Illumina sequencing (Figure 2E). To ensure that any potential differences in growth rates between barcoded clones cannot confound the expression measurement, barcodes are also amplified in parallel from genomic DNA and used to normalize transcript counts.

Quantification of GFP protein is achieved using fluorescence-activated cell sorting (FACS) followed by Illumina sequencing. The reporter population is partitioned into four bins based on GFP expression: reporters integrated into permissive chromatin environments are well expressed and are therefore likely to be found in the top bins, while reporters integrated into repressive chromatin environments are expressed at lower levels and are therefore likely to be found in the lower bins. Genomic DNA is extracted from the sorted cells collected in each bin, and barcode abundance across the four bins is quantified through PCR amplification and Illumina sequencing (Figure 2F). The distribution of sequencing reads across the four bins serves as a relative measure of GFP expression. For each individual barcoded reporter, we derive a gene expression metric (GEM) ranging between 1 (completely repressed; reads seen solely in bin 1) and 4 (fully expressed; reads seen solely in bin 4), as calculated by the equation $\mathrm{GEM} = \sum_{i=1}^{4} Ri \times i$

where i is the FACS bin and Ri is the fraction of the sequencing reads present for that reporter in the given bin i.

The output of a successful TRACE experiment is 2-fold. First, the genomic sites harboring responsive integrants can be correlated with various genomic and chromatin features to define those regions of the genome where chromatin regulators of interest exert their function; this is an advantage over biochemical methods that identify the genomic loci at which epigenetic factors reside, as binding does not necessarily equate to function. However, the real power of the TRACE approach lies in the subsequent utility of the reporter lines. CRISPR-mediated knockin at the genomic sites identified by TRACE generates individual fluorescent reporter lines that can be interrogated genetically to reveal the underlying biology. Reconstitution of the reporter phenotype using mutant versions of the epigenetic factor of interest can provide a rapid, functional assay to (1) delineate the protein domains required for function and (2) assess the functionality of disease-associated mutations. Moreover, the fluorescent phenotypic readout allows for genome-wide CRISPR screens, which have the power to identify collaborative factors required for protein recruitment to the reporter locus and downstream factors that translate this recruitment into a transcriptional response. The power of these assays has been demonstrated through our previous work on the HUSH (Human Silencing Hub) complex (Tchasovnikarova et al., 2015, 2017).

RESULTS

TRACE identifies reporters responsive to LSD1 inhibition

We performed a proof-of-principle TRACE experiment to find reporters responsive to pharmacological perturbation of LSD1-mediated histone lysine demethylation. LSD1 catalyzes the removal of mono- and di-methyl groups on lysine 4 of histone H3 (H3K4me1 and H3K4me2). Thus, treatment of cells with the LSD1 inhibitor GSK2879552 (hereafter LSD1i) results in an accumulation of H3K4me1 and H3K4me2 at LSD1 target sites and transcriptional derepression (Mohammad et al., 2015).

We transduced one million KBM-7 cells with the barcoded reporter library at an MOI of 0.1 (~10% GFP^+), such that the overwhelming majority (>99%) of transduced cells contained only a single integrant. The GFP^+ population was isolated by FACS to generate a TRACE library, which we estimated to contain ~100,000 unique integrations based on the starting proportion of GFP^+ cells. After expansion, the library was divided into two and treated with either DMSO or LSD1i, ensuring that 100-fold representation of each barcoded reporter was maintained in the two conditions for the duration of the treatment (Figures 3A and S1A). Seven days later, we assayed GFP expression at the protein level using FACS followed by Illumina sequencing (Figure 3A). To assess the reproducibility of the method, we performed the TRACE experiment in duplicate.

We sought to identify integrants that exhibited concordant derepression upon LSD1i treatment across duplicate experiments. While the vast majority of transgenes exhibited little change in expression upon treatment with LSD1i (Figures 3B and 3C), we found 981 integrants that were derepressed across the two replicate experiments (Figure 3D). There was no obvious pattern to the location of these integrants, with the hits distributed at approximately the expected frequency across genomic regions marked by different histone modifications (Figure S1B).

Next, we leveraged the screen results to generate LSD1i-responsive reporter clones. We used CRISPR/Cas9 technology to target a GFP expression construct to two genomic sites (Figures 3E and S1C). Owing to the low efficiency of homologous recombination in KBM-7 cells, we used a non-homologous end joining (NHEJ) knockin approach (Lackner et al., 2015). KBM-7 cells were nucleofected with a donor plasmid expressing GFP and two single guide RNA (sgRNA)/Cas9 ribonucleoprotein (RNP) complexes: one of the complexes contained an sgRNA targeting the intended genomic site, while the other was loaded with an sgRNA targeting a pair of sites flanking the donor cassette, thus liberating the reporter construct from the plasmid to facilitate NHEJ-mediated integration into the genome (Figure S1D). Two weeks after nucleofection, the resulting population of cells was treated with LSD1i and single-cell clones were isolated from the GFP^+ population. Although the knockin efficiency was low (~0.1%), we were readily able to isolate single-cell clones from the GFP^+ population that contained the P_{RSV}-GFP-pA insert at the intended target sites (Figure S1E). Critically,

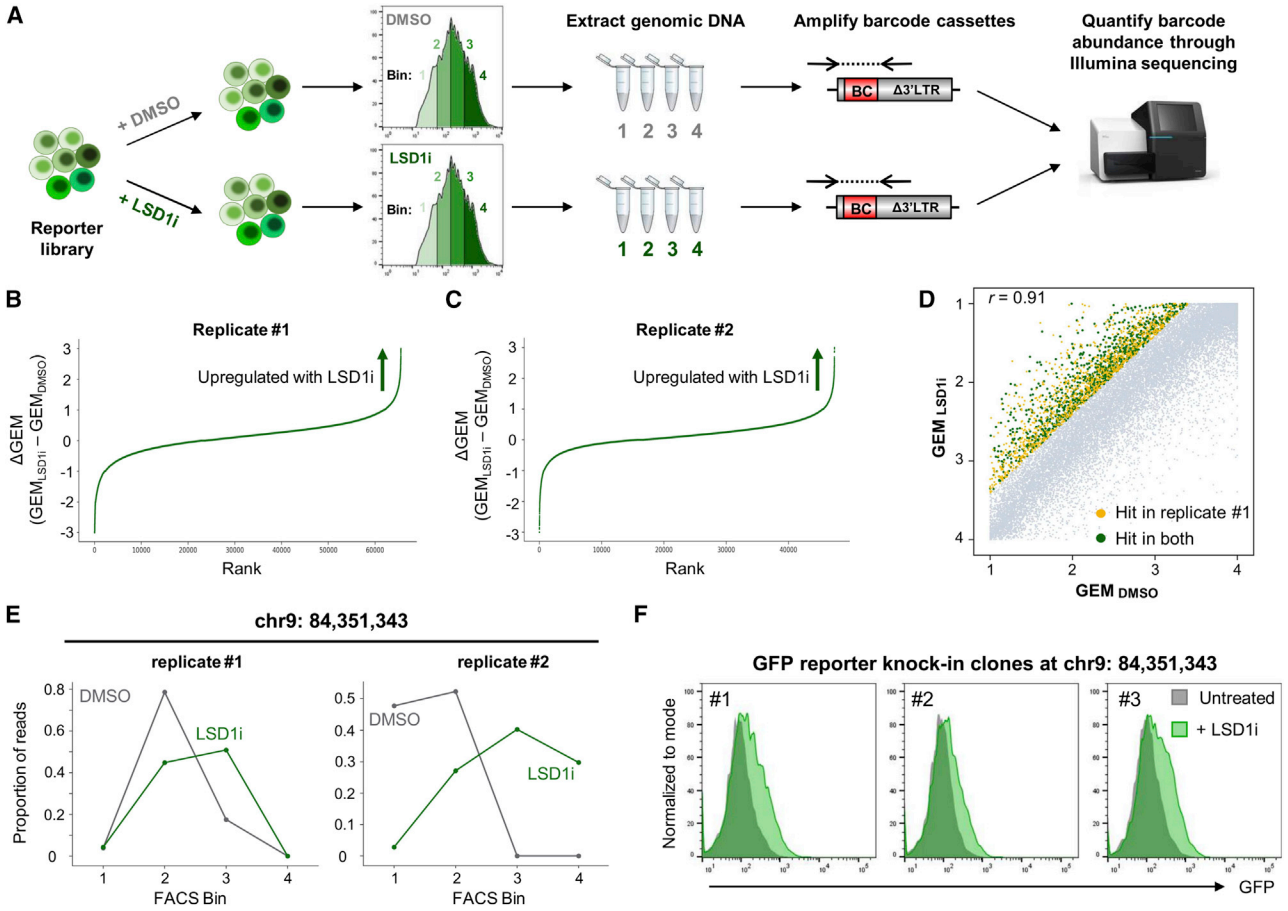


Figure 3. TRACE identifies reporters responsive to LSD1 inhibition

(A) Schematic representation of the TRACE experiment designed to identify reporters exhibiting increased expression upon treatment with the LSD1 inhibitor GSK2879552 (LSD1i).

(B–D) Identification of LSD1i-reponsive reporters. While the expression of the vast majority of integrants was unaffected by treatment with LSD1i (B and C), 2,066 integrants were upregulated in the presence of LSD1i in replicate #1, of which 981 were also derepressed in replicate #2 (D). We defined shared hits as the integrants ranked in the top 10% in replicate #1, which were also upregulated by a GEM of at least 0.3 in replicate #2.

(E and F) Generation of LSD1i-reponsive reporter clones. (E) Expression profile of a reporter (chromosome 9 [chr9]: 84,351,343) significantly derepressed upon treatment with LSD1i, as measured by FACS followed by Illumina sequencing. (F) CRISPR/Cas9-mediated knockin to chr9: 84,351,370 (27 bp away from the genomic site identified by TRACE) (see Figure S1) resulted in the generation of single-cell reporter clones responsive to LSD1i treatment.

treatment with LSD1i increased GFP expression in each clone (Figures 3F and S1F). Thus, a TRACE screen combined with a CRISPR/Cas9-mediated knockin approach can be used to generate reporters responsive to pharmacological perturbation of an epigenetic pathway.

Genomic distribution of TRACE reporters

To gain deeper insight into the genomic locations examined by TRACE, we took advantage of the fact that the transcript-based readout is not limited by the requirement for one reporter per cell and generated a large reporter library by transducing two million KBM-7 cells at an MOI of ~1.2 (68% GFP$^+$). Following isolation of the GFP$^+$ population, reporter expression was measured by PCR amplification of the barcode cassette from cDNA followed by Illumina sequencing (Figure 4A). The experiment was performed in triplicate (Figure S2) and identified a total of 1,256,599 integrants of which 612,246 could be assigned to genomic integration sites. We examined the distribution of these integrants across the genome by correlating their position with 8 chromatin states (Figures 4F–4H) and 10 histone modifications (Figures 4I–4K) (see STAR Methods).

TRACE reporters were heavily biased toward active chromatin, with 90.1% of integrations located in expressed genes (Figures 4B–4E). Integrants were strongly enriched across chromatin states and histone modifications associated with active transcription and depleted from repressive chromatin environments (Figures 4F–4K). Chromatin position effects clearly influenced reporter expression: transgene expression was maximal in proximity to transcriptional start sites (TSSs), but decreased along gene bodies, while among the repressive chromatin environments, the heterochromatic state (“ZNF genes/Repeats/Heterochromatin”) marked by H3K9me3 exerted a stronger

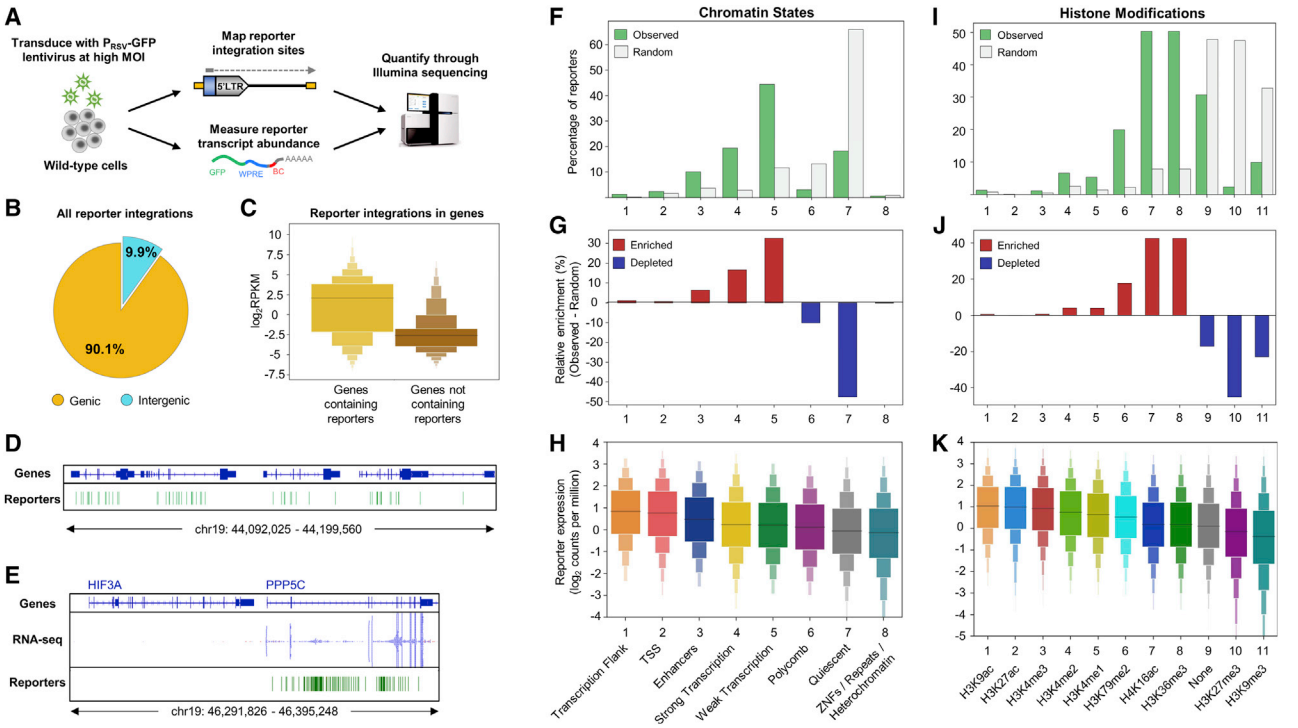


Figure 4. TRACE reporters are biased toward integration into expressed genes

(A) Generation of a high complexity reporter library.

(B–E) Preferential integration of lentiviral reporters into expressed genes. The vast majority of integrants map to genes (B and D) that are well expressed (C and E).

(F and G) Reporter integrants are biased toward euchromatic chromatin states. The barplots depict the distribution of reporter integration sites across eight chromatin states. "Observed" refers to the distribution of mapped integration sites, while "random" refers to the distribution of integration sites that would be expected if lentiviral integration was entirely random.

(H) Integrant expression is influenced by the overlapping chromatin environment. Reporter expression is highest in the vicinity of transcriptional start sites (TSSs) and lowest in repressive chromatin states.

(I and J) Reporter integrants are biased toward genomic sites marked by active histone modifications. The barplots depict the distribution of reporter integration sites across ten histone modifications. "Observed" refers to the distribution of mapped integration sites, while "random" refers to the distribution of integration sites that would be expected if lentiviral integration was entirely random.

(K) Integrant expression is influenced by the overlapping chromatin environment. Reporter expression is highest in proximity to TSSs (marked by H3K9ac and H3K4me1/2/3) and enhancers (marked by H3K27ac) and lowest in heterochromatic regions (marked by H3K27me3 and H3K9me3).

silencing effect than the "Polycomb" state marked by H3K27me3 (Figures 4H and 4K).

Expanding the scope of TRACE through depletion of LEDGF

While our initial TRACE screen was successful in identifying responsive reporters, the uneven distribution of lentiviral integration sites across the genome would likely limit the utility of the approach to a subset of chromatin pathways. Thus, we sought to alter the integration preferences of lentiviral transgenes to provide a more even coverage across the genome, thereby permitting TRACE to sample a wider variety of different chromatin environments. The transcriptional coactivator lens epithelium-derived growth factor (LEDGF) is a host factor known to modulate HIV-1 integration site selection (Cherepanov et al., 2003). LEDGF localizes to its genomic target sites through an N-terminal PWWP domain, which binds trimethylated lysine 36 of histone H3 (H3K36me3) found across the bodies of actively transcribed genes (Turlure et al., 2006). Through the binding of its integrase protein to the C terminus of LEDGF, HIV-1 hijacks LEDGF to facilitate its integration into actively transcribed genes; indeed, knockout of LEDGF has been shown to increase the proportion of HIV-1 integrations occurring outside of expressed genes (Shun et al., 2007). Therefore, we reasoned that performing TRACE in cells lacking LEDGF may allow us to sample a wider array of chromatin states, thereby broadening the scope and utility of the method.

We used CRISPR/Cas9-mediated gene disruption to generate a clonal LEDGF knockout (KO) KBM-7 cell line (Figure S3A). Wild-type and LEDGF KO cells were transduced in parallel at a high MOI, and the integration sites were mapped (Figure 5A). Overall, lentiviral integrations in LEDGF KO cells were more uniformly distributed across the genome: in contrast to wild-type cells where reporters were heavily biased toward expressed genes, LEDGF KO cells also contained reporters in inactive genes and intergenic regions (Figures 5B–5D). Indeed, we observed a greater proportion of "dim" reporters in LEDGF KO cells (30.7% in bin 1) than in wild-type cells (24% in bin

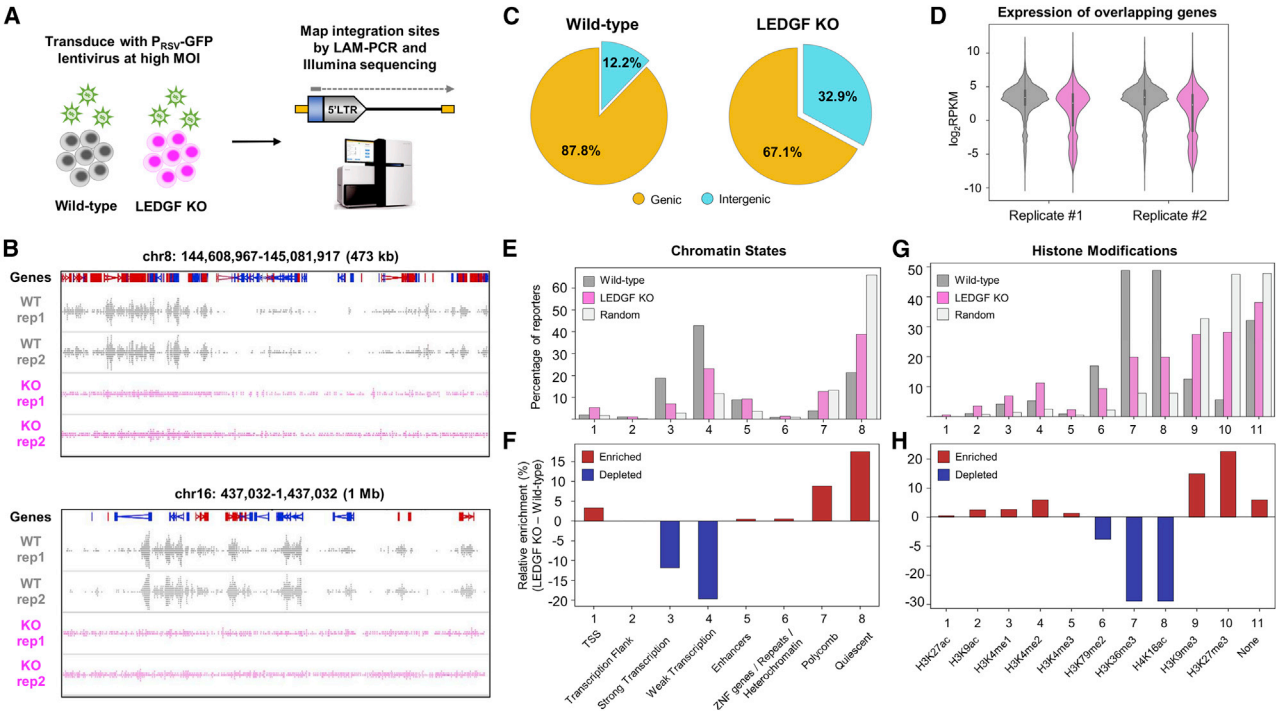


Figure 5. Loss of LEDGF redirects lentiviral integrants away from expressed genes

(A) Schematic representation of the experiment designed to compare the pattern of lentiviral integration sites in wild-type versus LEDGF KO cells.

(B and C) Lentiviral integrants in LEDGF KO cells are more evenly distributed between genes and intergenic regions compared with wild-type cells.

(D) Correlation between integrant transgene expression and that of their overlapping genes. Loss of LEDGF increased the proportion of lentiviral integrants found in poorly expressed genes.

(E–H) LEDGF KO redirects lentiviral integrants away from gene bodies, as assessed through analysis of overlapping chromatin states (E and F) and histone modifications (G and H).

1) (Figure S3B), suggesting that a higher number of integrants were located in repressed areas of the genome. This increase could not be explained by a general loss of transcription due to lack of functional LEDGF, as we did not observe any global trend toward lower GFP expression among the TRACE transgene population in LEDGF KO cells (Figure S3B, top). Loss of LEDGF redirected reporter integrations away from chromatin states and histone modifications associated with actively transcribed genes and instead resulted in an enrichment of integrations across all other types of chromatin (Figures 5E–5H, S3C, and S3D). Therefore, LEDGF ablation expands the scope of TRACE by altering the integration site preferences of lentiviral vectors.

Redirection of TRACE reporters through LEDGF chimeras

The ability to preferentially target reporters to a specific flavor of chromatin would enhance TRACE, as increasing the proportion of transgene integrants in particular types of chromatin would increase the likelihood of identifying relevant reporters. Thus, given the striking dependence of lentiviral integration site selection on LEDGF, we wondered whether this pathway could be manipulated to promote lentiviral integration into particular types of chromatin.

We postulated that replacing the H3K36me3-binding PWWP domain with alternative chromatin reader modules would bias TRACE transgene libraries toward desired types of chromatin. Indeed, there is preliminary evidence that replacing the N-terminal half of LEDGF, containing the DNA-binding AT hook and PWWP domain, with the PHD finger of inhibitor of growth protein 2 (ING2) is sufficient to bias lentiviral integrations toward CpG islands and TSSs, while the addition of the chromodomain of heterochromatin protein 1 homolog alpha (HP1α) or beta (HP1β) is sufficient to redirect integrations away from genes (Ferris et al., 2010; Gijsbers et al., 2010). We first reconstituted LEDGF KO cells with chimeric LEDGF constructs expressing the C-terminal integrase-binding domain (IBD) of LEDGF (residues 325–530) fused downstream of (1) the H3K27me3-binding chromodomain of CBX7$_{CD}$, (2) the H3K9me3-binding chromodomain of HP1α (HP1α_{CD}), or (3) the H3K4me3-binding PHD finger of ING2 (ING2$_{PHD}$) (Figures 6A and S4A). The chimera-expressing cells were then transduced at a high MOI, and lentiviral integration sites were mapped through linear amplification-mediated PCR (LAM-PCR) and Illumina sequencing (Figure 6B).

All three LEDGF chimeras successfully redirected lentiviral integrations to the expected genomic loci, as assessed by analysis of overlapping histone modifications and chromatin states (Figures 6C and 6D and S4C–S4E). Lentiviral integrations in

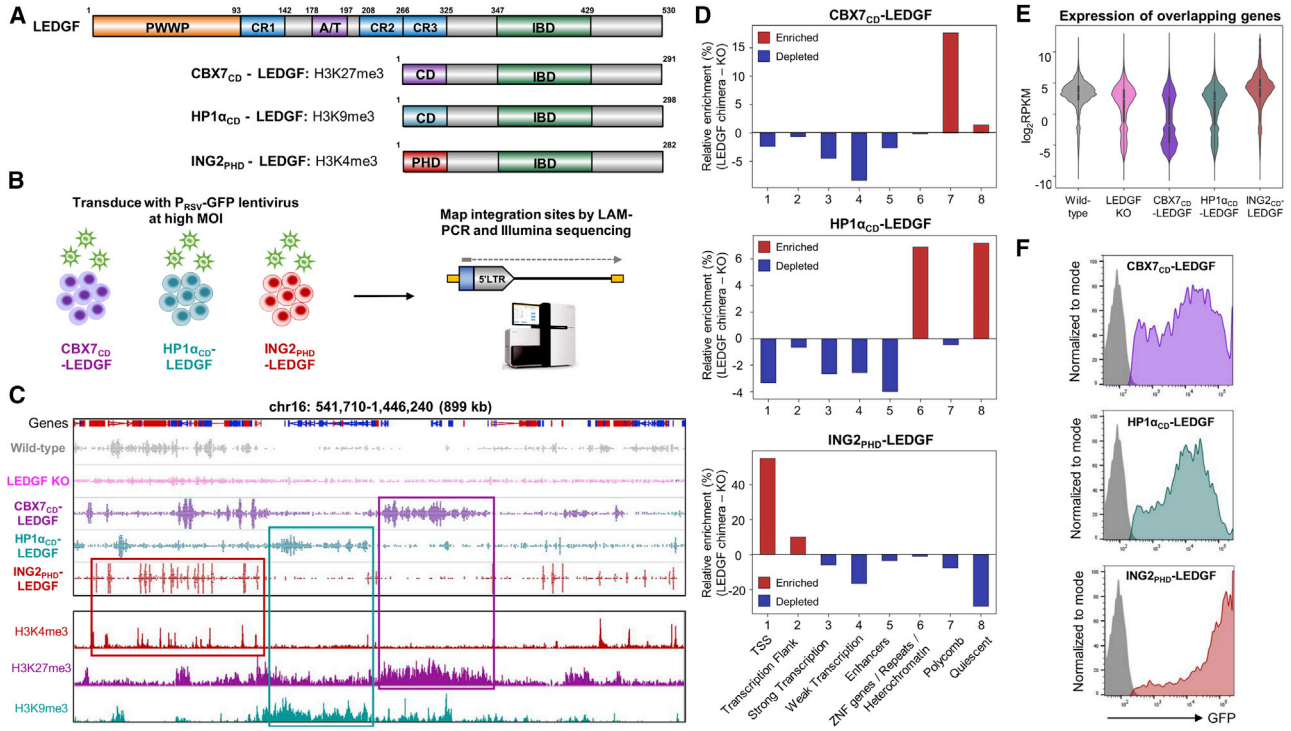


Figure 6. Retargeting lentiviral integrants through LEDGF chimeras

(A) Overview of LEDGF chimera design.

(B) Schematic representation of the experiment designed to map lentiviral integration sites in LEDGF chimera-expressing cells.

(C and D) LEDGF chimeras efficiently retarget lentiviral integrants to the expected histone modifications (C) and chromatin states (D).

(E and F) Correlation between the expression of integrants and that of their overlapping genes. (E) Violin plot summarizing the distribution of gene expression levels across all genes containing lentiviral integrants in cells expressing the indicated LEDGF chimeras. The CBX7$_{CD}$- and HP1α_{CD}-LEDGF chimeras facilitate lentiviral integration into poorly expressed genes, while expression of the ING2$_{PHD}$-LEDGF chimera biases integration toward more highly expressed genes. (F) The impact of this is mirrored in GFP transgene expression, as measured by flow cytometry (gray histograms indicate untransduced cells; colored histograms indicate GFP$^+$ cells).

CBX7$_{CD}$-LEDGF- and HP1α_{CD}-LEDGF-expressing cells were more frequently found in intergenic regions compared with integrations in LEDGF KO cells, whereas the majority of reporters in ING2$_{PHD}$-LEDGF cells were found in genes (Figure S4B). Integrants in CBX7$_{CD}$-LEDGF and HP1α_{CD}-LEDGF cells were enriched in both poorly expressed and silent genes, while integrants in ING2$_{PHD}$-LEDGF cells were biased toward well-expressed genes (Figure 6E). This bias appeared to be strongest in CBX7$_{CD}$-LEDGF cells and ING2$_{PHD}$-LEDGF cells; indeed, we observed a high number of "dim" reporters in CBX7$_{CD}$-LEDGF cells, whereas reporter expression was significantly higher in ING2$_{PHD}$-LEDGF cells (Figure 6F).

TRACE in CBX7$_{CD}$-LEDGF cells identifies reporter repressed by SUZ12

We sought to validate the utility of LEDGF chimera-expressing cells by performing a TRACE experiment in the CBX7$_{CD}$-LEDGF population to identify reporters of Polycomb-mediated repression. Polycomb-group (PcG) proteins form multimeric chromatin-bound complexes—prominently Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2)—that repress lineage-specific genes to ensure proper mammalian development (Schuettengruber et al., 2017). Despite the crucial role that these complexes play during development, a complete catalog of the proteins involved has not yet been delineated, and the spectrum of mechanisms of Polycomb-mediated repression are not fully understood (Fursova et al., 2019). Thus, the generation of Polycomb-repressed reporter clones would facilitate genetic dissection of Polycomb function in mammalian cells. We focused on PRC2, as the frequent dysregulation of PRC2 function in cancer has led to the development of multiple small molecule inhibitors that could be used in a TRACE assay: UNC1999 and GSK343 inhibit the catalytic function of enhancer of zeste homolog 2 (EZH2) and thus prevent H3K27me3 deposition (Konze et al., 2013; Verma et al., 2012), while EED226 blocks H3K27me3 recognition by embryonic ectoderm development protein (EED) (Qi et al., 2017).

We established a TRACE reporter library in CBX7$_{CD}$-LEDGF cells, wherein >80% of integrants were located in H3K27me3-marked genomic sites, and compared transgene expression in cells treated for 7 days with DMSO versus UNC1999, GSK343, and EED226 (Figure 7A). Although H3K27me3 was undetectable following treatment with each of the three inhibitors (Figure S5A), surprisingly we did not observe any global increase in GFP transgene expression (Figure S5B). In total, we identified only

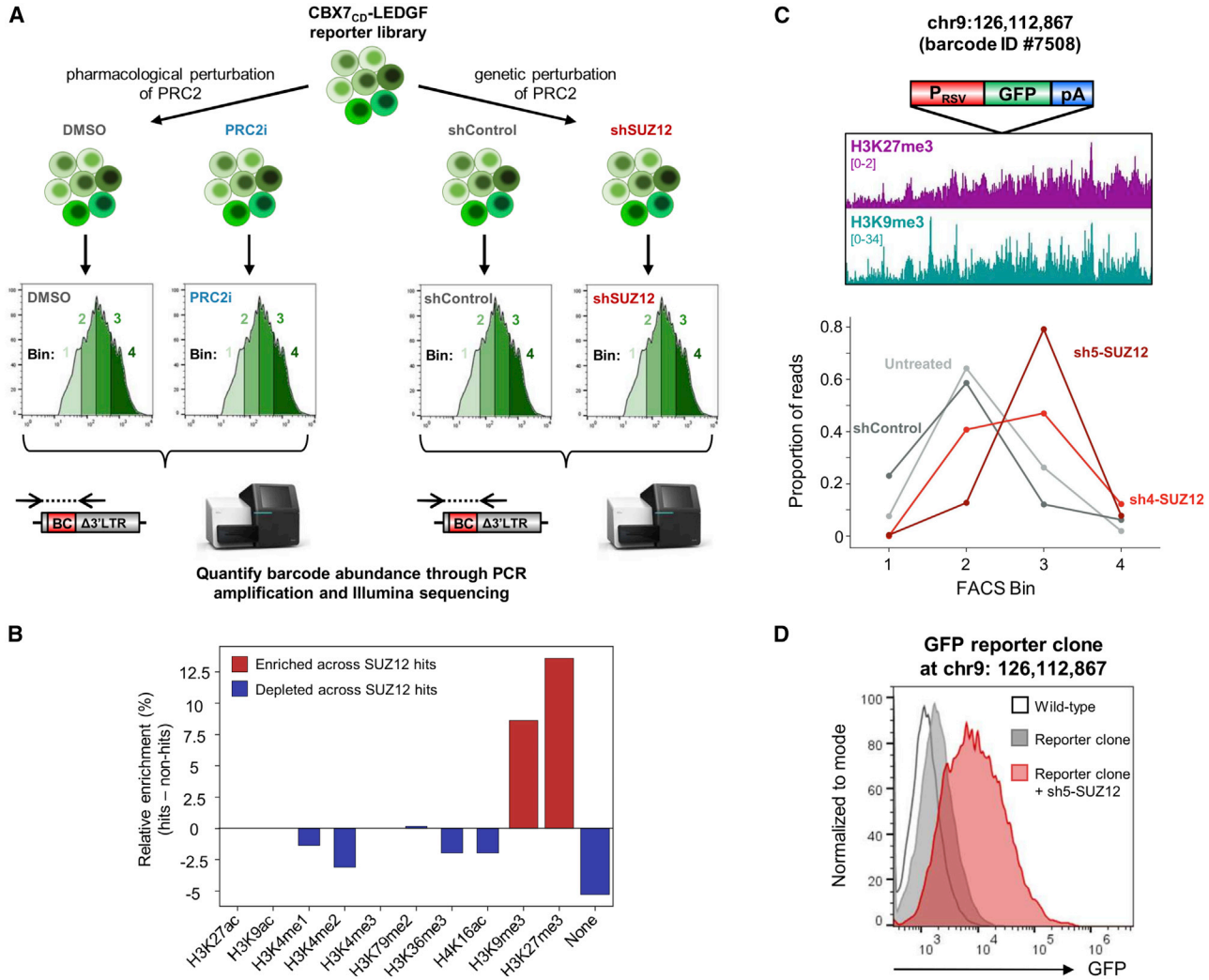


Figure 7. TRACE identifies reporters responsive to loss of the PRC2 subunit SUZ12

(A) Schematic representation of the TRACE experiment designed to identify reporters exhibiting increased expression upon treatment with PRC2 inhibitors (UNC1999, GSK343, and EED226) or shRNA-mediated depletion of the PRC2 subunit SUZ12.

(B) Genomic sites harboring reporters responsive to SUZ12 depletion are enriched for H3K27me3 and H3K9me3. Reporter integration sites were intersected with peaks of the indicated histone modifications, and the proportion of reporters overlapping each histone modification was calculated. The barplot depicts the enrichment or depletion of each modification comparing the genomic sites harboring reporters responsive to SUZ12 depletion ("hits") versus non-responsive reporters ("non-hits").

(C and D) Generation of a reporter clone responsive to SUZ12 disruption. (C) Expression profile of one of the leading reporters (chr9: 126,112,867), which exhibited significant derepression upon depletion of SUZ12. The reporter resides in a genomic locus marked by both H3K27me3 and H3K9me3. (D) CRISPR/Cas9-mediated knockin to the genomic site described in (C) resulted in the generation of a reporter clone responsive to SUZ12 depletion in KBM-7 cells.

37 reporters that were concordantly derepressed by all three inhibitors (Figure S5C). Thus, pharmacological inhibition of H3K27me3 deposition or recognition was insufficient to alleviate PRC2-mediated repression across transgenes located in PRC2-targeted genomic regions.

In parallel, we performed an analogous TRACE experiment by genetically perturbing PRC2 activity. We tested the efficacy of a panel of short hairpin RNAs (shRNAs) targeting the PRC2 subunit SUZ12 (Figure S5D) and chose the two optimal shRNAs to perform a TRACE experiment using the same $CBX7_{CD}$-LEDGF chimera transgene library (Figure 7A). In contrast to PRC2 inhibitor treatment, loss of SUZ12 resulted in a global increase in GFP expression; moreover, the degree of derepression was correlated with the extent of SUZ12 depletion (Figures S5D and S5E). Overall, the TRACE experiment identified 106 reporters that were strongly and concordantly derepressed by both shRNAs (Figure S5F), only one of which was also upregulated upon PRC2 inhibitor treatment (Figure S5G).

As expected, the genomic loci harboring SUZ12-responsive reporters were enriched for the H3K27me3 histone modification (Figure 7B). Surprisingly, however, we found that these genomic sites were also marked by H3K9me3, a histone modification not

normally associated with Polycomb-mediated silencing (Figures 7B and S6A–S6C). To validate the putative role of SUZ12 in mediating silencing across genomic sites marked by both repressive modifications, we used CRISPR/Cas9-mediated knockin to generate a fluorescent reporter clone harboring a GFP expression cassette at one of the top loci identified by the TRACE screen (Figure 7C). Two weeks after nucleofection, single-cell KBM-7 clones were isolated from the GFP$^+$ population by FACS, and integration of the P_{RSV}-GFP-pA insert at the intended target site was confirmed (Figure S5H). Crucially, loss of SUZ12 in this reporter clone—both through shRNA-mediated depletion and CRISPR/Cas9-mediated gene disruption—did indeed result in a robust increase in GFP expression (Figures 7D, S5I, and S5J). Thus, genetic perturbation of PRC2 function in $CBX7_{CD}$-LEDGF reporter cells successfully identified genomic sites responsive to SUZ12 depletion and facilitated the generation of a SUZ12-responsive fluorescent reporter clone. The colocalization of H3K27me3 and H3K9me3 in the vicinity of reporters derepressed upon loss of SUZ12 suggests that PRC2-mediated repression across certain genomic sites may be exerted through a non-canonical route involving H3K9me3; alternatively, it is possible that SUZ12-mediated repression across these genomic loci is independent of its role within PRC2. Future genetic interrogation of reporter clones at such H3K27me3/H3K9me3-marked sites would be necessary to distinguish between these two possibilities.

Generation of TRACE libraries using alternative promoters and cell lines

A wide range of GFP expression is critical to the resolution of TRACE. We had previously conducted a small-scale screen of promoters and cell lines that showed that the broadest spread of GFP expression was achieved using the RSV promoter in the KBM-7 cell line (Tchasovnikarova et al., 2015). However, a more comprehensive assessment of the performance of commonly used promoters across a range of standard human cell lines revealed multiple additional pairings that provide a broad spread of transgene expression (Figure S7A). Indeed, three combinations—the RSV promoter in the human monocytic THP-1 cell line, the cytomegalovirus (CMV) promoter in HeLa cells, and the mouse mammary tumor virus (MMTV) promoter in the karyotypically stable human telomerase reverse transcriptase (hTERT) RPE-1 cells—exhibited an even broader spread of GFP expression than that observed with the RSV promoter in KBM-7 cells (Figure S7A). Supporting the idea that TRACE will be successful with these promoter/cell type combinations, we found that lentiviral integrations in KBM-7, THP-1, HeLa, and RPE-1 cells were similarly distributed across the genome (Figures S7B–S7D). Thus, we anticipate that TRACE will be successful using other promoters and cell types (including non-cancerous cells), thereby permitting the study of cell-type-specific epigenetic factors.

DISCUSSION

TRACE allows for the identification of phenotypic fluorescent reporters of epigenetic pathways by sampling the expression of hundreds of thousands of GFP lentiviral transgenes distributed across the genome. Here, we show that TRACE can be used to generate reporters responsive to both pharmacological and genetic inhibition of epigenetic pathways, and demonstrate that manipulation of LEDGF allows TRACE to sample a wider range of chromatin environments. Thus, TRACE can be applied to generate fluorescent reporters for theoretically any epigenetic pathway of interest.

Choice of reporter expression readout

TRACE transgene expression can be assayed at both the transcript and protein level. A transcript-based readout allows for multiple replicate experiments to be performed in parallel with ease and does not rely on the availability of FACS sorting. However, it can be less cost-efficient due to the sequencing depth required to sample all reporters at sufficient coverage: well-expressed reporters make up the majority of reporter transcripts in the cell and will therefore take up a large proportion of the total sequencing reads. Thus, while 20 million sequencing reads may provide sufficient coverage of 200,000 equally expressed reporters, a far larger number would be required to achieve sufficient coverage of all reporters given the ~100-fold difference in expression across the reporter population.

A protein-based readout of GFP expression allows for even coverage of all reporters, as poorly expressed reporters are separated from well-expressed reporters in different collection bins. Thus, the extra cost associated with sorting is offset by the lower cost of Illumina sequencing required to achieve sufficient coverage of all reporters. Therefore, a transcript-based readout may be more appropriate when interrogating euchromatic processes, while a FACS-based readout may be more suited to identifying reporters of repressive pathways.

Choice of pathway perturbation

TRACE compares reporter expression in the presence and absence of a perturbation of a pathway of interest. Such perturbations can be achieved either through small molecule inhibitors, or genetic perturbations such as CRISPR interference (CRISPRi), RNA interference (RNAi), or the expression of exogenous constructs. The optimal method will likely depend on the specific goals of the experiment, but the key is that the chosen perturbation affects all the cells in the TRACE population. In this respect CRISPR/Cas9-mediated gene disruption is unlikely to be suitable, as only a proportion of the cells would be expected to contain the intended mutation.

We have demonstrated that both pharmacological and genetic perturbations can be used to interrogate an epigenetic pathway. Combining these strategies could identify potential off-target effects of small molecule inhibitors, and in the case of catalytic inhibitors, evaluate the requirement for enzymatic activity in the epigenetic process under investigation. Furthermore, as chromatin regulators often exert their activity as part of multisubunit complexes, genetic ablation of different subunits could identify locus-specific requirements for different complex members.

Choice of LEDGF background

The ability to alter the integration preferences of lentiviral vectors by manipulating the expression of LEDGF significantly expands

the scope of TRACE. The likelihood of identifying relevant reporters for euchromatic pathways may be increased by performing TRACE either in wild-type cells or cells expressing a LEDGF chimera targeting lentiviruses to an active histone modification of interest. By contrast, potential reporters for repressive pathways may make up a larger proportion of the TRACE population in the absence of LEDGF or upon expression of a LEDGF chimera targeting lentiviruses to the repressive histone modification of choice. Even if nothing is known about the pathway of interest, performing TRACE in LEDGF KO cells—where a more uniform sampling of the genome is achieved—should increase the probability of identifying a suitable reporter.

Comparison with TRIP

The TRIP methodology (Akhtar et al., 2013) was the first demonstration of the use of genetically barcoded reporters to measure chromatin position effects in a multiplexed fashion. Although the generation of barcoded TRACE integrant libraries was inspired by the TRIP approach, TRACE offers a number of advantages: (1) the purpose of TRACE is not only to quantify chromatin position effects but also to generate fluorescent reporter lines for epigenetic pathways of interest; (2) TRACE can be performed at a scale significantly larger than has been demonstrated for TRIP: here we report expression measurements at >600,000 reporter integration sites (Figure 4), whereas TRIP experiments were conducted by monitoring 10,000–20,000 sites (Akhtar et al., 2013); (3) transgene expression in TRACE can be measured at both the transcript and protein level, whereas the use of multi-copy transposable reporter constructs limits the readout of TRIP to the transcript level; and (4) the ability to alter the integration preferences of lentiviral vectors by manipulating the expression of LEDGF significantly expands the scope of TRACE. This is particularly powerful when studying repressive pathways, as heterochromatic areas of the genome are refractory to lentiviral integration.

Pharmacological inhibition of PRC2 function is insufficient to alleviate Polycomb-mediated transgene repression

The use of both small molecule inhibitors and shRNA allowed us to assess the requirements for H3K27me3 deposition in PRC2-mediated repression. Despite the large proportion of integrants located in H3K27me3-marked areas of the genome, we were surprised to find that loss of H3K27me3 through inhibition of EZH2 catalytic activity or EED-mediated chromatin localization did not result in significant transgene derepression. By contrast, depletion of SUZ12 did result in transgene derepression. We consider two possible explanations: (1) EZH2 enzymatic activity may be insufficient to establish repression, which would be consistent with previous reports that show that H3K27me3 is necessary, but not sufficient, to fully specify repressed domains across Polycomb-regulated regions (Pengelly et al., 2013) and that EZH2-mediated deposition of H3K27me3, but not EZH2 itself, is dispensable for differentiation (Lavarone et al., 2019); and alternatively (2) SUZ12 could perform additional repressive functions independent of PRC2. Thus, further genetic interrogation of SUZ12-responsive knockin clones would be necessary to dissect the role of Polycomb proteins at these sites.

Interestingly, some of the reporters responsive to SUZ12 depletion identified by our TRACE screen in $CBX7_{CD}$-LEDGF chimera cells resided in genomic loci marked by H3K9me3. While trimethylation of H3K9 is not usually associated with Polycomb-mediated silencing, a number of reports have hinted at cooperation between these two repressive pathways (Bernstein et al., 2006; Margueron et al., 2009; Mozzetta et al., 2014; Voigt et al., 2012; Wang et al., 2008). Functional interrogation of the TRACE-generated knockin clones harboring GFP reporters at H3K9me3-marked genomic sites will be needed to dissect the potential role of H3K9 methylation in SUZ12-mediated repression.

Limitations

The method to map lentiviral integration sites described here relies on the use of a single restriction enzyme, NlaIII, to digest the genomic DNA extracted from the reporter population. As a proportion of all the lentiviral integration sites will lie either too close or too far from an NlaIII recognition site to be mapped, the genomic DNA could also be digested with additional "4-cutter" restriction enzymes that are not affected by CpG methylation to increase the proportion of reporters that can be assigned to their genomic integration sites. Adapting the protocol for additional enzyme(s) would require redesign of the lentiviral vector to remove cleavage sites of those enzymes from the vector backbone, such that PCR amplification of the virus-genome junctions is not prevented.

Lentiviral integration into a gene might disrupt its function in a manner that impacts the downstream analysis of a pathway of interest. However, the lentiviral vectors used in TRACE do not contain the splice acceptor or donor sequences that feature in the majority of gene-trap vectors (Friedel and Soriano, 2010); hence, we do not expect them to efficiently disrupt gene function unless integrated into an exon. Nevertheless, care should be taken when choosing reporter integration sites for downstream analysis by prioritizing reporters that have integrated in a reverse orientation in a gene, or in intergenic regions of the genome.

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
  - Cell Culture
- METHOD DETAILS
  - TRACE
  - Lentiviral Mapping of TRACE Integrants
  - Small molecule inhibitors
  - Lentiviral Expression Vectors
  - Plasmids
  - Lentivirus Production
  - Flow Cytometry

- Immunoblotting
- Purification of Cas9
- Purification of Moloney Murine Leukemia Virus (MMLV) Reverse Transcriptase
- CRISPR/Cas9-mediated genome editing
- ChIP-seq

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Histone modifications analysis
  - Chromatin states analysis
- ADDITIONAL RESOURCES
  - Detailed protocol

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2021.11.035.

ACKNOWLEDGMENTS

We thank the HSCI-CRM Flow Cytometry Core Facility (MGH, Boston, USA) and the School of Biological Sciences Flow Cytometry Core Facility (University of Cambridge, UK) for FACS. We are grateful to Julie Ahringer for critical reading of the manuscript. I.A.T. is a Damon Runyon-Dale F. Frey Breakthrough Scientist supported (in part) by the Damon Runyon Cancer Research Foundation (DFS-2277-16). This work was supported by NIH grants R01-GM4390121A1 and R35-GM131743 to R.E.K.

AUTHOR CONTRIBUTIONS

Conceptualization, I.A.T. and R.E.K.; methodology, I.A.T.; investigation, I.A.T.; formal analysis, I.A.T. and M.D.; resources, S.K.M.; writing – original draft, I.A.T.; writing – review & editing, I.A.T. and R.E.K.; supervision, R.E.K.; funding acquisition, I.A.T. and R.E.K.

DECLARATION OF INTERESTS

R.E.K. is a member of Molecular Cell's Advisory Board.

Received: March 22, 2021
Revised: September 14, 2021
Accepted: November 29, 2021
Published: December 27, 2021

REFERENCES

Akhtar, W., de Jong, J., Pindyurin, A.V., Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L.F.A., van Lohuizen, M., and van Steensel, B. (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell *154*, 914–927.

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell. Biol. *26*, 2560–2569.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J. Biol. Chem. *278*, 372–381.

Eissenberg, J.C., Morris, G.D., Reuter, G., and Hartnett, T. (1992). The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. Genetics *131*, 345–352.

Ferris, A.L., Wu, X., Hughes, C.M., Stewart, C., Smith, S.J., Milne, T.A., Wang, G.G., Shun, M.C., Allis, C.D., Engelman, A., and Hughes, S.H. (2010). Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. Proc. Natl. Acad. Sci. USA *107*, 3135–3140.

Friedel, R.H., and Soriano, P. (2010). Gene trap mutagenesis in the mouse. Methods Enzymol. *477*, 243–269.

Fukuda, K., Okuda, A., Yusa, K., and Shinkai, Y. (2018). A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. Genome Res. *28*, 846–858.

Fursova, N.A., Blackledge, N.P., Nakayama, M., Ito, S., Koseki, Y., Farcas, A.M., King, H.W., Koseki, H., and Klose, R.J. (2019). Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. Mol. Cell *74*, 1020–1036.e8.

Gijsbers, R., Ronen, K., Vets, S., Malani, N., De Rijck, J., McNeely, M., Bushman, F.D., and Debyser, Z. (2010). LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. Mol. Ther. *18*, 552–560.

Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, Mdel.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat. Biotechnol. *32*, 267–273.

Konze, K.D., Ma, A., Li, F., Barsyte-Lovejoy, D., Parton, T., Macnevin, C.J., Liu, F., Gao, C., Huang, X.P., Kuznetsova, E., et al. (2013). An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1. ACS Chem. Biol. *8*, 1324–1334.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Lackner, D.H., Carré, A., Guzzardo, P.M., Banning, C., Mangena, R., Henley, T., Oberndorfer, S., Gapp, B.V., Nijman, S.M.B., Brummelkamp, T.R., and Bürckstümmer, T. (2015). A generic strategy for CRISPR-Cas9-mediated gene tagging. Nat. Commun. *6*, 10237.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lavarone, E., Barbieri, C.M., and Pasini, D. (2019). Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. Nat. Commun. *10*, 1679.

Liu, N., Lee, C.H., Swigut, T., Grow, E., Gu, B., Bassik, M.C., and Wysocka, J. (2018). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. Nature *553*, 228–232.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10–12.

Mehta, S., and Zhang, J. (2011). Reporting from the field: genetically encoded fluorescent reporters uncover signaling dynamics in living biological systems. Annu. Rev. Biochem. *80*, 375–401.

Mohammad, H.P., Smitheman, K.N., Kamat, C.D., Soong, D., Federowicz, K.E., Van Aller, G.S., Schneck, J.L., Carson, J.D., Liu, Y., Butticello, M., et al. (2015). A DNA Hypomethylation Signature Predicts Antitumor Activity of LSD1 Inhibitors in SCLC. Cancer Cell *28*, 57–69.

Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., Margueron, R., and Ait-Si-Ali, S. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. Mol. Cell *53*, 277–289.

Pengelly, A.R., Copur, Ö., Jäckle, H., Herzig, A., and Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. Science *339*, 698–699.

Qi, W., Zhao, K., Gu, J., Huang, Y., Wang, Y., Zhang, H., Zhang, M., Zhang, J., Yu, Z., Li, L., et al. (2017). An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED. Nat. Chem. Biol. *13*, 381–388.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. (2000).

Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

Reuter, G., and Wolff, I. (1981). Isolation of dominant suppressor mutations for position-effect variegation in Drosophila melanogaster. Mol. Gen. Genet. *182*, 516–519.

Robbez-Masson, L., Tie, C.H.C., Conde, L., Tunbak, H., Husovsky, C., Tchasovnikarova, I.A., Timms, R.T., Herrero, J., Lehner, P.J., and Rowe, H.M. (2018). The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. Genome Res. *28*, 836–845.

Schuettengruber, B., Bourbon, H.M., Di Croce, L., and Cavalli, G. (2017). Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. Cell *171*, 34–57.

Shun, M.C., Raghavendra, N.K., Vandegraaff, N., Daigle, J.E., Hughes, S., Kellam, P., Cherepanov, P., and Engelman, A. (2007). LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. Genes Dev. *21*, 1767–1778.

Starmer, J., and Magnuson, T. (2016). Detecting broad domains and narrow peaks in ChIP-seq data with hiddenDomains. BMC Bioinformatics *17*, 144.

Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M.A., and Lehner, P.J. (2015). GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science *348*, 1481–1485.

Tchasovnikarova, I.A., Timms, R.T., Douse, C.H., Roberts, R.C., Dougan, G., Kingston, R.E., Modis, Y., and Lehner, P.J. (2017). Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. Nat. Genet. *49*, 1035–1044.

Timms, R.T., Tchasovnikarova, I.A., Antrobus, R., Dougan, G., and Lehner, P.J. (2016). ATF7IP-Mediated Stabilization of the Histone Methyltransferase SETDB1 Is Essential for Heterochromatin Formation by the HUSH Complex. Cell Rep. *17*, 653–659.

Timms, R.T., Tchasovnikarova, I.A., and Lehner, P.J. (2019). Differential viral accessibility (DIVA) identifies alterations in chromatin architecture through large-scale mapping of lentiviral integration sites. Nat. Protoc. *14*, 153–170.

Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G., and Reuter, G. (1994). The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. EMBO J. *13*, 3822–3831.

Turlure, F., Maertens, G., Rahman, S., Cherepanov, P., and Engelman, A. (2006). A tripartite DNA-binding element, comprised of the nuclear localization signal and two AT-hook motifs. mediates the association of LEDGF/p75 with chromatin in vivo *34*, 1653–1665.

Verma, S.K., Tian, X., LaFrance, L.V., Duquenne, C., Suarez, D.P., Newlander, K.A., Romeril, S.P., Burgess, J.L., Grant, S.W., Brackley, J.A., et al. (2012). Identification of potent, selective, cell-Active inhibitors of the histone lysine methyltransferase EZH2. ACS Med. Chem. Lett. *3*, 1091–1096.

Voigt, P., LeRoy, G., Drury, W.J., 3rd, Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181–193.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897–903.

Weekes, M.P., Tan, S.Y., Poole, E., Talbot, S., Antrobus, R., Smith, D.L., Montag, C., Gygi, S.P., Sinclair, J.H., and Lehner, P.J. (2013). Latency-associated degradation of the MRP1 drug transporter during latent human cytomegalovirus infection. Science *340*, 199–202.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-H3K27ac	Active Motif	Cat#39134; RRID:AB_2722569
Rabbit monoclonal anti-H3K9ac	Cell Signaling Technology	Cat#9649; RRID:AB_823528
Rabbit polyclonal anti-H3K4me1	Abcam	Cat#ab8895; RRID:AB_306847
Rabbit polyclonal anti-H3K4me2	Abcam	Cat#ab7766; RRID:AB_2560996
Rabbit monoclonal anti-H3K4me3	Cell Signaling Technology	Cat#9751; RRID:AB_2616028
Rabbit monoclonal anti-H3K79me2	Cell Signaling Technology	Cat#5427; RRID:AB_10693787
Rabbit polyclonal anti-H3K36me3	Abcam	Cat#ab9050; RRID:AB_306966
Rabbit polyclonal anti-H4K16ac	EMD Millipore	Cat#07-329; RRID:AB_310525
Rabbit monoclonal anti-H3K27me3, used for ChIP	GeneTex	Cat#GTX60892; RRID:AB_2888006
Rabbit monoclonal anti-H3K27me3, used for immunoblot	Cell Signaling Technology	Cat#9733; RRID:AB_2616029
Rabbit polyclonal anti-LEDGF	Bethyl Laboratories	Cat#A300-848A-T; RRID:AB_2171223
Rabbit polyclonal anti-SUZ12	Proteintech	Cat#20366-1-AP; RRID:AB_10694152
Mouse monoclonal anti-β-actin	Sigma-Aldrich	Cat#A2228; RRID:AB_476697
Peroxidase AffiniPure Donkey Anti-Rabbit IgG (H+L)	Jackson ImmunoResearch	Cat#711-035-152; RRID:AB_10015282
Peroxidase AffiniPure Donkey Anti-Mouse IgG (H+L)	Jackson ImmunoResearch	Cat#715-035-150; RRID:AB_2340770
Bacterial and virus strains		
NEB 5-alpha Competent *E. coli*	New England Biolabs	Cat#C2987H
Chemicals, peptides, and recombinant proteins		
GSK2879552 2HCl	Selleck Chemical	Cat#S7796
UNC1999	Sigma-Aldrich	Cat# SML0778
GSK343	Sigma-Aldrich	Cat#SML0766
EED226	Cayman Chemical	Cat#22031
Cas9	This paper	N/A
Critical commercial assays		
NEBuilder HiFi DNA Assembly Cloning Kit	New England Biolabs	Cat#E5520S
Gentra Puregene Cell Kit	QIAGEN	Cat#158745
QIAquick PCR purification kit	QIAGEN	Cat#28104
Macherey-Nagel NucleoSpin RNA Midi Kit	Fisher Scientific	Cat#10551443
SF Cell Line 4D-NucleofectorTM X Kit L	Lonza	Cat#V4XC-2012
NEBNext Ultra II DNA Library Prep Kit for Illumina	New England Biolabs	Cat#E7645L
Deposited data		
Raw and analyzed data	This paper	GEO: GSE182748
Experimental models: Cell lines		
KBM-7	Paul Lehner laboratory; Tchasovnikarova et al., 2015	N/A
HEK293T	Paul Lehner laboratory; Tchasovnikarova et al., 2015	N/A

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
HeLa	Paul Lehner laboratory; Tchasovnikarova et al., 2015	N/A
THP-1	Paul Lehner laboratory; Weekes et al., 2013	N/A
hTERT RPE-1	ATCC	Cat#CRL-4000
Oligonucleotides		
See Table S1	This paper	N/A
Recombinant DNA		
pHRSIN-P_{RSV}-GFP-WPRE-BC	This paper	N/A
pHRSIN-P_{MMTV}-GFP-WPRE	This paper	N/A
pHRSIN-P_{CMV}-GFP-WPRE	This paper	N/A
pHRSIN-P_{SFFV}-$CBX7_{CD}$-LEDGF-P_{PGK}-$Puro^R$	This paper	N/A
pHRSIN-P_{SFFV}-$HP1\alpha_{CD}$-LEDGF-P_{PGK}-$Puro^R$	This paper	N/A
pHRSIN-P_{SFFV}-$ING2_{PHD}$-LEDGF-P_{PGK}-$Puro^R$	This paper	N/A
pC-SIREN-sh1-SUZ12	This paper	N/A
pC-SIREN-sh2-SUZ12	This paper	N/A
pC-SIREN-sh3-SUZ12	This paper	N/A
pC-SIREN-sh4-SUZ12	This paper	N/A
pC-SIREN-sh5-SUZ12	This paper	N/A
pC-SIREN-sh6-SUZ12	This paper	N/A
pC-SIREN-sh7-SUZ12	This paper	N/A
pKLV-U6gRNA(BbsI)-PGKpuro2ABFP	Addgene	50946
pET-NLS-Cas9-6xHis	Addgene	62934
pCR28-MLV RT-M5	Michael Marr laboratory	N/A
P_{RSV}-GFP-pA	This paper	N/A
Software and algorithms		
Cutadapt	Martin, 2011	https://cutadapt.readthedocs.io/en/stable/index.html
Bowtie 2	Langmead and Salzberg, 2012; Langmead et al., 2009	http://bowtie-bio.sourceforge.net/index.shtml
Seaborn visualization library for Python	Seaborn	https://seaborn.pydata.org/
hiddenDomains	Starmer and Magnuson, 2016	http://hiddendomains.sourceforge.net
bedtools	bedtools	http://bedtools.readthedocs.io/en/latest/
SeqMonk	Babraham Bioinformatics Group	https://www.bioinformatics.babraham.ac.uk/projects/seqmonk
FlowJo	FlowJo	https://www.flowjo.com
Other		
Detailed TRACE protocol	This paper	Methods S1

RESOURCE AVAILABILITY

Lead contact

Further information and request for resources and reagents should be directed to the lead contact, Iva A. Tchasovnikarova (it257@cam.ac.uk).

Materials availability

Commercially available reagents are listed in the Key resources table. All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement when applicable.

Data and code availability

- The ChIP-seq data have been deposited to GEO under GSE182748.
- Routine bioinformatics approaches used to analyze the data are described in relevant STAR Methods sections.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Culture

KBM-7 cells, HEK293T, HeLa and THP-1 cells were a kind gift from Prof. Paul Lehner (CITIID, Cambridge, UK). KBM-7 and HEK293T cells were cultured in IMDM supplemented with 10% fetal calf serum (FCS) and penicillin/streptomycin (100 U/ml), while HeLa and THP-1 cells were cultured in RPMI supplemented with 10% fetal calf serum (FCS) and penicillin/streptomycin (100 U/ml). hTERT RPE-1 cells were purchased from ATCC (CRL-4000) and cultured in DMEM/F-12 supplemented with 10% fetal calf serum (FCS) and penicillin/streptomycin (100 U/ml). All cell lines were routinely tested for mycoplasma contamination (ATCC Universal Mycoplasma Detection Kit).

METHOD DETAILS

TRACE

For the TRACE experiment aimed at identifying reporters responsive to LSD1i treatment (Figures 3 and S1), one million KBM-7 cells were transduced with the barcoded lentiviral library at a range of MOIs, and the proportion of GFP^+ cells in each population was assessed by flow cytometry 2 days later. The population containing ~10% GFP^+ cells (equating to ~100,000 unique integrants) was cultured for a further 5 days, following which GFP^+ cells were isolated by FACS. The resulting population was expanded and frozen in aliquots of 20 million cells, such that at least 100-fold coverage of the library would be maintained upon defrosting. To identify LSD1i-responsive reporters, a vial of the GFP^+ cells was defrosted, expanded for 3 days and then divided in two (10 million cells each): one half was treated with DMSO and served as a control, while the other half was treated with 1 μM LSD1i. Cells were cultured in the presence of either DMSO or LSD1i for 7 days, at which point GFP reporter expression was analyzed by FACS followed by Illumina sequencing (described below).

For the TRACE experiment aimed at identifying reporters responsive to pharmacological inhibition of PRC2 function treatment (Figures 7 and S5), one million $CBX7_{CD}$-LEDGF-expressing KBM-7 cells were transduced with the barcoded lentiviral library at a range of MOIs, and the population containing ~20% GFP^+ cells (equating to ~200,000 unique integrants) was cultured for a further 5 days, following which GFP^+ cells were isolated by FACS. The resulting population was expanded and frozen in aliquots of 30 million cells, such that at least 100-fold coverage of the library would be maintained upon defrosting. To identify LSD1i-responsive reporters, a vial of the GFP^+ cells was defrosted, expanded for 7 days and split across five conditions (20 million cells each): untreated cells served as a control, while the rest were treated either with DMSO or with 1 μM of UNC1999, GSK343 or EED226. Cells were cultured for 7 days, at which point GFP reporter expression was analyzed by FACS followed by Illumina sequencing (described below).

For the TRACE experiment aimed at identifying reporters responsive to loss of SUZ12 (Figures 7 and S5), a vial of the $CBX7_{CD}$-LEDGF TRACE population was defrosted, expanded for 7 days and split across four conditions (20 million cells each): untreated cells served as a control, while the remaining three populations were transduced either with a non-targeting shRNA (shControl) or one of two SUZ12-targeting shRNAs (sh4-SUZ12 and sh5-SUZ12). Two days after transduction, cells were selected with 0.75 μg/ml Puromycin (Thermo Fisher) and following a further five days of culture, GFP expression was analyzed by FACS followed by Illumina sequencing (described below).

To assay GFP expression by FACS, the untreated population was used to set four equal gates of 25% each and 20 million cells were sorted, collecting 5 million cells in each bin. Treated samples were sorted in the same way, using the gates set on the untreated population. Cells were lysed immediately after sorting, and genomic DNA was extracted using the Gentra Puregene Cell Kit (QIAGEN). Barcode cassettes were amplified through 18 cycles of exponential PCR using Q5 High-Fidelity DNA Polymerase (NEB), a pool of staggered primers binding upstream of the 24 bp barcode sequence and a single primer binding downstream of the 24 bp barcode sequence (see Table S1 for primer sequences). Following a column-based clean-up (QIAquick PCR purification kit), PCR products were quantified and 200 ng used to add the Illumina P5 and P7 sequences through 7 cycles of further exponential PCR. Following a final clean-up (QIAquick PCR purification kit), PCR products were quantified, pooled and sequenced on an Illumina HiSeq4000 instrument (GENEWIZ). The resulting sequence reads were trimmed of invariant lentiviral sequences using Cutadapt, yielding a 24-nucleotide barcode sequence. Barcodes were mapped to a custom index describing the barcode sequences and their corresponding genomic locations using Bowtie 2. The resulting raw count table was normalized for sequencing depth. The relative GFP expression of each integrant was determined by assessing the distribution of sequencing reads for each barcode across the four FACS bins, by calculating a gene expression metric (GEM): $GEM = \sum_{i=1}^{4} Ri \times i$ (where i is the FACS bin and Ri is the fraction of the sequencing reads present for that reporter in the given bin i).

To assay GFP transcript abundance, RNA and genomic DNA were extracted from the TRACE cell population described in Figure 4A using the Macherey-Nagel NucleoSpin RNA Midi Kit (Fisher Scientific) or Gentra Puregene Cell Kit (QIAGEN), respectively. All of the RNA was converted to cDNA using MMLV RT and a custom primer binding downstream of the 24 bp barcode (see Table S1 for primer sequences). Following reverse transcription, any remaining RNA was digested with RNaseA for 30 min at 37°C, and cDNA was purified using Agencourt AMPure XP beads (Beckman Coulter) at a ratio of 1.8X. Barcode cassettes were then amplified through 11 cycles of exponential PCR, followed by a further 7 cycles as described above. Following a final clean-up (QIAquick PCR purification kit), PCR products were quantified, pooled and sequenced on an Illumina HiSeq4000 instrument (GENEWIZ). The resulting sequence reads were trimmed of invariant lentiviral sequences using Cutadapt, yielding a 24-nucleotide barcode sequence. Barcodes were mapped to a custom index describing the barcode sequences and their corresponding genomic locations using Bowtie 2. The resulting raw count table was normalized for sequencing depth. The relative GFP expression of each integrant was determined by normalizing barcode transcript counts to their abundance in the corresponding genomic DNA samples.

Lentiviral Mapping of TRACE Integrants

Following lentiviral transduction of target cells, genomic DNA was extracted using the Gentra Puregene Cell Kit (QIAGEN) and digested with NlaIII overnight at 37°C. Due to the location of NlaIII sites in the lentiviral vector, DNA fragments containing virus-genome junctions are larger than ~1.8 kb; therefore, large fragments were enriched using Agencourt AMPure XP beads (Beckman Coulter) at a ratio of 0.5X. NlaIII overhang-compatible double-stranded DNA adaptors were then ligated onto the remaining DNA fragments using T4 DNA ligase (NEB) overnight at 16°C. Unbound adaptors were removed through a further AMPure bead clean-up at a ratio of 0.65X. Virus-genome junctions were then amplified through 200 cycles of linear PCR using Herculase II Fusion DNA Polymerase (Agilent) and a biotinylated primer binding upstream of the 5′LTR in the integrated provirus (see Table S1 for primer sequences). The resulting biotinylated products were annealed to streptavidin-coated M-280 beads (Thermo Fisher Scientific) and washed at least 5 times in PBS plus 0.1% Tween-20 for 10 minutes each. Virus-genome junctions were then amplified through 18 cycles of on-bead exponential PCR using Herculase II Fusion DNA Polymerase (Agilent), a pool of staggered forward primers binding upstream of the 24 bp barcode sequence and a pool of staggered reverse primers binding the adaptor. Following an AMPure bead clean-up at a ratio of 0.9X, PCR products were quantified and 200 ng used to add the Illumina P5 and P7 sequences through 7 cycles of further exponential PCR. Following a final AMPure bead clean-up at a ratio of 0.9X, PCR products were quantified, pooled and sequenced on an Illumina HiSeq4000 instrument (GENEWIZ) generating 150 bp paired-end reads. The resulting sequence reads were trimmed of invariant lentiviral sequences using Cutadapt, yielding a 24-nucleotide barcode sequence and flanking genomic sequence for each read. Genomic sequences were mapped to the human genome (GRCh38) using Bowtie 2, such that barcoded integrants could be matched to their corresponding genomic location. Barcodes which could not be unambiguously assigned to a single chromosomal location were discarded.

Small molecule inhibitors

The following small molecule inhibitors were used: GSK2879552 2HCl (Selleck Chemical, S7796), UNC1999 (Sigma-Aldrich, SML0778), GSK343 (Sigma-Aldrich, SML0766) and EED226 (Cayman Chemical, 22031). KBM-7 cells were cultured in the presence of 1 μM of each inhibitor for 7 days.

Lentiviral Expression Vectors

Expression of GFP in all TRACE experiments was achieved using the pHRSIN-P_{RSV}-GFP-WPRE vector described previously (Tchasovnikarova et al., 2015), with the barcode cassette inserted into a KpnI site in the U3 region of the 3′LTR. For expression of GFP in THP-1, HeLa and hTERT RPE-1 cells, we replaced the RSV promoter in the pHRSIN-P_{RSV}-GFP-WPRE vector with either the CMV or MMTV promoters. For exogenous expression of LEDGF, we used the pHRSIN-P_{SFFV}-GFP-P_{PGK}-$Puro^R$ lentiviral vector (Tchasovnikarova et al., 2015): the fragments encoding the relevant chromatin reader domain ($CBX7_{CD}$, $HP1\alpha_{CD}$ or $ING2_{PHD}$) and the C terminus of LEDGF were amplified by PCR from cDNA and inserted in place of GFP using the HiFi Assembly method (NEB). For lentiviral expression of shRNA constructs, the pHR-SIREN vector was used with hairpins cloned in as BamHI-EcoRI fragments (see Table S1 for shRNA sequences).

Plasmids

The donor plasmid used for NHEJ knock-ins contains the RSV promoter driving GFP followed by the SV40 polyadenylation signal. These cassettes were amplified by PCR and inserted into the pMAX-GFP vector (Amaxa) cut with NsiI and PciI, using the HiFi Assembly method (NEB).

Lentivirus Production

Lentiviral particles were generated through triple transfection of HEK293T cells with the relevant lentiviral transfer vector plus the pCMVΔR8.91 and pMD.G packaging plasmids using TransIT-293 transfection reagent (Mirus) according to the manufacturer's instructions. Two days following transfection, viral supernatant was harvested and cleared of cell debris using a 0.45 μm filter. Target cells were transduced by spinoculation at 800 x *g* for 1 hour. For the generation of TRACE reporter libraries, GFP^+ cells were enriched using FACS. For expression of LEDGF chimeras, transduced KBM-7 cells were selected with 0.75 μg/ml puromycin (Thermo Fisher).

Flow Cytometry

For flow cytometry analysis, cells were washed once in PBS, fixed in 1% paraformaldehyde and analyzed on an LSR II (BD), FACSMelody (BD) or CytoFLEX (Beckman Coulter) instruments. For cell sorting, cells were washed once in PBS, resuspended in PBS + 2% FCS and sorted on a FACSAria II (BD) or FACSMelody (BD) instrument. Data analysis was carried out using FlowJo.

Immunoblotting

Cells were lysed in 1% SDS supplemented with 25 units of benzonase (Merck Millipore #70746-4) at room temperature for 20 min. Lysates were heated to 70°C for 10 minutes following addition of Laemmli buffer (Bio-Rad), and separated by SDS-PAGE. Proteins were transferred onto a PVDF membrane using the Wet Tank Blotting System (Bio-Rad). Membranes were blocked in 5% milk in PBS plus 0.2% Tween-20 for a minimum of 30 min at room temperature, probed overnight with the indicated antibodies at 4°C, washed 3-5 times in PBS plus 0.2% Tween-20 at room temperature, and incubated with the appropriate HRP-conjugated secondary antibodies (Jackson ImmunoResearch) for 40 min at room temperature. Reactive bands were visualized using Pierce ECL, SuperSignal West Pico or SuperSignal West Dura substrates (Thermo Fisher Scientific).

Purification of Cas9

E. coli BL21 Star (DE3) pRARE competent cells (Life Technologies) were transformed with pET-NLS-Cas9-6xHis, a gift from Prof. David Liu (Addgene plasmid #62934). Transformed cells were inoculated into 2YT broth containing 100 μg/ml carbenicillin and 30 μg/ml chloramphenicol and grown at 37°C overnight. Cells were then diluted 100-fold and grown further at 37°C until they reached the mid-log stage. Cas9 protein expression was induced by the addition of 1mM isopropyl β-D-1- thiogalactopyranoside (IPTG) followed by growth at 18°C overnight. Cells were collected by centrifugation at 5000 x *g* and resuspended in 50 mM Tris pH 7.5, 150 mM NaCl. Lysozyme (1 mg/ml) was added and incubated on ice for 30 minutes. Sodium chloride was added to a final concentration of 500 mM and cells were sonicated for 2.5 minutes. The lysate was centrifuged at 48,000 x *g* for 15 minutes. The supernatant was passed through a 0.45 micron syringe filter and loaded onto a 5 mL HisTrap Fast Flow column (Cytiva). The column was washed with 500 mM NaCl, followed by 150 mM NaCl in 50 mM Tris pH 7.5, 10% glycerol, 10 mM imidazole. Cas9 protein was eluted with a linear gradient from 0-500 mM imidazole in 50 mM Tris pH 7.5, 10% glycerol, 150 mM NaCl. Peak fractions were combined and passed over a HiTrap Q column (Cytiva) to remove contaminating nucleic acid. Column flowthrough was dialyzed overnight into 20 mM HEPES pH 7.8, 200 mM KCl, 5% glycerol, 1mM DTT. Dialyzed Cas9 protein was centrifuged at 48,000 x *g* for 15 minutes. The supernatant was concentrated to 5 mg/ml, syringe-filtered, and snap-frozen in liquid nitrogen.

Purification of Moloney Murine Leukemia Virus (MMLV) Reverse Transcriptase

Plasmid pCR28-MLV RT-M5 - encoding a Histidine-tagged, thermostable, RNaseH-deficient MMLV Reverse Transcriptase - was a kind gift from Dr. Michael Marr (Brandeis University, Boston, USA). BL21 Star (DE3) pRARE competent cells were transformed with pCR28-MLV RT-M5 and inoculated into 2YT broth containing 30 μg/ml kanamycin and 30 μg/ml chloramphenicol. Expression and purification of RT-M5 was performed as described above. Following HiTrap Q column, RT-M5 was dialyzed overnight into 40 mM Tris pH 7.5, 200 mM NaCl, 0.2 mM EDTA, 2mM DTT, 0.02% IGEPAL. Dialyzed protein was mixed with glycerol in a 1:1 ratio (final storage conditions were 20 mM Tris pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 0.01% IGEPAL, 50% glycerol) and stored at −20°C.

CRISPR/Cas9-mediated genome editing

To generate SUZ12 knockout clones, sgRNA-encoding oligonucleotides (IDT) (see Table S1 for sequences) were phosphorylated with T4 PNK (NEB), annealed by heating to 95°C followed by slow cooling to room temperature, and then cloned into the lentiviral sgRNA expression vector pKLV-U6gRNA(BbsI)-PGKpuro2ABFP (Addgene #50946, kindly deposited by Dr. Kosuke Yusa (Koike-Yusa et al., 2014)). The GFP reporter clone at chr9:126,112,867 was transduced with a lentiviral vector expressing Cas9 (Tchasovnikarova et al., 2017) and transduced cells were selected with Blasticidin (Thermo Fisher). Cas9-expressing cells were then transduced with the lentiviral sgSUZ12 expression vectors and transduced cells were selected with Puromycin (Thermo Fisher). Six days later, GFP^+ cells were single-cell cloned on a FACSAria II (BD) cell sorter and the resulting clones were validated by immunoblotting.

To generate the LEDGF knockout KBM-7 clone, an sgRNA targeting LEDGF was assembled from crRNA (see Table S1 for sequences) and tracrRNA (IDT) by heating to 95°C for 5 min and allowing to cool slowly to room temperature. The sgRNA was incubated with purified Cas9 protein for 20 min at room temperature to assemble RNP complexes, which were then introduced in KBM-7 cells using the SF Cell Line 4D-Nucleofector Kit (Lonza) and a 4D-Nucleofector instrument (Lonza) set at program EO-100. One week later, nucleofected cells were single-cell cloned and the resulting clones were validated by immunoblotting.

For knock-in experiments, two Cas9 RNP complexes were assembled as described above: one contained an sgRNA targeting the genomic site of interest, while the other contained a 'generic' sgRNA targeting a pair of unique sequences flanking the expression cassette within the donor plasmid (see Table S1 for sequences). Cells were then nucleofected with the two Cas9 RNP complexes plus 2 μg of donor plasmid using the SF Cell Line 4D-Nucleofector Kit (Lonza) and a 4D-Nucleofector instrument (Lonza) set at program EO-100. Cells were cultured for two weeks to allow for loss of any unintegrated donor plasmid, before GFP^+ cells were single-cell cloned on a FACSAria II (BD) or MoFlo Astrios (Beckman Coulter) cell sorter; successful integration in the resulting clones was confirmed by PCR.

ChIP-seq

Twenty million KBM-7 cells per ChIP were washed once in PBS, resuspended in 15 mL of complete IMDM, and then cross-linked in 1% formaldehyde (Thermo Fisher) for 10 min at room temperature. The reaction was quenched by adding glycine to a final concentration of 0.125 M for 5 min at room temperature. The crosslinked cells were washed twice in ice-cold PBS and lysed in 0.5% IGEPAL supplemented with 1 mM HEPES and 85 mM KCl. Following a 10 minute incubation on ice, nuclei were isolated by centrifugation at 800 x *g* for 5 minutes at 4°C, and then incubated in nuclear lysis solution (5 mM Tris, 10 mM EDTA, 1% SDS) for 10 min at 4°C. The samples were sheared using a Bioruptor (Diagenode; high power, 20 cycles of 30 s on, 30 s off) and insoluble material was removed by centrifugation at 16,000 x *g* for 10 min at 4°C. The supernatant, containing sheared chromatin, was pre-cleared with Pierce Protein G magnetic beads (Thermo Fisher Scientific) for a minimum of 2 h at 4°C. Chromatin carrying the desired histone modifications was immunoprecipitated overnight using 5 μg of primary antibody and Pierce Protein G magnetic beads (Thermo Fisher Scientific). On the following day, the magnetic bead-bound protein-DNA complexes were washed a total of five times and eluted in 150 mM $NaHCO_3$ plus 1% SDS. Cross-links were reversed at 67°C for a minimum of 4 h in the presence of 0.3 M NaCl and any remaining RNA was removed by RNaseA. Proteins were then digested in the presence of Proteinase K for a further 2 h at 45°C. The remaining DNA was isolated (QIAGEN PCR purification kit) and served as a template for the generation of Illumina sequencing libraries using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). Libraries were quantified using a TapeStation 2200 (Agilent), pooled, and submitted for 150 bp paired-end sequencing using an Illumina HiSeq 4000 instrument (GENEWIZ).

QUANTIFICATION AND STATISTICAL ANALYSIS

Histone modifications analysis

ChIP-seq reads were aligned to the human genome (GRCh38) using Bowtie 2, and reads with a MAPQ score > 20 were imported into SeqMonk for further analysis. Peaks were called using the implementation of the MACS peak caller in SeqMonk, with the exception of H3K9me3 and H3K27me3 for which peaks were called using the hiddenDomains tool (Starmer and Magnuson, 2016). To compute the percentage of reporters across different histone modifications, the bedtools intersect command was used to annotate reporter integration sites with overlapping peaks histone modifications. To calculate the distribution of reporters that would be expected were lentiviral integration to be entirely unbiased, the total nucleotides encompassed by each histone modification was divided by the total number of bases in the human genome and expressed as a percentage; due to the co-occurrence of marks at some regions, altogether these percentages sum to > 100.

Chromatin states analysis

Chromatin states were assigned from data generated by the epigenomics roadmap project (Kundaje et al., 2015) for K562 cells (E123, used for KBM-7 cells), primary monocytes from peripheral blood (E029, used for THP-1 cells), HeLa S3 cells (E117, used for HeLa cells) and fetal brain (E082, used for hTERT RPE-1 cells). The core 15 state model was downloaded from the epigenomics roadmap project and merged into 8 states to simplify the analysis and to remove states that encompassed only a small fraction of the genome. The 15 states were combined into 8 states as follows:

KBM-7 State	K562 State	K562 State Description
TSS	TssA + TssAFlnk + TssBiv	Active TSS Flanking Active TSS Bivalent/Poised TSS
Transcription Flank	TxFlnk + BivFlnk	Transcr. at gene 5′ and 3′ Flanking Bivalent TSS/Enh
Strong Transcription	Tx	Strong transcription
Weak Transcription	TxWk	Weak transcription
Enhancers	EnhG + Enh + EnhBiv	Genic enhancers Enhancers Bivalent Enhancer
ZNF genes/Repeats/Heterochromatin	ZNF/Rpts + Het	ZNF genes & repeats Heterochromatin
Polycomb	ReprPC + ReprPCWk	Repressed PolyComb Weak Repressed PolyComb
Quiescent	Quies	Quiescent/Low

To compute the percentage of reporters across different chromatin states, the bedtools intersect command was used to annotate reporter integration sites with overlapping chromatin states. To calculate the distribution of reporters that would be expected were lentiviral integration to be entirely unbiased, the total nucleotides encompassed by each chromatin state was divided by the total number of bases in the human genome.

ADDITIONAL RESOURCES

Detailed protocol

A detailed description of TRACE can be found in Methods S1.

Molecular Cell, Volume *82*

Supplemental information

TRACE generates fluorescent human reporter cell lines to characterize epigenetic pathways

Iva A. Tchasovnikarova, Sharon K. Marr, Manashree Damle, and Robert E. Kingston

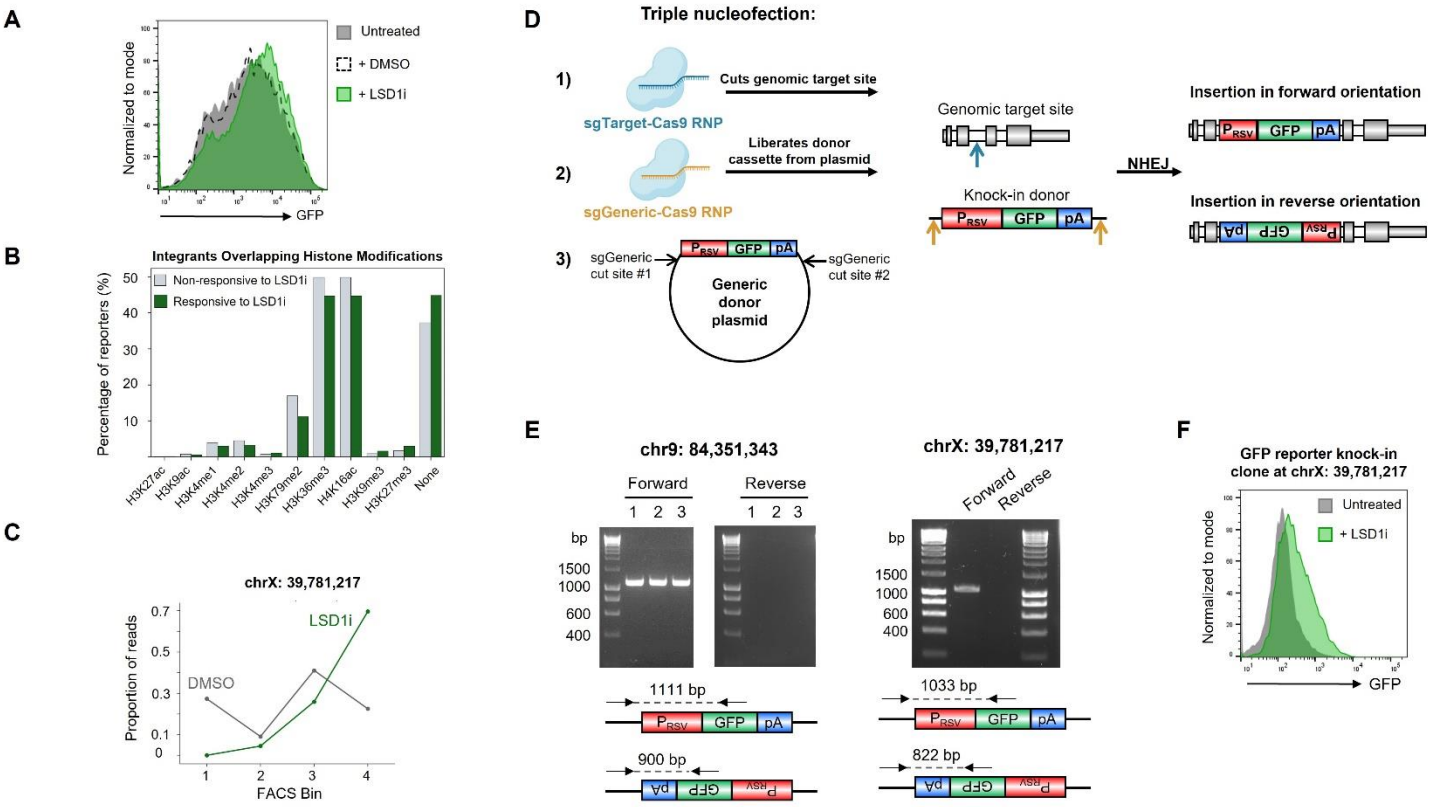


Figure S1. Generation of LSD1i-responsive reporter clones. (Related to Figure 3)

(A) Small molecule inhibition of LSD1 catalytic activity in the TRACE population. KBM-7 cells were treated with 1 μM LSD1i for 7 days, at which point GFP levels were assessed by flow cytometry.

(B) Genomic sites harboring reporters responsive to pharmacological inhibition of LSD1i are not enriched in specific types of chromatin.

(C) Expression profile of a reporter (chrX: 39,781,217) derepressed upon treatment with LSD1i, as measured by FACS followed by Illumina sequencing.

(D) Overview of NHEJ-mediated knock-in approach. KBM-7 cells are nucleofected with three species: (1) RNPs comprised of Cas9 and sgRNA targeting intended genomic site, (2) RNPs comprised of Cas9 and a 'generic' sgRNA targeting flanking regions in the plasmid donor, and (3) a plasmid donor carrying P_{RSV}-GFP-pA flanked by sgGeneric recognition sites. Following nucleofection, the donor cassette is liberated from the plasmid and inserted into the intended genomic site through NHEJ-mediated repair.

(E) Validation of donor integration at the intended genomic sites**.** P_{RSV}-GFP-pA integration was validated through PCR from genomic DNA extracted from the four LSD1i-responsive knock-in clones (Figure 3F and Figure S1F). All four clones harbored the donor cassette in the forward orientation.

(F) Generation of an LSD1i-reponsive reporter clone. CRISPR/Cas9-mediated knock-in to chrX: 39,781,209 (8 bp away from the genomic site identified by TRACE) resulted in the generation of a single cell reporter clone responsive to LSD1i treatment.

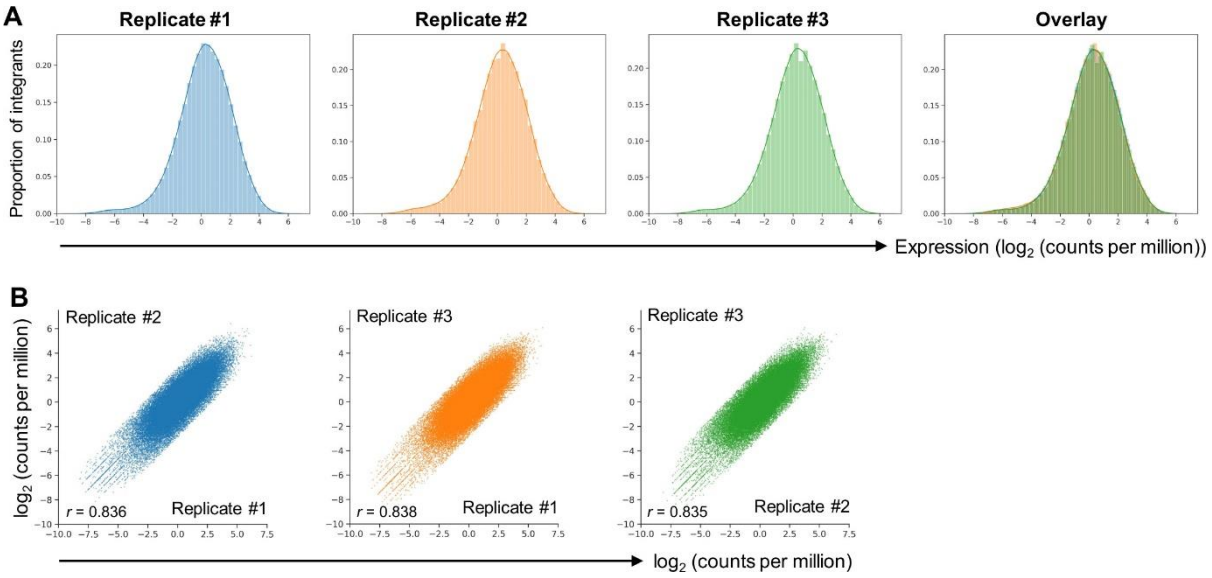


Figure S2. Transcript-based readout of reporter expression is consistent between replicate experiments. (Related to Figure 4)

(A) Distribution of integrant GFP expression across three replicates.

(B) Pairwise comparisons of integrant expression between replicates.

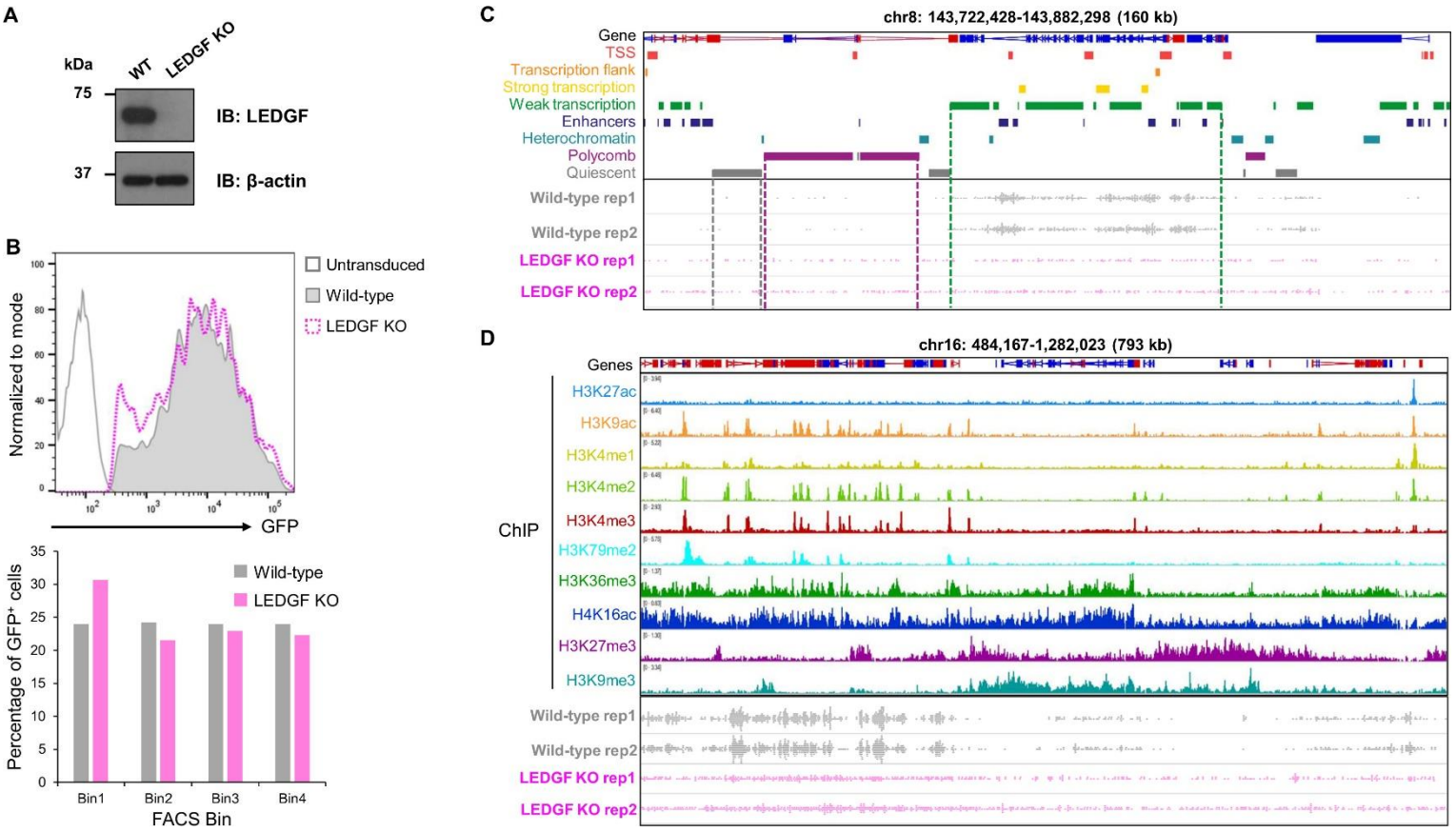


Figure S3. Redistribution of lentiviral integrations in LEDGF KO cells. (Related to Figure 5)

(A) Generation of a LEDGF KO KBM-7 clone. KBM-7 cells were nucleofected with RNPs containing Cas9 and an sgRNA targeting *LEDGF* and single-cell cloned 7 days later. Successful disruption of LEDGF was validated by immunoblot (IB).

(B) Loss of LEDGF results in a greater proportion of lentiviral integrations into repressive chromatin. LEDGF KO cells were transduced with the lentiviral TRACE library and GFP expression was measured by flow cytometry; a larger proportion of the GFP population was found in the lowest bin by FACS when compared to wild-type cells.

(C and D) Lentiviral integrations in LEDGF KO cells are more evenly distributed across different chromatin states (C) and histone modifications (D).

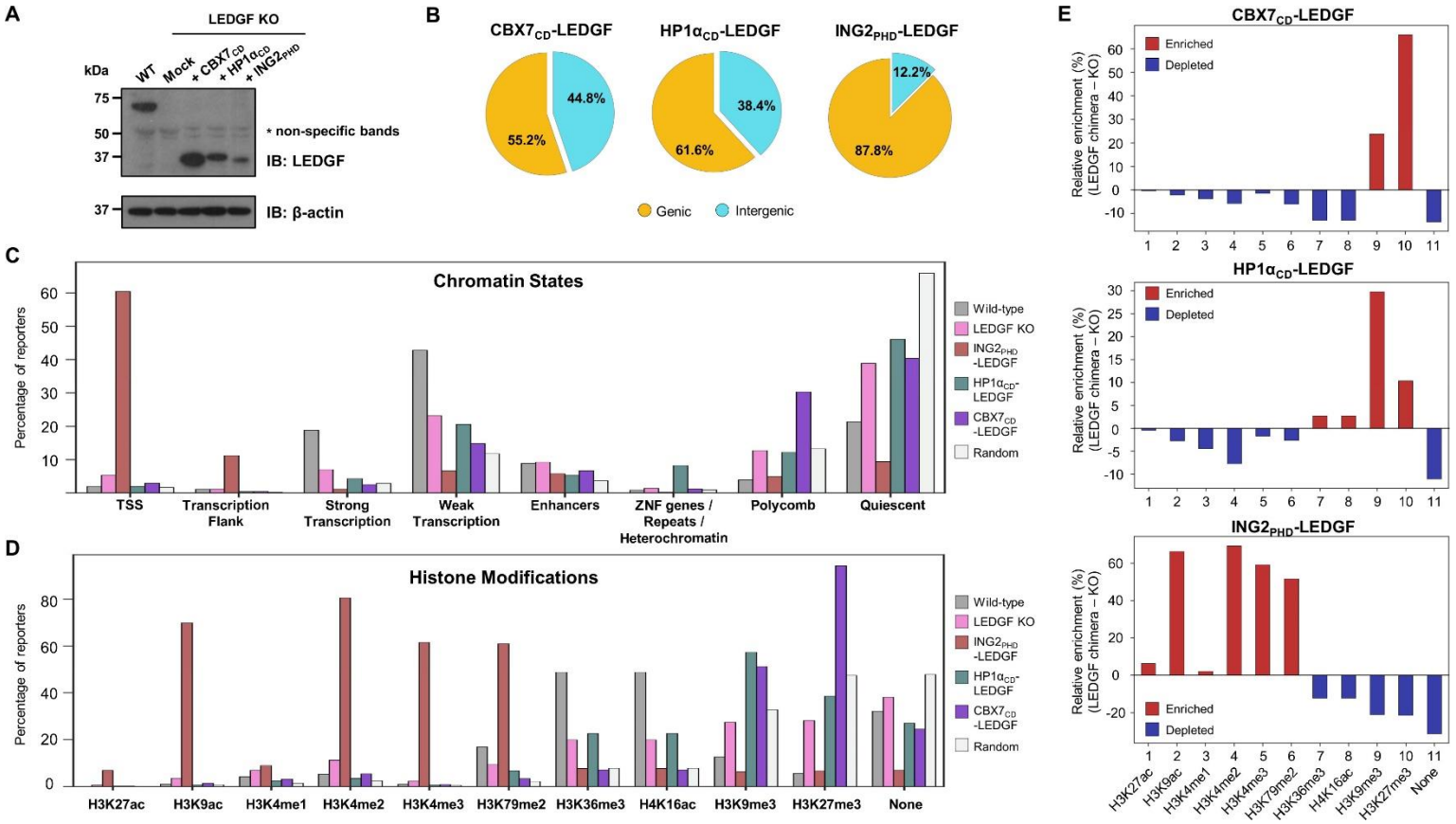


Figure S4. Redistribution of lentiviral integrations through expression of LEDGF chimeras. (Related to Figure 6)

(A) Expression of LEDGF chimeras in LEDGF KO KBM-7 cells. LEDGF KO cells were transduced with lentiviral constructs expressing three different chromatin reader domains ($CBX7_{CD}$, $HP1\alpha_{CD}$ and $ING2_{PHD}$) fused to the C-terminus of LEDGF. Following puromycin selection, expression was validated by immunoblot using an anti-LEDGF antibody that recognizes the C-terminus of LEDGF.

(B-D) Distribution of lentiviral integrants across genes and intergenic regions (B), chromatin states (C) and histone modifications (D).

(E) Enrichment of lentiviral integrants at genomic sites marked by LEDGF chimera-bound histone modifications.

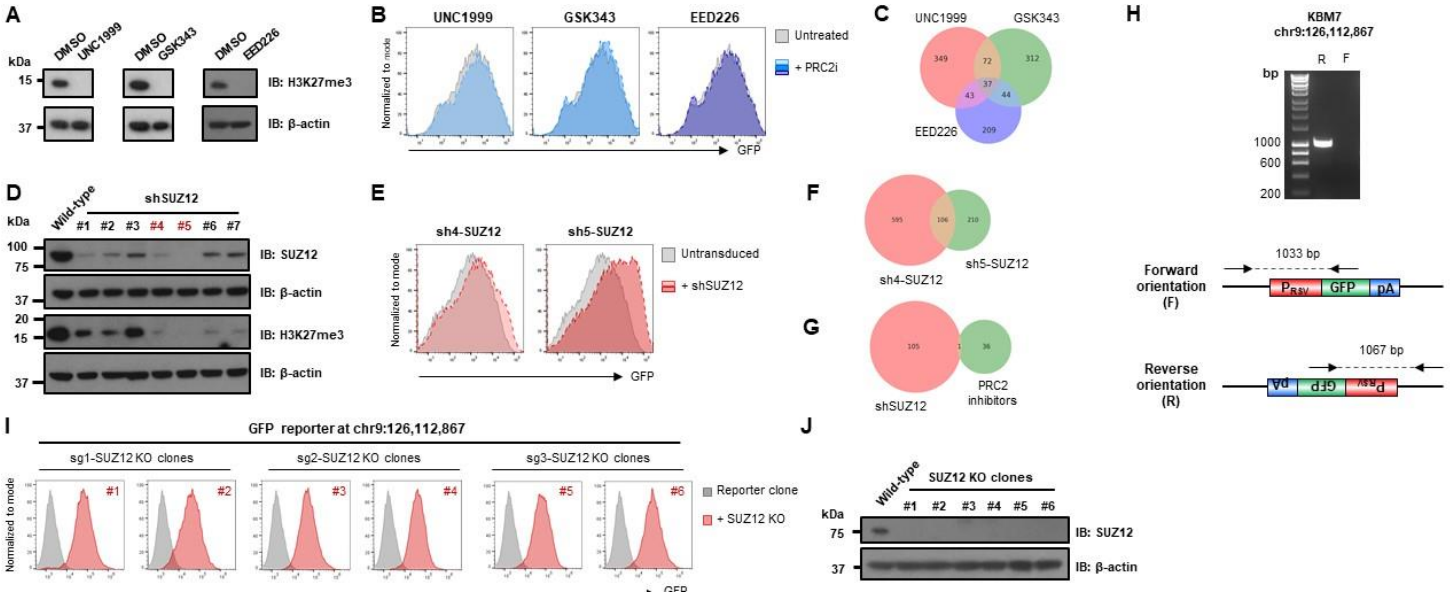


Figure S5. Loss of SUZ12 but not EZH2 enzymatic activity results in transgene derepression. (Related to Figure 7)

(A) Small molecule inhibition of PRC2 catalytic activity. KBM-7 cells were treated with the indicated inhibitors at 1 μM for 7 days, at which point H3K27me3 levels were assessed by immunoblot.

(B) Loss of H3K27me3 does not result in global derepression of TRACE reporters in $CBX7_{CD}$-LEDGF chimera cells.

(C) TRACE identified 37 reporters exhibiting concordant derepression upon treatment with UNC1999, GSK343 or EED226. We defined shared hits as the integrants upregulated by a GEM of at least 0.6 in each of the UNC1999, GSK343 and EED226 conditions when compared to DMSO, which were also unchanged (GEM < 0.25) in the DMSO condition when compared to untreated cells.

(D) Validation of shRNA-mediated depletion of SUZ12. KBM-7 cells were transduced with 7 different shRNAs targeting SUZ12. Following puromycin selection, SUZ12 and H3K27me3 levels were assessed by immunoblot. The two most efficacious shRNAs (sh4 and sh5) were selected for a TRACE experiment aimed at identifying SUZ12-responsive reporters.

(E) Loss of SUZ12 resulted in global derepression of TRACE reporters in $CBX7_{CD}$-LEDGF chimera cells.

(F) TRACE identified 106 reporters exhibiting concordant derepression upon depletion of SUZ12. We defined shared hits as the integrants upregulated by a GEM of at least 0.5 in sh4-SUZ12 condition and a GEM of at least 0.7 in the sh5-SUZ12 condition when compared to shControl, and unchanged (GEM <0.25) in the shControl condition when compared to untreated cells. The difference in GEM threshold for sh4-SUZ12 and sh5-SUZ12 reflects the difference in knockdown efficiency between the two shRNAs.

(G) There was little concordance between pharmacological and genetic perturbation of PRC2 function. Indeed, just a single reporter was derepressed by all three PRC2 small molecule inhibitors, sh4-SUZ12 and sh5-SUZ12.

(H) Validation of donor integration at the intended genomic site. P_{RSV}-GFP-pA integration was validated through PCR from genomic DNA extracted from the SUZ12-responsive knock-in clone (KI clone) (Figure 7D). The KI clone harbored the donor cassette in the reverse orientation.

(I-J) CRISPR/Cas9-mediated disruption of SUZ12 results in reporter derepression. (I) SUZ12 knockout (KO) in the reporter clone harboring a GFP expression cassette at chr9: 126,112,867 (Figure 7D) was achieved by sequential transduction of the reporter cells with Cas9 and one of three sgRNA-expressing lentiviral vectors. GFP expression in the resulting clones was assessed by flow cytometry. (J) Confirmation of loss of SUZ12 by immunoblot.

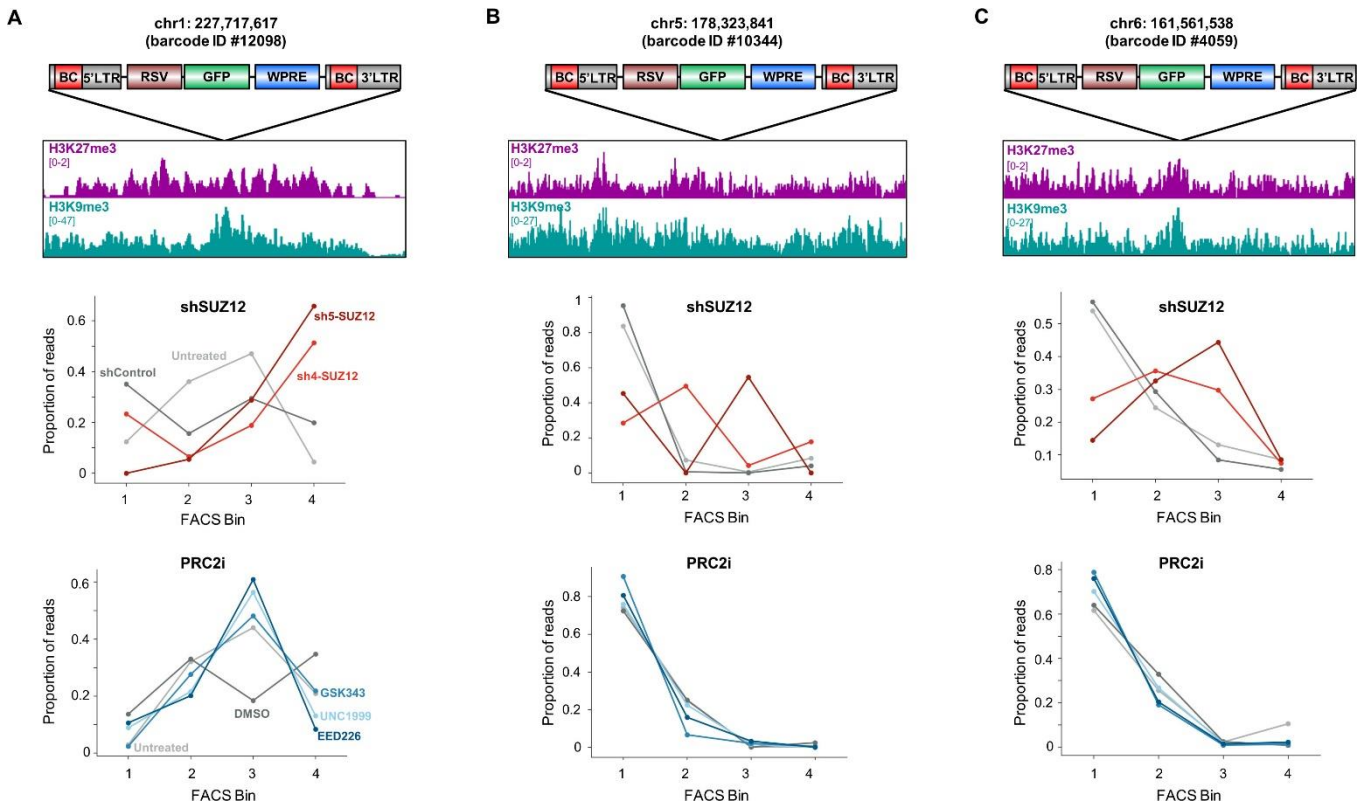


Figure S6. Reporters responsive to SUZ12 depletion reside in genomic loci marked by H3K27me3 and H3K9me3. (Related to Figure 7)

(A-C) IGV screenshots depicting H3K27me3 and H3K9me3 distribution in genomic regions harboring TRACE reporters responsive to loss of SUZ12. Reporter expression profiles are shown below.

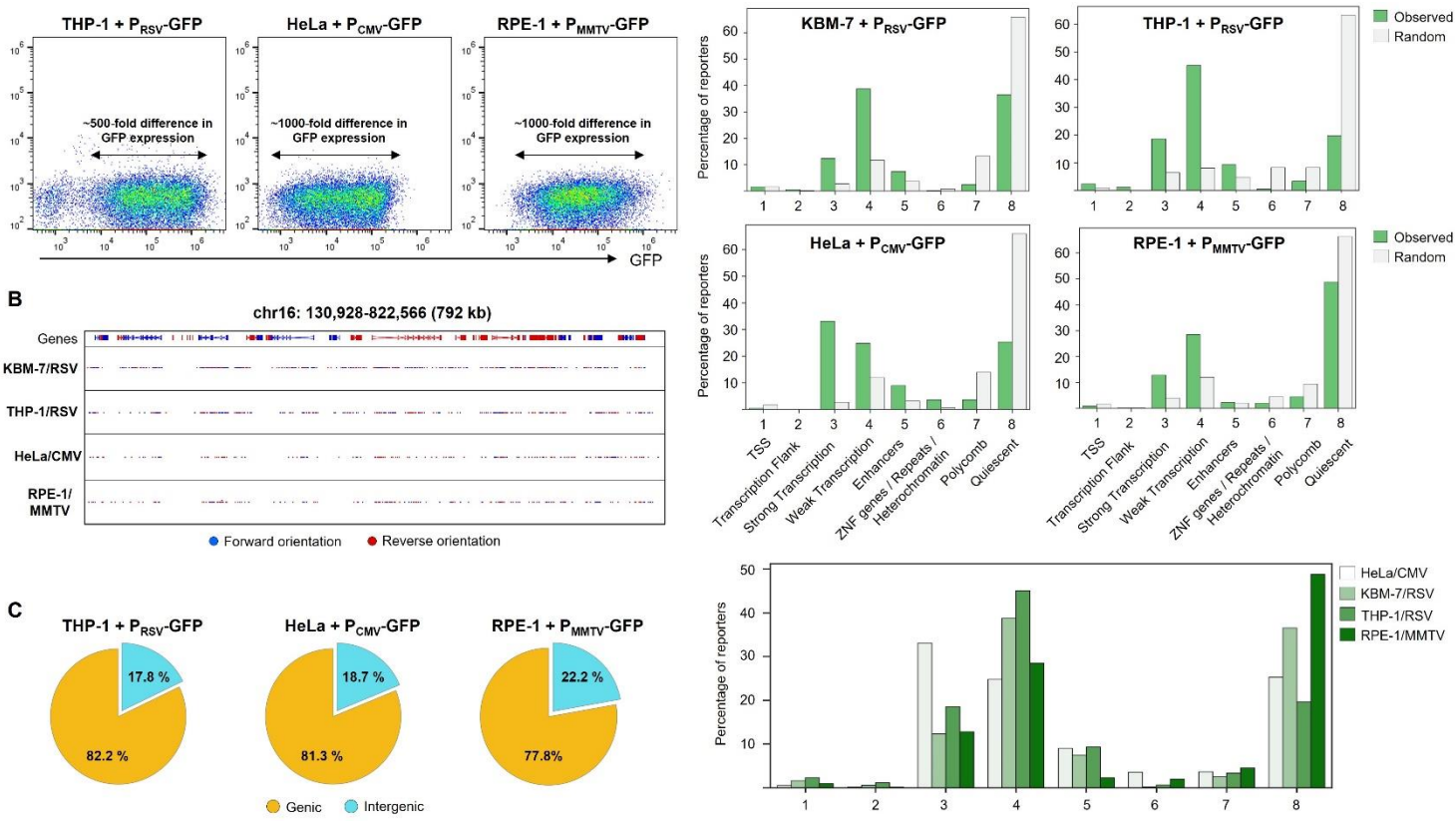


Figure S7. Generation of TRACE libraries using alternative promoters and cell lines.

(Related to Figure 2)

(A) Transduction of THP-1, HeLa and hTERT RPE-1 cells with lentiviral expression vectors driven by either the RSV, CMV or MMTV promoters results in a broad range of GFP expression.

(B-C) Lentiviral integrants in KBM-7, THP-1, HeLa and hTERT RPE-1 cells are similarly distributed across genes and intergenic regions (B), with the vast majority (~80%) of integrants mapping to genes (C).

(D) Lentiviral integrants in KBM-7, THP-1, HeLa and hTERT RPE-1 cells are similarly distributed across chromatin states. The barplots depict the distribution of reporter integration sites across eight chromatin states. 'Observed' refers to the distribution of mapped integration sites, while 'Random' refers to the distribution of integration sites that would be expected if lentiviral integration was entirely random. The barplot in the bottom panel depicts the observed distribution of reporter integration sites in the four cell lines across eight chromatin states.