Article
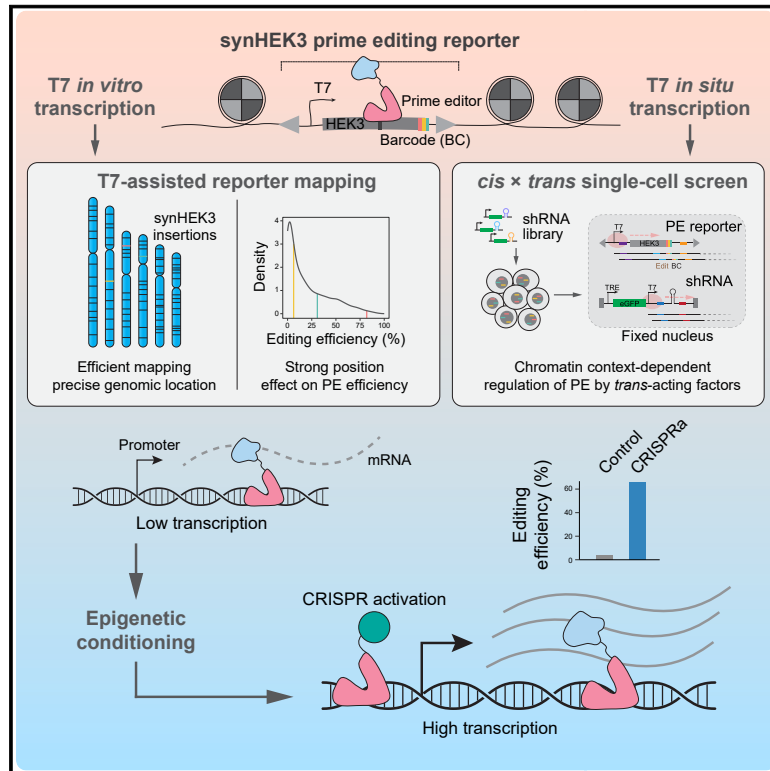
Chromatin context-dependent regulation and epigenetic manipulation of prime editing

Graphical abstract

Highlights

- T7-assisted reporter mapping assay determines precise locations of integrated constructs

- Active transcriptional elongation enhances prime editing

- Pooled shRNA screens reveal HLTF as a context-dependent suppressor of prime editing

- Epigenetic conditioning of a locus modulates prime editing efficiency

Authors

Xiaoyi Li, Wei Chen, Beth K. Martin, ..., Jenny F. Nathans, David S. Lee, Jay Shendure

Correspondence

xyli10@uw.edu (X.L.), shendure@uw.edu (J.S.)

In brief

Prime editing exhibits strong, chromatin-related position effects and is promoted by active transcriptional elongation. Leveraging these features via epigenetic conditioning of a locus modulates prime editing efficiency.

CellPress

Cell

Article

Chromatin context-dependent regulation and epigenetic manipulation of prime editing

Xiaoyi Li,[1,*] Wei Chen,[1,2] Beth K. Martin,[1] Diego Calderon,[1] Choli Lee,[1] Junhong Choi,[1,3] Florence M. Chardon,[1] Troy A. McDiarmid,[1] Riza M. Daza,[1] Haedong Kim,[1] Jean-Benoît Lalanne,[1] Jenny F. Nathans,[1,4] David S. Lee,[1,5] and Jay Shendure[1,3,6,7,8,9,*]

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[2]Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA 98195, USA
[3]Howard Hughes Medical Institute, Seattle, WA 98195, USA
[4]Medical Scientist Training Program, University of Washington, Seattle, WA 98195, USA
[5]Institute for Protein Design, University of Washington, Seattle, WA 98195, USA
[6]Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA
[7]Allen Discovery Center for Cell Lineage Tracing, Seattle, WA 98109, USA
[8]Seattle Hub for Synthetic Biology, Seattle, WA 98109, USA
[9]Lead contact
*Correspondence: xyli10@uw.edu (X.L.), shendure@uw.edu (J.S.)
https://doi.org/10.1016/j.cell.2024.03.020

SUMMARY

We set out to exhaustively characterize the impact of the *cis*-chromatin environment on prime editing, a precise genome engineering tool. Using a highly sensitive method for mapping the genomic locations of randomly integrated reporters, we discover massive position effects, exemplified by editing efficiencies ranging from ∼0% to 94% for an identical target site and edit. Position effects on prime editing efficiency are well predicted by chromatin marks, e.g., positively by H3K79me2 and negatively by H3K9me3. Next, we developed a multiplex perturbational framework to assess the interaction of *trans*-acting factors with the *cis*-chromatin environment on editing outcomes. Applying this framework to DNA repair factors, we identify HLTF as a context-dependent repressor of prime editing. Finally, several lines of evidence suggest that active transcriptional elongation enhances prime editing. Consistent with this, we show we can robustly decrease or increase the efficiency of prime editing by preceding it with CRISPR-mediated silencing or activation, respectively.

INTRODUCTION

Prime editing facilitates the precise installation of diverse genetic variants with minimal off-target effects.[1–5] The prime editor consists of a fusion of a Cas9 nickase and reverse transcriptase (RTase), while the prime editing guide RNA (pegRNA) specifies both the target site and desired edit. The high programmability and specificity of prime editing, together with its avoidance of double-stranded breaks (DSBs), offer considerable advantages over alternatives in the context of therapeutic genome editing,[6,7] molecular recording,[8–10] and the functional characterization of genetic variants.[11]

A counterpoint to this promise is that prime editing's efficiency is highly variable and often low.[1,12–14] Relevant factors likely include (1) properties of the prime editing ribonucleoprotein (RNP) complex; (2) the target sequence and edit type; (3) *trans*-acting factors, e.g., endogenous DNA repair proteins; and (4) the *cis*-chromatin environment of the target site. The first three factors have been explored to great effect, resulting in optimized and/or specialized prime editors,[1,13,15,16] sequence-based machine learning models

that facilitate pegRNA design,[12,17–20] and mismatch repair (MMR)-targeting strategies that enhance prime editing efficiency,[1,13,14] respectively. However, how the *cis*-chromatin environment impacts prime editing remains largely unexplored.

It is well established that conventional Cas9-mediated genome editing is strongly influenced by chromatin, both through the steric effects of nucleosomes[21,22] and epigenetic effects on the balance of endogenous DNA repair pathways used to repair Cas9-mediated DSBs.[23–26] However, as prime editing leverages a different set of endogenous DNA repair pathways than conventional Cas9-mediated genome editing, we hypothesized that prime editing's outcome would also be strongly influenced by the chromatin.

RESULTS

Design and efficient mapping of prime editing reporters

To measure *cis*-chromatin effects, we developed a strategy for efficiently mapping the precise genomic locations of randomly integrated reporters (Figure 1A). The method relies on linear
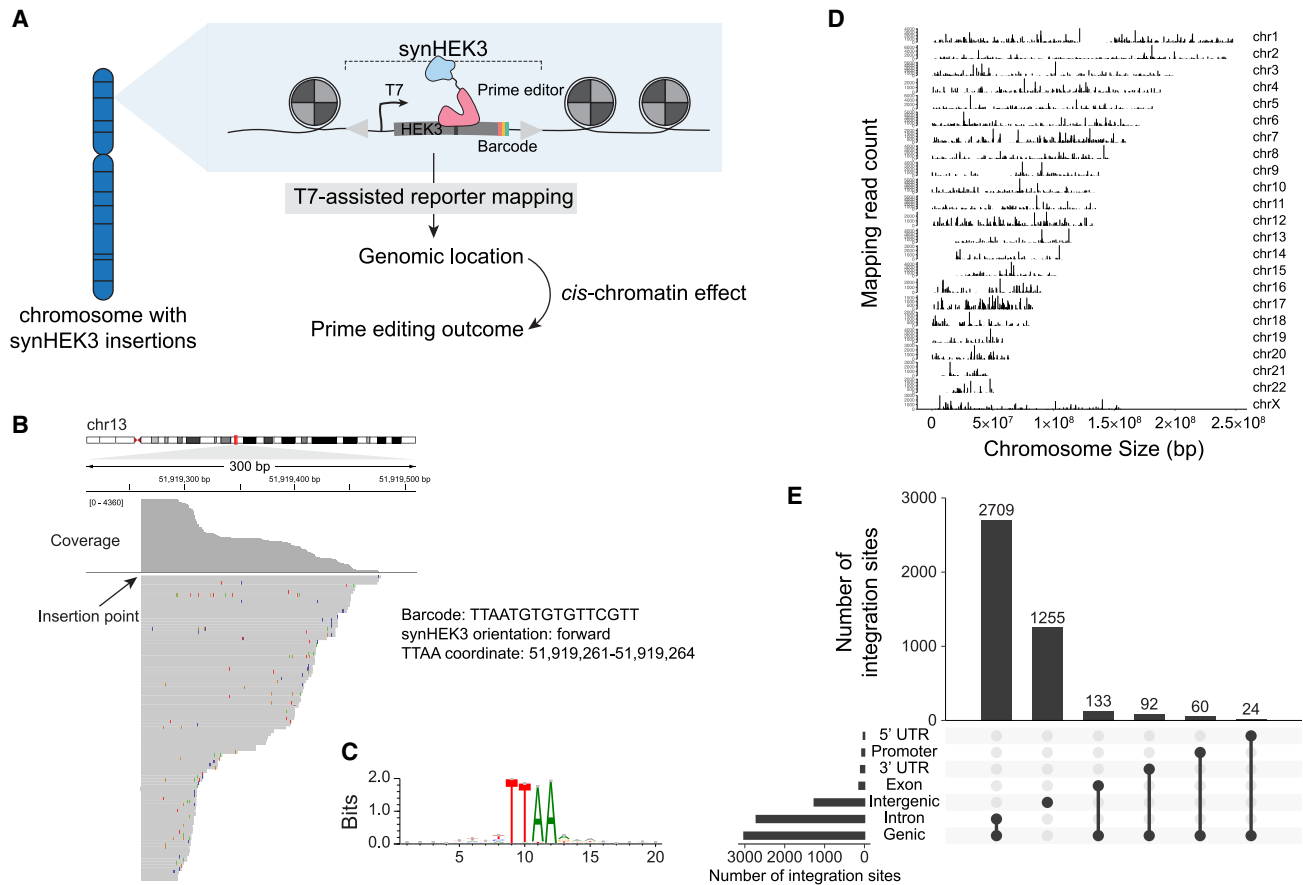
Figure 1. Efficient genome-wide mapping of integration sites of synHEK3 reporters

(A) Measuring *cis*-chromatin effects on prime editing efficiency. A library of synHEK3 reporters is randomly inserted throughout the genome. Genomic locations of individual reporters are determined with a T7-assisted reporter mapping method. The *cis*-chromatin contexts of mapped reporters are used to model prime editing outcomes as measured from each reporter.

(B) Genome browser view of a read pileup pinpointing the precise coordinates of a synHEK3 reporter integration. Barcode sequence, orientation, and coordinates of the reporter are annotated.

(C) Motif enrichment analysis of 20-bp windows surrounding synHEK3 integration sites.

(D) Coverage plot of unique synHEK3 reporters identified in the bottlenecked pool (*n* = 4,273). Vertical bar lengths correspond to read counts.

(E) UpSet plot of genomic annotations of synHEK3 integration sites.

See also Figure S1.

amplification of insertion junctions by *in vitro* transcription (IVT) on genomic DNA (gDNA) from a bacteriophage T7 promoter embedded within the reporter. *In vitro* transcribed RNAs are then reverse transcribed (RT), amplified by semi-specific PCR, and sequenced to associate reporter-specific barcodes (BCs) with neighboring genomic sequences. Sequencing reads span insertion junctions, such that integration sites are mapped at base-pair (bp) resolution (Figure S1A). T7-assisted reporter mapping offers two key advantages over conventional reporter mapping approaches.[27–30] First, it eliminates digestion and ligation steps that are characterized by bias and/or suboptimal efficiency. Second, T7 polymerase linearly amplifies molecules carrying positional information before RT-PCR, which further increases sensitivity.

We developed a compact *piggyBac* (PB)-based prime editing reporter bearing the T7 promoter, the target sequence for a highly efficient pegRNA (HEK3), and a 16-bp degenerate BC. The final construct was 358 bp and lacked any known regulatory elements that could potentially interfere with the local chromatin environment (Figure S1A). We randomly integrated a complex library of these "synHEK3" reporters into a K562 cell line constitutively expressing the PE2 prime editor[2] (Figure S1B). After PB transposase was diluted out and all synHEK3 integrations were stable, we estimated 15.5 reporter copies (mean) per cell by qPCR (Figure S1C). We bottlenecked these cells to ∼500 clones, each presumably containing a unique set of integrated reporters. Downstream work was performed on this bottlenecked population (Figure S1B).

To map the genomic locations of synHEK3 integrations, we performed T7-assisted reporter mapping. Subsets of aligned reads defined sharp boundaries when visualized on a genome browser, each corresponding to the precise insertion point of a
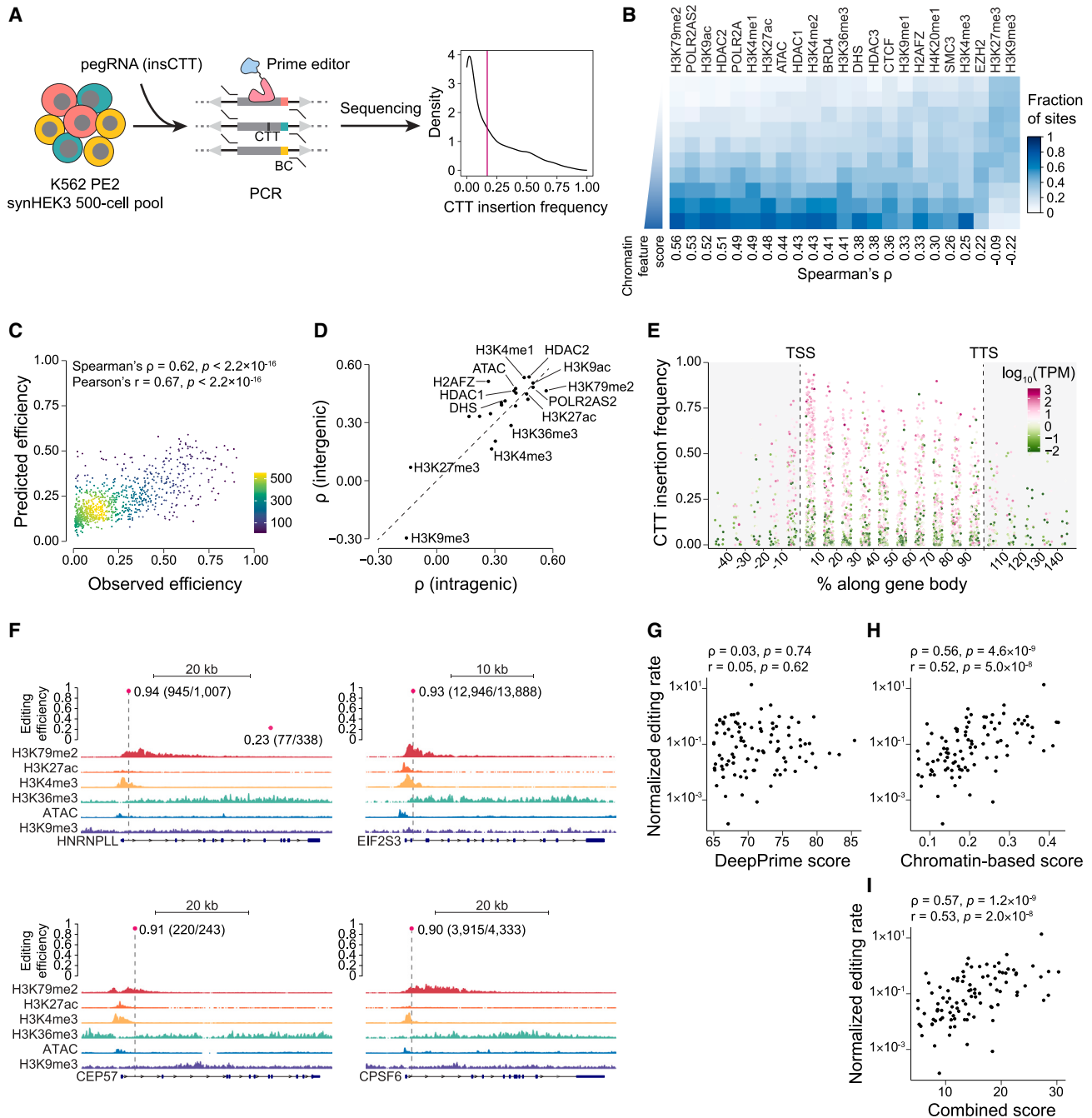
Figure 2. Chromatin context has a major impact on prime editing efficiency

(A) Left: workflow of the experiment. Right: density plot of CTT insertion frequency in all uniquely barcoded synHEK3 reporters ($n = 4,273$). Red line indicates prime editing efficiency (17%) at the endogenous HEK3 locus in K562 cells.

(B) Heatmap of fractions of highly editable (>25%) sites in synHEK3 sites stratified by chromatin feature scores. SynHEK3 reporters are binned into 10 equally sized bins with increasing chromatin feature scores. The chromatin features are ordered left-to-right by their correlation coefficient (Spearman's ρ) with prime editing efficiency.

(C) Scatter plot of observed vs. predicted prime editing efficiencies using reporters in a holdout test set. Points colored by the number of neighboring points. p values determined by the Spearman correlation test or by the Pearson correlation test.

(D) Scatter plot of Spearman's ρ between chromatin feature scores and prime editing efficiencies, calculated separately for intragenic and intergenic reporters.

synHEK3 reporter (Figure 1B). After BC error correction, we identified 10,095 insertion sites (Figure S1D). Saturation analysis indicated that the mapping was near complete (Figure S1E). Motif analysis of insertion junctions revealed the expected TTAA motif for PB transposition[31] (Figure 1C). We removed sites lacking this motif (6.4%) and assigned the genomic coordinates of TTAA motifs as locations of individual reporters. Of note, only 4,273 mapped sites (42.3%) bore a unique BC (Figure 1D), while the remaining 5,177 (51.3%) bore one of 880 BCs that recurred at multiple integration sites (Figure S1F). These recurrent BCs are not due to limited BC complexity in the plasmid library but rather reflect PB transposon excision and re-integration events occurring after DNA replication but before the transposase was fully diluted out.[32]

SynHEK3 reporters were distributed across all major genomic annotations, with a bias toward genic regions (Figures 1E, S1G, and S1H). Compared with randomly selected genomic or TTAA sites, PB integrations were most strongly enriched near active transcriptional start sites (TSSs) and enhancers and most strongly depleted from quiescent regions, which are mostly constitutive heterochromatin[33] (Figure S1I). Given these biases, we sought to assess how broad a range of epigenetic environments was sampled by integrated reporters. For this, we computed epigenetic scores for 2-kb windows surrounding synHEK3 integrations for various chromatin features in K562 cells[34] (Tables S2 and S3) and compared these against equivalent scores for randomly selected genomic and TTAA sites. While synHEK3 integration sites tended to have higher scores for markers of open chromatin due to PB integration bias, they nonetheless sample the full dynamic ranges of the chromatin features surveyed, including those associated with poorly accessible chromatin environments (Figure S1J).

Impact of chromatin features on prime editing efficiency
To measure prime editing outcomes across different chromatin contexts, we transfected the bottlenecked cell population with a pegRNA designed to install a CTT insertion at the HEK3 target sequence.[1] After 4 days, we measured prime editing efficiencies for all uniquely barcoded reporters (Figure 2A, left). This pegRNA drove 17% editing at the endogenous *HEK3* locus. However, a broad range of efficiencies (~0%–94%, mean: 19.8%) was observed for the HEK3 target when integrated at different genomic locations via the synHEK3 reporter, demonstrating prime editing outcomes are highly influenced by position effects (Figure 2A, right).

We examined correlations between chromatin feature scores[34] and prime editing efficiencies across synHEK3 genomic locations. The chromatin feature scores of H3K79me2, POLR2AS2, H3K9ac, and HDAC2 were most positively correlated, while H3K9me3 was most negatively correlated (Figures S2A and S2B). When we binned synHEK3 reporters based on each chromatin feature score, the same positive features were strongly associated with the proportion of sites that were highly edited (>25%; Figure 2B).

Among all features examined, H3K79me2 levels were most strongly correlated with prime editing efficiency (Spearman's $\rho = 0.56$; Figures S2A and S2B). The lowest and highest deciles of H3K79me2 scores exhibited median efficiencies of 1.7% and 54.5%, respectively, a 32-fold difference (Figure S2B). Upon training a beta regression model based on the chromatin feature scores, H3K79me2 ($p = 1.47 \times 10^{-12}$) and H3K9me3 ($p = 1.15 \times 10^{-14}$) were the best positive and negative predictors of efficiency, respectively (Table S2). This model accurately predicted prime editing efficiencies on a holdout test set (Spearman's $\rho = 0.62$, Pearson's r = 0.67, $p < 2.2 \times 10^{-16}$ for both; Figure 2C).

Surprisingly, a small fraction (11%) of H3K79me2-high sites (top 10%) had near-zero editing efficiencies (Figure S2A). We surmised that some of these underedited outliers might be due to a technical artifact, wherein some clones lose expression of PE2 due to excision during synHEK3 integration, as both were integrated via PB. To check this, we engineered two new pools of synHEK3 reporters in wild-type (WT) K562 cells, mapped reporter locations, and measured editing efficiencies with transiently expressed PE2 (Table S3). Consistent with our hypothesis, a smaller fraction of H3K79me2-high sites had near-zero editing efficiencies (4.7% vs. 11%). Importantly, chromatin features predicting high prime editing efficiencies were similar, and the beta regression model trained on the original data accurately predicted the editing efficiencies of these two independently generated sets of synHEK3 reporters (Spearman's $\rho = 0.71$, Pearson's r = 0.70, $p < 2.2 \times 10^{-16}$ for both; Figures S2C and S2D).

In mammalian cells, H3K79 methylation is solely deposited by DOT1L in a process coupled to RNA polymerase II (RNA Pol II) transcriptional elongation.[35–37] As H3K79me2 is overwhelmingly intragenic, a global analysis might mask differences in the interaction of chromatin and prime editing in intragenic vs. intergenic regions. We therefore repeated the correlation analysis and beta regression modeling on intragenic and intergenic synHEK3 reporters separately (Figure 2D; Table S2). These analyses showed that H3K79me2's value as a positive predictor is restricted to

(E) Prime editing efficiency for gene-proximal reporters. Distance was scaled by gene length and binned. Negative values refer to synHEK3 sites located upstream of TSS. Values >100% refer to synHEK3 sites located downstream of transcription termination site (TTS). Points colored based on the expression levels ($\log_{10}$) of the genes. TPM, transcripts per million.

(F) Genome browser views of the 4 most highly editable sites. Sites of integration and measured editing efficiencies are shown as a dot plot at top and aligned with selected epigenetic tracks. For each synHEK3 insertion, editing efficiency, number of reads with edit (numerator), and the total number of reads (denominator) are annotated. The dashed vertical lines mark locations of the insertions.

(G) Scatter plot of sequence-based prediction (DeepPrime, x axis) vs. normalized editing rate (y axis, $\log_{10}$ scale) for epegRNAs designed for prime editing at endogenous genomic sites. p values determined by the Spearman correlation test or by the Pearson correlation test.

(H) Scatter plot of chromatin-based prediction (our model, x axis) vs. normalized editing rate (y axis, $\log_{10}$ scale). p values determined by the Spearman correlation test or by the Pearson correlation test.

(I) Scatter plot of combined score (x axis) vs. normalized editing rate (y axis, $\log_{10}$ scale). p values determined by the Spearman correlation test or by the Pearson correlation test.

See also Figure S2.

intragenic regions of the genome, while H3K9me3 is a strong negative predictor of prime editing efficiency in both intragenic and intergenic regions.

The chromatin immunoprecipitation (ChIP)-seq signal of H3K79me2 is more spread out and downstream-biased than that of H3K4me3, a marker of promoters (Figure S2E). We hypothesized that prime editing is most efficient in transcriptionally active regions and particularly so immediately downstream of active TSS. To evaluate this, we stratified synHEK3 reporters by distance from the nearest TSS as well as the mRNA expression level of the overlapping or nearest gene. Prime editing efficiencies were indeed correlated with both expression levels and TSS proximity (Figures 2E, S2F, and S2G). In addition, we identified a weak correlation between prime editing efficiency and transcript orientation, as synHEK3 sites on the same strand as the coding strand of transcription exhibited slightly higher editing efficiencies ($p = 0.05$; Figure S2H).

To visualize the contrasting epigenetic environments, we prepared genome browser views of the top 4 highly editable synHEK3 sites (90%–94% PE) and 4 poorly editable synHEK3 sites (~1% PE) (Figures 2F and S2I). The top sites are intragenic in highly expressed genes and within 3.5 kb of the TSS. The poorly editable examples are also intragenic but within unexpressed genes and in some cases hundreds of kilobases from the TSS. Similar visualizations of the epigenetic environments of all synHEK3 sites from this experiment, together with their efficiencies, can be browsed at https://genome.ucsc.edu/s/lxyttpp912/hg38_synHEK3_hub.

Impact of chromatin environment on prime editing in diverse endogenous target sites

To ask whether learnings based on synHEK3 reporters generalize, we designed epegRNAs with DeepPrime[19] to install 3-bp CCT insertions at 121 endogenous genomic target sites and tested these in K562 cells (predicted efficiencies 65%–85%; Tables S1 and S4). Interestingly, the resulting editing efficiencies were uncorrelated with DeepPrime scores, probably because they were all designed for high efficiency (Spearman's $\rho = 0.03$, $p = 0.74$, Pearson's r = 0.05, $p = 0.62$; Figure 2G). However, despite variation in guide and target sequences, editing efficiencies were well predicted by the synHEK3-derived beta regression model (Spearman's $\rho = 0.56$, $p = 4.6 \times 10^{-9}$, Pearson's r = 0.52, $p = 5.0 \times 10^{-8}$; Figures 2H and S2J). Combining the DeepPrime score with our model did not further increase prediction accuracy (Spearman's $\rho = 0.57$, $p = 1.2 \times 10^{-9}$, Pearson's r = 0.53, $p = 2.0 \times 10^{-8}$; Figure 2I). We conclude that the epigenetic factors shaping prime editing efficiencies at synHEK3 reporters also apply to endogenous target sites. Furthermore, a prediction model based on chromatin features can be used to rank the outputs of sequence-based pegRNA design tools.[12,18,19]

Differential position effects on Cas9 vs. prime editing

To the extent that the wide range of prime editing efficiencies at integrated synHEK3 reporters is due to differential accessibility for the prime editor, we would expect Cas9 editing to be similarly affected. We therefore transfected the K562 synHEK3 reporter pool with Cas9 RNP bearing a gRNA with the same HEK3-target-ing spacer and compared editing outcomes for prime vs. Cas9 editing across an identical set of genomic locations. Cas9 editing was more efficient than prime editing. Frequent indels were observed as early as day 1. 75% of reporters had indel frequencies >90% by day 4, and Cas9 successfully edited synHEK3 reporters that were resistant to prime editing (Figures 3A and S3A). As expected, synHEK3 reporters near highly expressed genes were more frequently edited by Cas9. Reporters immediately downstream of TSSs of highly expressed genes had higher indel frequencies at day 1, but the differences were negligible at later time points (Figure S3B).

To explore differences, we clustered synHEK3 reporters into 6 groups based on measured efficiencies of prime and Cas9 editing (Figure 3B). The reporters in each group naturally clustered in principal-component analysis (PCA) plots of chromatin feature scores, suggesting that the epigenetic environment shapes the editing efficiencies exhibited by members of each group (Figures S3C and S3D). We further clustered the larger, more highly edited groups (groups 3–6) into subclusters, resulting in 14 clusters overall (Figures 3B and S3E).

While prime and Cas9 editing efficiencies were generally correlated, there were a few pairs of groups whose properties were informative beyond that correlation: (1) those highly edited by both prime and Cas9 editing (groups 6.1 and 6.2), (2) those highly edited by Cas9 but poorly by prime editing (groups 4.2 and 4.3), and (3) those poorly edited by both prime and Cas9 editing (groups 1.0 and 2.0) (Figure 3C).

Groups 6.1 and 6.2 exhibited similarly high Cas9 editing (mean: 96.1% vs. 96.8%, $p = 0.1$), but group 6.2 exhibited significantly higher prime editing (mean: 69.3% vs. 82.5%, $p = 6.6 \times 10^{-16}$). Both group 6.1 and group 6.2 are overwhelmingly intragenic (Figure 3D) and similarly accessible ($p = 0.63$; Figure 3E). However, group 6.2 sites are more highly expressed than group 6.1 sites, suggesting that higher levels of transcription specifically promote prime editing ($p = 2.8 \times 10^{-3}$; Figure 3F).

Groups 4.2 and 4.3 exhibit high levels of Cas9 RNP editing (mean: 86.4% and 89.9%) but low levels of prime editing (mean: 17.1% and 2.2%). Many synHEK3 sites in group 4.3 exhibited near-zero efficiency for prime editing, which we attributed to the technical artifact of PB excision of PE2 construct, discussed above. However, group 4.2 sites tended to be intragenic and accessible (Figures 3D and 3E) but lowly expressed ($p = 8.3 \times 10^{-14}$ for group 4.2 vs. 6; Figure 3F) and far from any TSS ($p = 4.1 \times 10^{-11}$ for group 4.2 vs. 6; Figure 3G). The fact that group 4.2 sites are amenable to Cas9 editing but not prime editing suggests that both Cas9 and prime editing benefit from chromatin accessibility, but more active transcription specifically promotes prime editing.

The most poorly editable sites fell into groups 1.0 and 2.0 (Figure 3C). Both exhibited very-low prime editing efficiencies (mean: 0.8% vs. 2.2%, $p = 9.3 \times 10^{-9}$), but group 1.0 had substantially lower Cas9 editing (mean: 25.8% vs. 38.3%, $p < 2.2 \times 10^{-16}$). We suspect group 1.0 corresponds to constitutive heterochromatin marked by higher H3K9me3 that remains silenced throughout the cell cycle and development.[38] In contrast, group 2.0 is marked by higher H3K27me3 and might correspond to facultative heterochromatin, which is silenced upon differentiation[39,40] (Figure S3F).
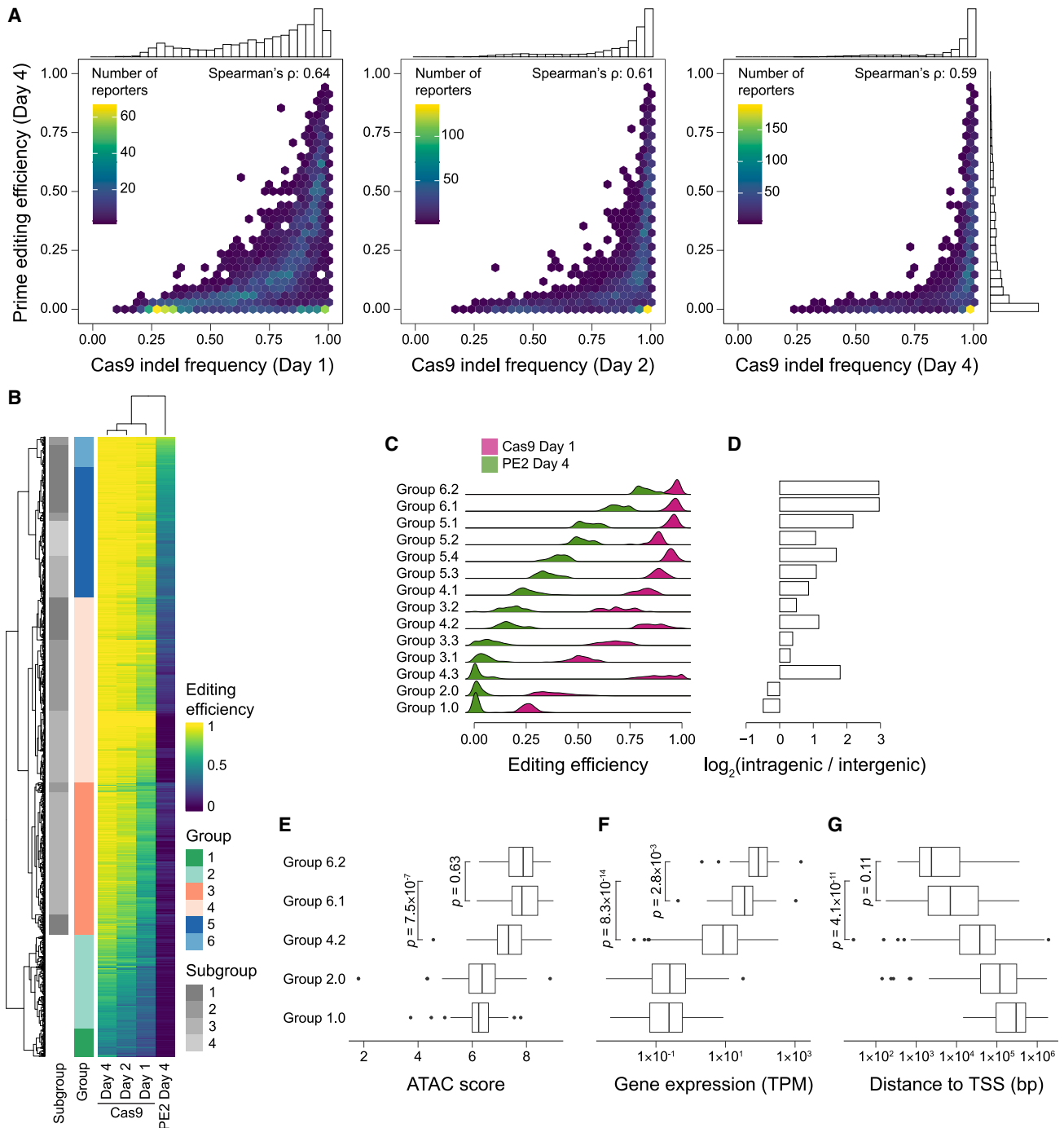
Figure 3. Comparison between Cas9 and prime editing, leveraging a common set of integrated reporters

(A) Comparison of editing efficiencies for Cas9 (day 1, 2, or 4 after transfection) vs. prime editing (day 4 after transfection) at an identical set of synHEK3 reporters. Plots colored based on the number of reporters assigned to each bin. 1D histograms of editing efficiencies plotted at top and right.

(B) Hierarchical clustering of synHEK3 reporters based on prime and Cas9 editing efficiencies.

(C) Density plot of prime and Cas9 editing efficiencies for 14 groups of synHEK3 reporters, ordered by mean prime editing efficiency.

(D) Bar graph of the log_2 ratio between number of intragenic vs. intergenic sites in each of the 14 groups.

(E–G) Comparison of properties of intragenic sites in selected groups. p value: two-sided Kolmogorov-Smirnov test. (E) Boxplot of assay for transposase accessible chromatin with sequencing (ATAC-seq) scores of selected groups. (F) Expression levels of the overlapping genes in TPM of selected groups (x axis; log_{10} scale). (G) Distance (bp) between the synHEK3 reporters in selected groups and the nearest TSS (x axis; log_{10} scale).

See also Figure S3.

Finally, we examined frequencies of alleles resulting from Cas9 editing inferred to derive from non-homologous end joining (NHEJ) vs. microhomology-mediated end joining (MMEJ) repair pathways. We found that MMEJ usage was significantly higher in synHEK3 sites near H3K27me3 and H3K9me3, consistent with observations by Schep et al.[24] (Figure S3G). Moreover, MMEJ usage increased as editing efficiencies decreased, with group 1.0 sites exhibiting the highest MMEJ/(MMEJ + NHEJ) ratios (median: 0.54; Figure S3H). Finally, within genes, MMEJ usage steadily increased along the gene body (Figures S3I and S3J), suggesting a gradient effect posed by chromatin on DNA break repair kinetics such that faster repair mechanisms are favored immediately downstream of TSSs.[24,41] We speculate that a similar TSS-distance gradient may affect which repair mechanisms ultimately resolve prime editing-induced single strand breaks (SSBs) and intermediate structures.

Investigating the *cis*-chromatin context dependencies of *trans*-acting factors influencing prime editing

DNA damage repair (DDR) is tightly regulated by local chromatin context.[42] To explore mechanisms mediating chromatin context-specific regulation of prime editing, we sought to develop a multiplex perturbational framework with which we could capture differences in prime editing outcomes in various chromatin environments upon knocking down various DDR-relevant factors. Potential strategies for this goal include Repair-seq[13] or well-based screening of pre-integrated DNA repair reporters.[43] However, the former averages over genomic contexts, while the latter does not scale effectively. We therefore designed a DDR-focused genetic screen, in which perturbations are coupled to outcomes at pre-integrated synHEK3 reporters with single-cell molecular profiling.[44–46] To co-profile the perturbation(s) received by a given cell and prime editing outcomes at its synHEK3 reporters, we sought to use T7 *in situ* transcription (IST)[47,48] to drive RNA production from both perturbational and reporter constructs in fixed, permeabilized nuclei and to then capture these molecules with sci-RNA-seq3, a scalable method for single-cell transcriptional profiling.[49]

We focused on 22 and 28 uniquely barcoded synHEK3 reporters in two monoclonal K562 lines derived from the original pool ("clone 3" and "clone 5"), whose integration sites spanned a range of chromatin environments (Figure 4A; Table S3). Prime editing efficiencies in these monoclonal lines were highly correlated with those estimated for the same reporters in the original polyclonal population (Figure S4A).

To perturb *trans*-acting factors, we chose a doxycycline (Dox)-inducible short hairpin RNA (shRNA) system, which can trigger a potent knockdown effect at single copy and is orthogonal to prime editing.[50,51] We constructed a library of 304 shRNAs against 76 genes, including 74 DDR-related genes (10 unexpressed) and 2 luciferase genes (Table S5). The DDR-related, expressed genes comprised hits found by Repair-seq,[13,52] genes in other major DDR pathways, and epigenetic factors involved in H3K79me2 metabolism (Figure S4B). To enable post-fixation enrichment and targeted capture of the shRNA transcripts, we modified the lentiviral construct to contain a T7 promoter upstream of the shRNA and a RT primer-binding site (PBS) (Figure S4C).

We chose to model 6-bp insertions for several reasons. First, random insertions minimized potential biases due to the insertion sequence.[18] Second, 6-bp insertions achieved high baseline editing efficiencies (~20% across 50 synHEK3 sites), increasing signal-to-noise. Third, the random 6-mers serve as unique molecular identifiers (UMIs) for counting unique prime editing events at each integration site from sequencing data.

To co-capture perturbations and prime editing outcomes, cells were profiled via a modified sci-RNA-seq3 workflow (Figure 4A). The first modification occurred before the RT step, in that we performed T7 IST on methanol-fixed nuclei to drive RNA expression from both synHEK3 reporters and the shRNA construct (Figure 4B). Second, at the RT step, a primer cocktail targeting mRNAs (oligo-dT), synHEK3, and shRNA transcripts was used to simultaneously capture the transcriptome, editing outcome, and perturbation(s) in a given cell (Figure 4C). Finally, toward the end of the protocol, nuclear lysate was split; half was subjected to conventional sci-RNA-seq3 library preparation and half to enrichment PCR targeting the synHEK3 and shRNA-derived transcripts (Figure 4D).

For each experiment, three libraries were sequenced, corresponding to the cellular transcriptome, shRNA integrants and synHEK3 integrants (Figure S4D). Cellular transcriptomes were used to determine high-quality cells, and captured shRNA and synHEK3 transcripts were assigned to these cells based on matching combinatorial indices. Thus, for each profiled cell, we obtained the identity of shRNA(s) present in that cell as well as the editing outcome at each of its 22 (clone 3) or 28 (clone 5) synHEK3 sites. We then applied a quantitative trait locus-style analysis to this matrix, for each synHEK3-shRNA pair, comparing the editing frequency at the synHEK3 reporter for cells that did vs. did not receive the shRNA (Figure 4E).

Identification of *trans*-acting regulatory factors of prime editing

We recovered 667,810 high-quality cells as assessed by complexity of the cellular transcriptome. SynHEK3-derived reads assigned to the two clones were well separated, indicating a low collision rate (Figure S5A). After quality control filters, we retrieved 229,657 cells (105,583 and 124,074 for clones 3 and 5, respectively), for which at least one shRNA and one synHEK3 integrant were detected. For clones 3 and 5, respectively, the median numbers of cells per shRNA were 1,604 and 2,208, shRNAs detected per cell were 4 and 5, and synHEK3 BCs detected per cell were 7 and 11 (Figure S5B). To determine whether we could reliably estimate editing efficiency with this approach, we sampled a small fraction of the cells in this screen and performed amplicon sequencing of the synHEK3 reporters. Prime editing efficiencies estimated from single-cell data were almost perfectly correlated with those estimated by bulk amplicon sequencing (Figure S5C).

To detect shRNA-mediated knockdown effects on prime editing outcomes, we calculated and compared editing efficiencies in cells with vs. without a given shRNA (binomial test). Across all synHEK3-shRNA combinations in the two monoclonal lines, we observed a clear excess of significant synHEK3-shRNA pairs compared with pairs involving control shRNAs (Figure 5A). 337/7,168 and 250/5,632 synHEK3-shRNA pairs were significant at a
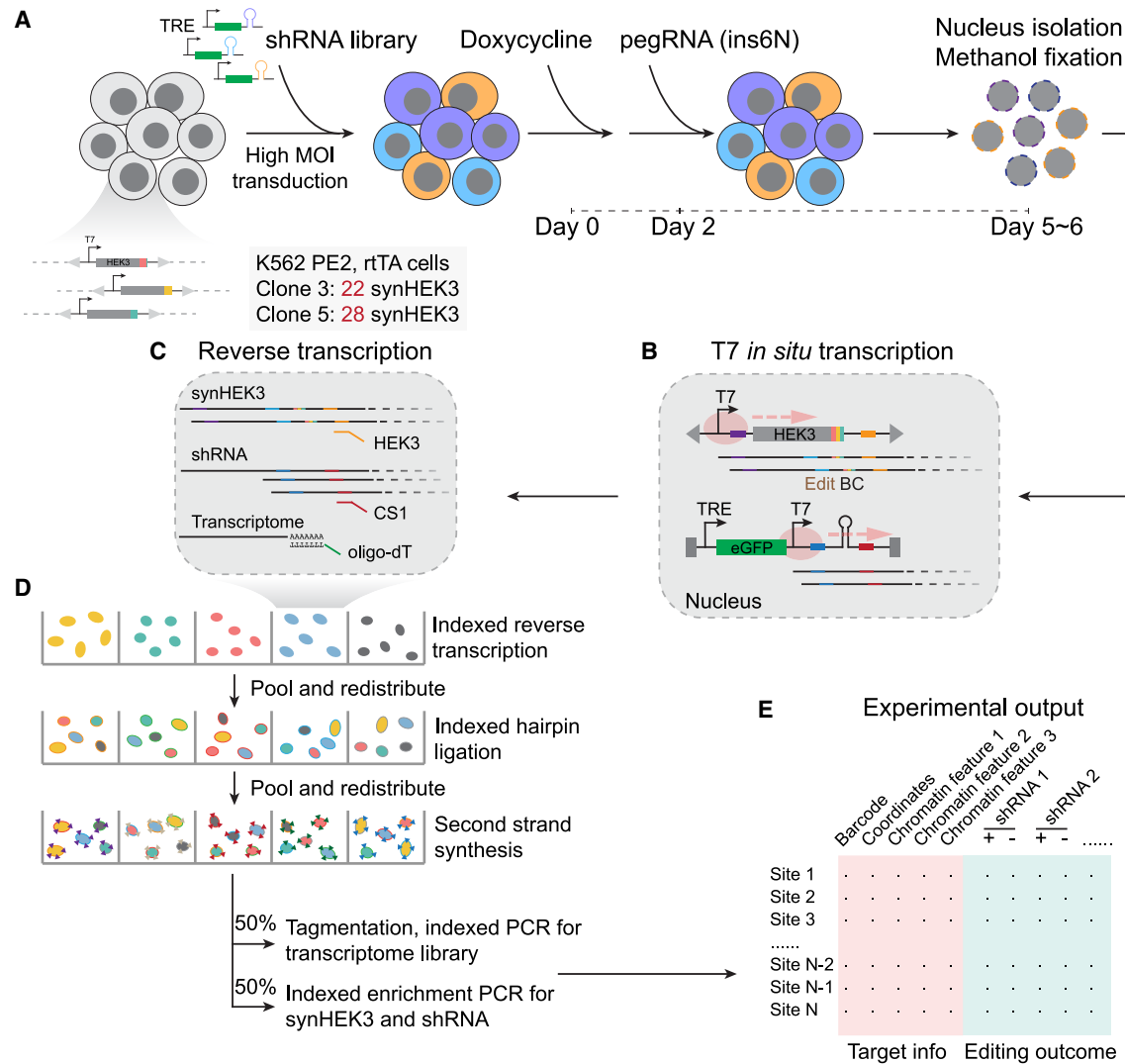
Figure 4. Dissecting chromatin context-dependent regulation of prime editing using a modified sci-RNA-seq3 workflow
Experimental workflow of the pooled shRNA screen.
(A) The two monoclonal K562 lines used in this experiment stably expressed PE2 and reverse tetracycline transactivator (rtTA) and together contained 50 unique synHEK3 reporters. Cells were transduced with the TRE-shRNA library at a high multiplicity of infection (MOI) and treated with doxycycline. On day 2, cells were transfected with pegRNAs to introduce random 6-bp insertions at synHEK3 reporters. After 3–4 days, nuclei were extracted and fixed. TRE, tetracycline response element.
(B) Fixed nuclei were subjected to IST with T7 polymerase (pink circle) to produce transcripts from synHEK3 and shRNA constructs.
(C) Nuclei were distributed to 96-wells for indexed RT. In each well, a cocktail of three indexed RT primers were used: oligo-dT primers and synHEK3- and shRNA-specific primers.
(D) After RT, nuclei were pooled and redistributed into 96-well plates for indexed hairpin ligation. Then, they were pooled and split to final 96-well plates. After second-strand synthesis, nuclei were lysed, and the resulting lysates were split into two plates. One plate underwent Tn5 tagmentation and indexed PCR to generate a transcriptome library. The other plate was used for indexed enrichment PCR targeting the synHEK3 and shRNA transcripts.
(E) For each synHEK3 reporter, editing outcomes were computed and compared between cells with and without a specific shRNA.
See also Figure S4.

false discovery rate (FDR) of 5% in clone 5 and clone 3, respectively. This excess of highly significant p values does not appear to be driven by differences in the number of cells associated with candidate-targeting vs. control shRNAs (Figure S5D).

shRNAs targeting several genes were both highly significant and reproducible across the two clones (Figure 5B). The most prominent targets include major components of the MMR

pathway (PMS2 and MLH1), previously reported to influence prime editing outcomes.[13,14] To better visualize these effects, we consolidated synHEK3-shRNA pairs targeting each MMR-related factor (Figures 5C and S5E). The strongest prime editing-promoting effects were observed for shRNAs against PMS2 and MLH1 (homologs of bacterial MutLα; Figure 5D), which form a heterodimer and coordinate multiple repair steps
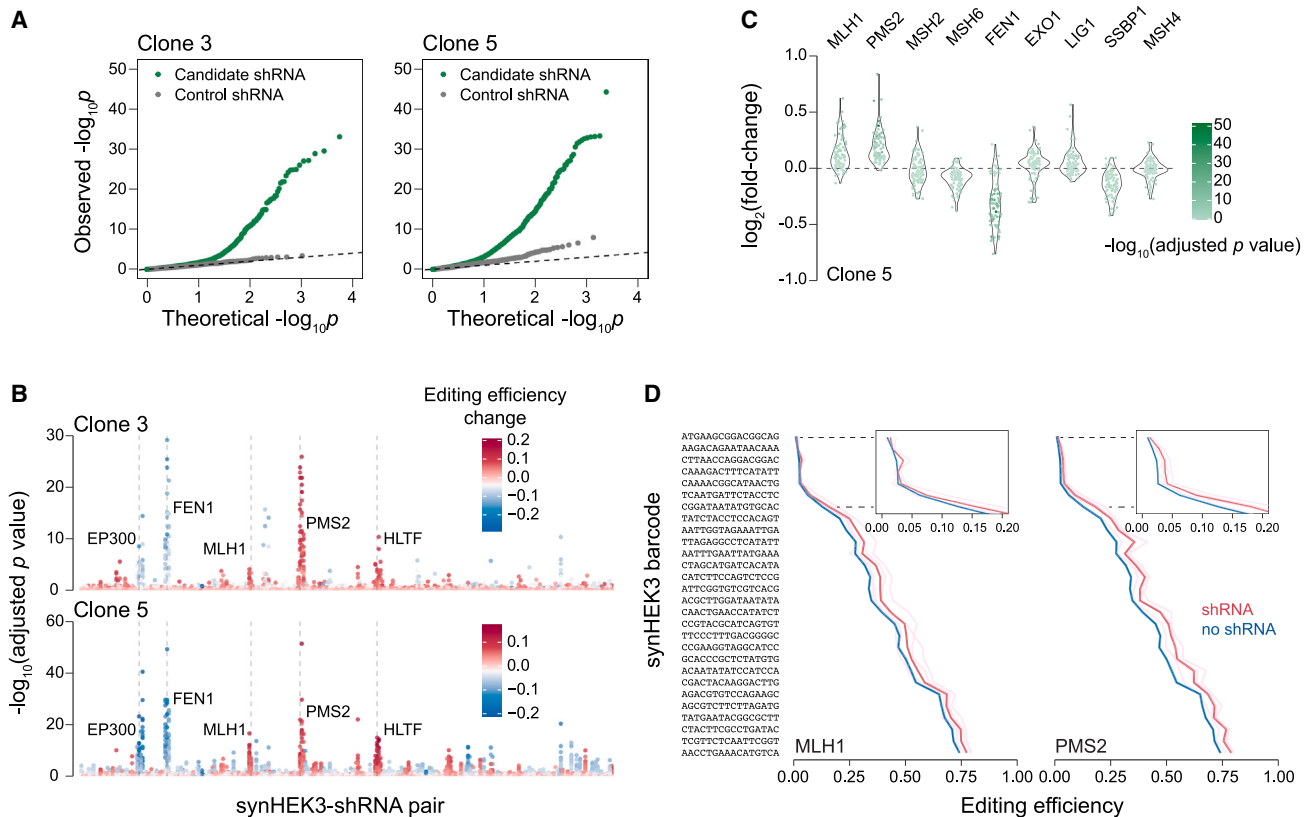
Figure 5. Effects of perturbing MMR-related genes on prime editing

(A) Quantile-quantile plot (Q-Q plot) of statistical significance ($-\log_{10}$) of synHEK3-shRNA pairs in clones 3 (left) and 5 (right). Candidate shRNAs (green) and control shRNAs (gray) are plotted separately. p value: binomial test.

(B) Plots of adjusted p values ($-\log_{10}$) of all synHEK3-shRNA pairs. Target genes with high statistical significance are annotated. Points are colored by editing efficiency changes caused by corresponding shRNAs.

(C) Effects of shRNAs targeting MMR-related genes in clone 5. $\log_2$ fold-changes of prime editing efficiencies of synHEK3-shRNA pairs are plotted and colored by their corresponding adjusted p values ($-\log_{10}$).

(D) Effects of shRNAs against MLH1 and PMS2. Pink lines: editing frequencies in cells with individual shRNAs; red line: mean editing frequencies of the gene-targeting shRNAs; light blue lines: control editing frequencies for individual shRNAs (not visible because low variance relative to mean line, shown in blue); and blue line: mean control editing frequencies.

See also Figure S5.

after mismatch recognition.[53] Knocking down FEN1, a 5′ DNA flap endonuclease, led to strong suppression of prime editing (Figure S5F), consistent with its reported role.[13] Perturbing the bacterial MutSα homologs MSH2 and MSH6[53] did not yield consistent effects, with MSH6 even leading to a slight decrease in prime editing efficiency in clone 5. MSH2/MSH6 dimer recognizes 1–2-bp mismatches, while MSH2/MSH3 dimer detects longer indels (>2 bp). We speculate that knocking down MSH6 releases sequestered MSH2, allowing for more efficient detection and correction of the 6-bp insertions. The mild effect for MSH2 might suggest a different mechanism used in recognizing 6-bp insertions. Knocking down other MMR-related factors (e.g., EXO1, LIG1, SSBP1, MSH4) did not yield consistent effects (Figures 5C and S5E).

In addition to MMR genes, targeting EP300 decreased prime editing efficiencies across all sites (Figure S5G). EP300 encodes the p300 transcriptional coactivator,[54] and its direct involvement in DDR remains elusive. We cannot rule out the

possibility that its effects here result from downregulation of global transcriptional output including expression of the prime editor complex.

To our surprise, knocking down the methyltransferase (DOT1L) and demethylase (KDM2B)[55] of H3K79 did not lead to significant changes in the prime editing efficiencies (Figure S5H). To confirm this result, we performed both bulk RNA sequencing (RNA-seq) and measurements of prime editing efficiencies in synHEK3-bearing K562 cells treated with EPZ-5676, a potent DOT1L inhibitor.[56] Expression levels of genes containing synHEK3 insertions in the single-cell screen were not significantly affected, despite massive expression changes induced by EPZ-5676 (Figure S5I). There was a global reduction in prime editing efficiency, but no significant relationship between changes in prime editing efficiencies and baseline H3K79me2 levels at intragenic synHEK3 reporters or gene expression changes (Figure S5J). These results suggest that H3K79me2 is a bystander of high prime editing efficiency, rather than playing a direct role.
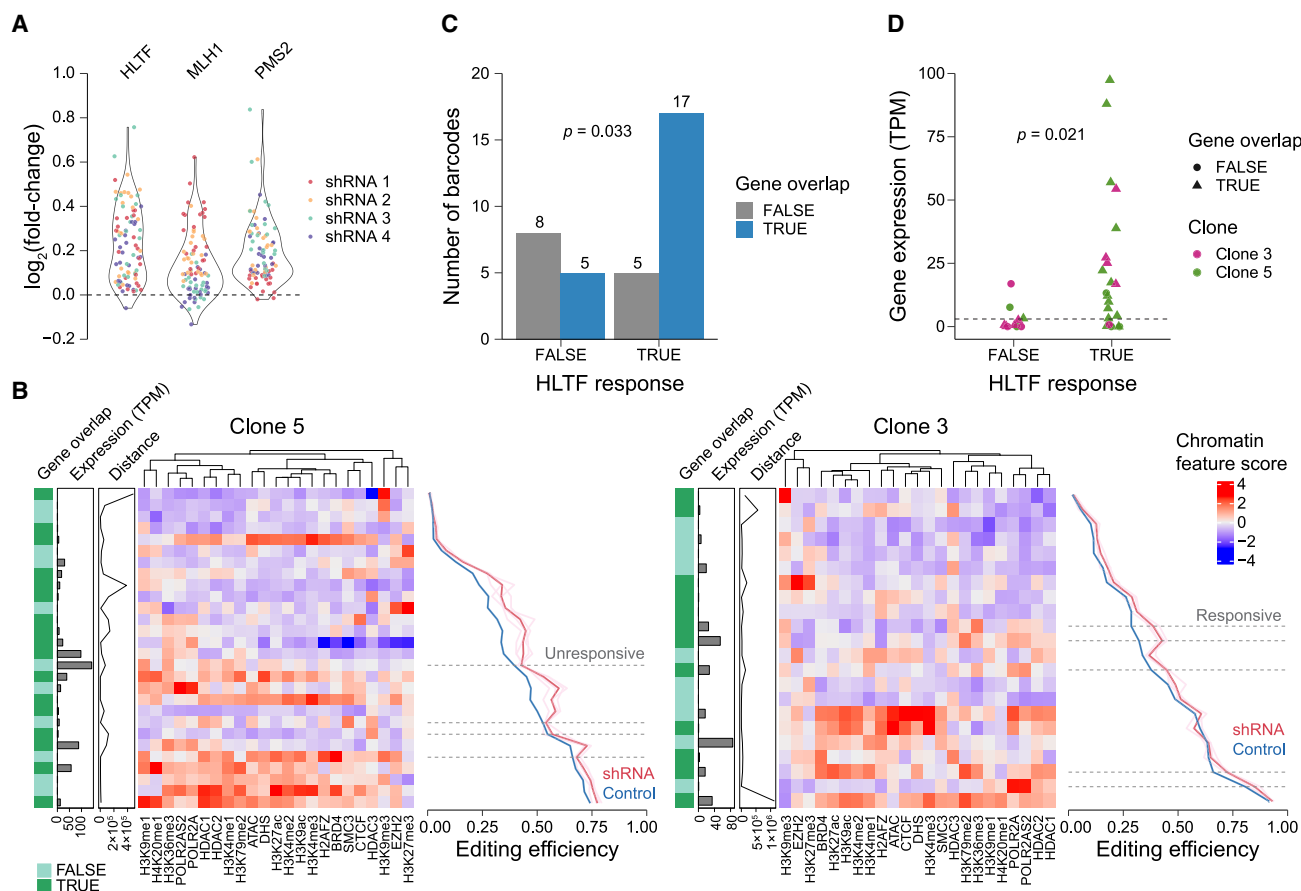
Figure 6. Chromatin context-specific response to HLTF inhibition

(A) Violin plot of fold-changes of editing efficiency of synHEK3 sites in response to inhibition of HLTF, MLH1, and PMS2. Points colored by shRNA identity.

(B) Heatmap of synHEK3 reporters (row) and their responses to shRNAs against HLTF (left: clone 5; right: clone 3). Leftmost bar annotates the overlapping status of synHEK3 reporters with GRCh38 gene annotations. Second left bar indicates the expression status of the overlapping or nearest gene in TPM. Third left bar indicates distances (bp) to corresponding TSS of gene-overlapping synHEK3 reporters or nearest genes for synHEK3 reporters outside genes. Middle heatmap was generated using scaled chromatin feature scores of synHEK3 reporters and clustered by column. Right line plot shows effects of shRNAs against HLTF. Pink lines: editing frequencies in cells with individual shRNAs; red line: mean editing frequencies of the gene-targeting shRNAs; light blue lines: control editing frequencies for individual shRNAs (not visible because low variance relative to mean line, shown in blue); blue line: mean control editing frequencies. Dashed lines: sites showing differential response to HLTF inhibition.

(C) Bar plot of synHEK3 reporter counts based on their responsiveness to HLTF inhibition and overlapping status with GRCh38 gene annotations. p value: Fisher's exact test.

(D) Expression of genes overlapping or proximal (within 10 kb) to a synHEK3 reporter. Genes above the dashed line (TPM = 3) are considered expressed. p value: two-sided Kolmogorov-Smirnov test.

See also Figure S6.

HLTF works as a chromatin context-dependent repressor of prime editing

Another strong hit in the screen was HLTF, a member of the SWI/SNF2 family that contains an E3 ubiquitin ligase domain. HLTF mediates polyubiquitination of PCNA, which promotes replication through DNA lesions in an error-free manner.[57] A previous report has shown knocking down HLTF led to a slight decrease in the frequency of prime editing-mediated single-nucleotide substitutions.[13] In contrast, we observed strong upregulation of 6-bp insertions, comparable to changes induced by shRNAs against MLH1 and PMS2 (Figure 6A), suggesting differences in mechanisms involving HLTF in the repair of short vs. longer mismatches.

Inspection of individual synHEK3 insertion sites revealed heterogeneous responses to HLTF inhibition (Figures 6B and S6A). Due to editing frequencies being bounded on [0, 1], there is an inverse correlation between baseline editing efficiencies and fold-changes in responses to perturbations (Figure S6C). Therefore, we focused on synHEK3 reporters with intermediate baseline editing frequencies (0.2–0.9). In clone 5, 4/21 of these sites showed much weaker increases in editing frequencies compared with reporters with similar starting editing frequencies. In clone 3, reporters were overall less responsive to HLTF inhibition, but 5/14 sites showed stronger and statistically significant upregulation of editing frequencies (Figures S6A and S6D). Sites that were less
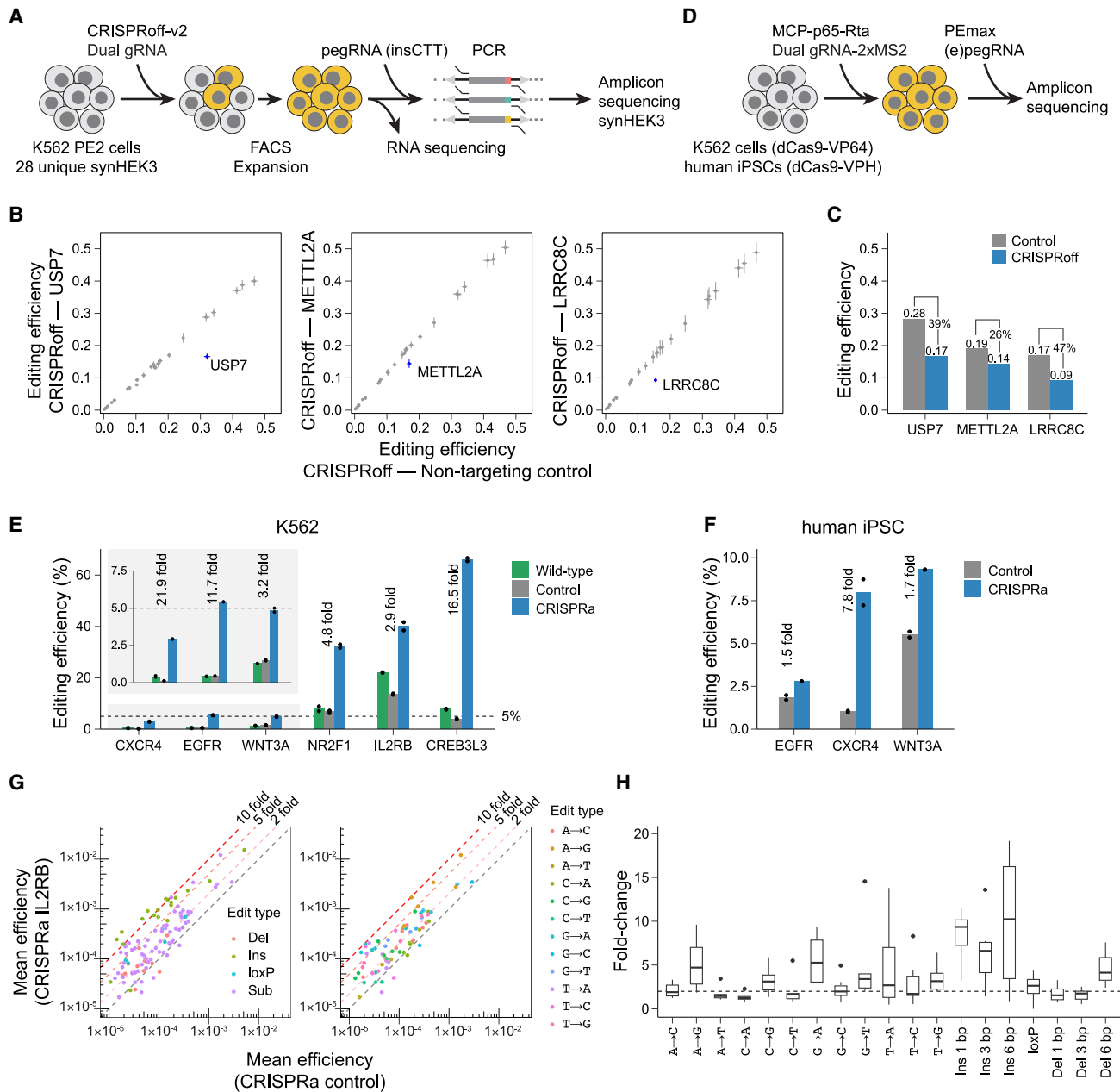
Figure 7. Modulating prime editing outcomes by epigenetic conditioning

(A) Workflow of the CRISPRoff experiment.

(B) Scatter plots of mean prime editing efficiencies of synHEK3 reporters in cells transfected with CRISPRoff gRNAs targeting *USP7*, *METTL2A*, and *LRRC8C* promoters. SynHEK3 reporters in corresponding target genes are labeled. Error bars correspond to standard deviation of measured editing efficiencies.

(C) Bar plot of prime editing efficiency changes of synHEK3 reporters in CRISPRoff experiment. Control editing efficiencies are predicted efficiencies using linear models trained in synHEK3 reporters that are not in the corresponding CRISPRoff target genes as shown in (B).

(D) Workflow of the CRISPRa experiments.

(E) Prime editing efficiency (%) at endogenous gene targets in K562 cells, with or without epigenetic editing via CRISPRa. Green bars show mean prime editing efficiencies in a wild-type K562 cell line, which received only PEmax and (e)pegRNA. Gray bars show mean prime editing efficiencies when control promoters were activated via CRISPRa. Blue bars show mean prime editing efficiencies when target gene promoters were activated. Fold-changes are calculated between the CRISPRa (blue) and control (gray) groups. Inset shows a zoomed view of the first three genes.

(F) Prime editing efficiency (%) at endogenous gene targets in a human iPSC line, with or without epigenetic editing via CRISPRa. Gray bars show mean prime editing efficiencies when control promoters were activated via CRISPRa. Blue bars show mean prime editing efficiencies when target gene promoters were activated.

responsive to HLTF inhibition showed slightly higher response to shRNAs against PMS2, an effect more obvious in clone 5 (Figure S6B). We grouped the selected synHEK3 reporters based on their responsiveness to HLTF shRNA knockdown and counted their overlapping status with annotated genes. The responsive group was enriched for gene-overlapping sites (5.4-fold, Fisher's exact test $p = 0.033$; Figure 6C). Furthermore, when taking expression status of overlapping or nearby genes (within 10 kb) into consideration, responsive sites tended to be near actively transcribed genes, while the unresponsive sites were mostly in non-transcribing regions (9.2-fold mean gene expression difference; $p = 0.021$; Figure 6D). We validated these observations in cells transduced with individual shRNAs (Figure S6E).

HLTF might suppress prime editing in transcribed regions through its role in epigenetic regulation or alternatively through its role in DNA repair. To distinguish between these possibilities, we performed bulk RNA-seq and ATAC-seq in cells expressing shRNAs targeting HLTF (shHLTF.2367, shHLTF.2623). However, we observed minimal changes upon HLTF knockdown, including at genes and ATAC peaks near HLTF-responsive synHEK3 sites (Figures S6F and S6G; Table S6). This result suggests that the context-specific effect of HLTF on prime editing is not mediated through its role in chromatin remodeling.

Modulating prime editing outcomes by targeted epigenetic reprogramming

Several lines of evidence presented here suggest that active transcriptional elongation enhances prime editing. First, H3K79me2 is deposited by DOT1L, which is part of the RNA Pol II transcriptional elongation complex. H3K79me2 is strongly predictive of intragenic prime editing efficiency. At the same time, shRNA-mediated knockdown of DOT1L failed to appreciably alter prime editing efficiencies, a result corroborated by pharmacological inhibition of DOT1L. Second, intragenic prime editing efficiencies were also correlated with transcription levels, proximity to TSS, and, more weakly, transcript orientation. Third, our analyses of the differential outcomes of prime and Cas9 editing on identical synHEK3 reporters suggested that while chromatin accessibility may enhance both prime and Cas9 editing, higher levels of transcription and TSS proximity appear to specifically promote prime editing.

We next sought to exploit the apparent relationship between active transcriptional elongation and prime editing to modulate the efficiency of prime editing. We chose three intronic synHEK3 reporters identified in the monoclonal line, clone 5, used above. These reporters were in the genes USP7, METTL2A, and LRRC8C (Figure S7A). We transiently transfected these cells with CRISPRoff-v2, an epigenetic editing tool,[58] and pairs of gRNAs to silence these genes' promoters, and subsequently measured prime editing efficiency across synHEK3 reporters (Figure 7A). After calibrating to synHEK3 reporters outside of target genes (Figure 7B; Table S7), we observed 39%, 26%, and 47% reductions in editing frequencies of synHEK3 reporters in USP7, METTL2A, and LRRC8C, respectively (Figure 7C). These reductions are likely direct effects of gene silencing, as other synHEK3 reporters were not affected, and RNA-seq indicated highly specific target gene silencing (Figure S7B). Interestingly, the magnitude of editing efficiency reductions was on par with the reductions in gene expression induced by CRISPRoff (53%, 30%, and 55%, respectively).

Enhancing prime editing efficiency by transient gene activation

Encouraged by these results, we next examined the possibility of increasing intragenic prime editing efficiency by preceding it with transient gene activation by CRISPRa.[59] For these experiments, we sought to enhance prime editing of the endogenous genome, rather than synHEK3 reporters. We initially selected four target genes, CXCR4, IL2RB, EGFR, and CDKL5, and used prime editing to introduce a different substitution per target, some clinically relevant.[60–62] We first transfected a K562 cell line stably expressing the dCas9-VP64 fusion[63] with a pair of gRNAs (2XMS2) targeting the target gene's promoter, along with an MCP-p65-Rta fusion protein. Then, 2 days later, we transfected the cells with PEmax and pegRNA to program the desired mutations (Figure 7D). For three of the targets, we observed increases in prime editing efficiency, ranging from 40% (EGFR, $p = 0.076$) to 21.9-fold (CXCR4, $p = 0.001$) (Figures S7C–S7E; Table S7).

A potential limitation is that these pegRNAs had low baseline efficiencies. Focusing on a subset of previous targets and some new targets that could be strongly upregulated by CRISPRa (Figure S7D), we designed new (e)pegRNAs targeting them with DeepPrime. The newly designed (e)pegRNAs had baseline editing efficiencies ranging from 0.5% to 13.5%. After CRISPRa of corresponding promoters, we observed 2.9- to 16.5-fold upregulation in prime editing efficiencies (Figure 7E; Table S7). The newly designed pegRNA for IL2RB had the highest baseline efficiency (13.5%), which was boosted to 40% (2.9-fold) by CRISPRa. Our best post-CRISPRa efficiency was for a target in CREB3L3 and was boosted from 4% to 66% (16.5-fold). Of note, we measured editing efficiencies of the same (e)pegRNAs in a WT K562 cell line and observed similar baselines, suggesting that prime editing is not suppressed by the CRISPRa machinery (Figure 7E). To test whether this strategy was effective in a different cellular context, we preceded prime editing of CXCR4, WNT3A, and EGFR with CRISPRa in human iPSCs[64] and observed 7.8-, 1.7-, and 1.5-fold increases in editing, respectively (Figure 7F; Table S7).

To evaluate whether this strategy could facilitate various types of edits, we designed 7 libraries of epegRNAs to install 19 different variants within 7 different exons of IL2RB, including all possible substitutions, small insertions and deletions, and a long insertion (Figures S7F and S7G; Tables S1 and S7). Preceding prime editing with CRISPRa resulted in 1.6- to 5.8-fold increases in editing efficiency for the 7 exons (Figure S7H). The epegRNAs targeting

---

(G) Scatter plots of mean prime editing efficiencies measured in control vs. CRISPRa cells in exons of IL2RB. Points colored by edit type. x and y axes in $\log_{10}$ scale. The pink and red lines indicate 2-, 5-, and 10-fold differences between the CRISPRa and control groups.

(H) Boxplot of prime editing efficiency fold-change for all variants. The dashed line indicates a fold-change of 2.

See also Figure S7.

Cell
Article

CellPress
OPEN ACCESS

exon 1 of *IL2RB* were the least responsive to CRISPRa, potentially due to the CRISPRa and editing target sites being only 300 bp apart. The installation of small insertions (1, 3, and 6 bp) were most responsive to CRISPRa (median fold-changes ranging from 6.6- to 10.2-fold), followed by other edit types (Figures 7G and 7H). Among single-nucleotide substitutions, $G \rightarrow A$ and $A \rightarrow G$ edits were most responsive to CRISPRa (median fold-changes of 5.3- and 4.7-fold, respectively; Figure 7H). We conclude that epigenetic reprogramming strategy is effective for enhancing the prime editing efficiencies of all edit types.

Taken together, these experiments reinforce the strong link between transcriptional activity and prime editing outcome. They also demonstrate the feasibility of enhancing prime editing outcomes via transient activation of the gene to which edits are being introduced.

DISCUSSION

At the core of this study is a simple T7 promoter-bearing reporter construct and straightforward protocol that leverages T7 IVT and IST to allow classic questions surrounding chromatin position effects to be tackled in either bulk or single-cell format. When applied in bulk, T7 IVT enabled near-complete mapping of densely integrated reporters and measurement of position-dependent regulatory effects on prime editing. When incorporated into single-cell RNA-seq protocol, T7 IST enabled co-profiling of non-transcribing genomic constructs. By further incorporation into an shRNA genetic screen, the interaction between *trans*-acting factors and genome editing could be stratified by chromatin context. Of note, the method is straightforward to adapt, requiring only the inclusion of a 19-bp T7 promoter to any reporter construct, such that it has the potential to make capturing precise genomic coordinates routine for any bulk or single-cell assay in which reporter or effector constructs are randomly integrated.

We measured prime editing efficiency at over 4,000 genomic locations and quantified the correlation of 23 chromatin features with those efficiencies. Through beta regression modeling, we show that H3K79me2 is the best predictor of efficacious prime editing, but multiple lines of evidence suggest that this is probably due to its strong correlation with active transcriptional elongation rather than a direct effect. Importantly, this same beta regression model, based entirely on position effects on prime editing of the synHEK3 reporter, successfully predicts the relative editing efficiencies of high-quality epegRNAs targeting endogenous genomic target sites, providing value that is orthogonal to sequence-based prediction tools.

Examining the intersection of *trans*-acting factors with the *cis*-chromatin environment in shaping prime editing, we identified HLTF as a context-dependent repressor of prime editing, as knocking it down preferentially enhanced prime editing at sites undergoing active transcription. Although we cannot rule out a mechanism involving some unknown function of HLTF, our observations more plausibly relate to HLTF's role in DNA repair[57,65] than its role in chromatin remodeling. HLTF has previously been shown to repress PE3 but not PE2.[13] Taken together with our observations here, we hypothesize that HLTF might be unevenly distributed along chromatin or preferentially recognize special SSB repair intermediates. Actively transcribed regions of the genome

might be under intense scrutiny by HLTF such that DNA lesions can be corrected immediately by DDR to avoid accumulation of mutations in the transcribed genome over cell cycles. However, further studies, in particular biochemical profiling of HLTF's transient interactions during prime editing of transcribed regions, will be necessary to elucidate the underlying mechanism.

The correlation between active transcriptional elongation and prime editing efficiency prompted us to explore whether manipulating the epigenetic context of a target site could alter the efficiency with which it was edited. Indeed, with CRISPRoff, we showed that targeted gene silencing reduces prime editing efficiency for intragenic targets, with effect sizes on par with the fold-reduction in gene expression. Conversely, prime editing efficiencies of endogenous, intragenic targets can be markedly enhanced by first "conditioning" the locus with CRISPRa. We validated the effectiveness of this approach for multiple loci, at a range of distances from the TSS, in both K562 and iPSC lines, and for all edit types.

Across target sites in 6 genes that were successfully upregulated with CRISPRa in K562 cells, we observed a mean 10.2-fold (median 8.4-fold) increase in prime editing efficiency (Figure 7E), while across 133 combinations of edit types and target sites within the *IL2RB* locus, we observed a mean 4.1-fold (median 2.6-fold) increase (Figure 7G). These effect sizes are comparable to that of MMR inhibition (mean 7.7-fold increase).[13,14] Of note, the effectiveness of MMR inhibition is highly mutation dependent, with certain types of edits (e.g., longer indels) being less sensitive. While more experiments are required to fully understand the edit biases of our own strategy, we showed all prime editing-mediated edit types are robustly enhanced by CRISPRa conditioning, e.g., 6-bp insertions in *IL2RB* exhibited mean 10.0-fold (median 10.2-fold) enhancement (Figure 7H). The strategies are thus complementary, but given that this approach probably works through mechanisms orthogonal to MMR, they can potentially be combined.

In summary, we show that prime editing's efficiency is strongly impacted by the *cis*-chromatin landscape and promoted by active transcription in particular. Leveraging that insight, we further show how CRISPRa conditioning can be used to enhance the efficiency of intragenic prime editing, a strategy that may be useful in both basic research and therapeutic contexts. Finally, we note that the methods described here, in particular T7-assisted reporter mapping, may be generally useful for studying position effects on other chromatin-modulated processes.

Limitations of the study
Several limitations and technical nuances merit discussion. First, it is likely that there is more to the epigenome than is captured by 23 chromatin features. We note that synHEK3 reporters in distal intergenic regions exhibit highly variable efficiencies (Figure S2K) and wonder if heterogeneity within heterochromatin beyond what is captured by ENCODE assays[34] or, alternatively, higher-order structures (e.g., chromatin looping), might explain some of this.

Second, our initial experiment leveraged a single pegRNA introducing the same insertion to a constant target site in a single-cell type, thereby neglecting potential interactions between the mutation type, target sequence, *cis*-chromatin environment, and cell type. Although downstream experiments suggest that

our results generalize, there may be nuances that we have missed.

Third, in our shRNA single-cell screen, the effect sizes of the perturbations on prime editing efficiencies were not as large as those measured with Repair-seq.[13] However, we do not attribute this to inefficient knockdown by shRNAs (Figure S6H). Instead, this observation likely reflects the physiological regulation of 6-bp insertions, as they tend to escape MMR and have higher baseline editing efficiencies (20%–90% for the visualized sites), which may mute effect sizes (Figure S6C).

Finally, a reasonable worry in applying epigenetic reprogramming to condition loci for prime editing is the possibility of interference between editors. We do not observe any consistent suppression of prime editing in cells constitutively expressing the CRISPRa machinery (Figure 7E), but an anecdotal example (exon 1 of *IL2RB*) is consistent with steric interference if the CRISPRa and prime editing target sites are too close (Figure S7H). Using orthogonal (epi)genome engineering tools and/or transiently present CRISPR RNPs might further reduce the risk of cross-talk between different CRISPR-mediated modules.

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 STUDY PARTICIPANT DETAILS
  - Cell lines and cell culture
- METHOD DETAILS
  - Molecular cloning
  - General transfection strategies
  - Cell line generation
  - Prime editing experiment in synHEK3 reporters
  - Prime editing experiment in endogenous sites
  - Cas9 RNP editing experiment
  - CRISPRoff experiment
  - CRISPRa experiments
  - Design and test of IL2RB epegRNA libraries
  - Quantitative PCR analysis
  - T7-assisted reporter mapping
  - synHEK3 editing library preparation
  - Lentivirus production and transduction
  - Pooled shRNA screen
  - Amplicon sequencing
  - RNA sequencing
  - ATAC sequencing
  - T7-assisted reporter mapping library analysis
  - ENCODE dataset analysis
  - synHEK3 editing library analysis
  - sci-RNA-seq3 transcriptome library analysis
  - Pooled shRNA screen analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

REFERENCES

1. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature 576, 149–157. https://doi.org/10.1038/s41586-019-1711-4.

2. Choi, J., Chen, W., Suiter, C.C., Lee, C., Chardon, F.M., Yang, W., Leith, A., Daza, R.M., Martin, B., and Shendure, J. (2022). Precise genomic deletions using paired prime editing. Nat. Biotechnol. 40, 218–226. https://doi.org/10.1038/s41587-021-01025-z.

3. Jiang, T., Zhang, X.-O., Weng, Z., and Xue, W. (2022). Deletion and replacement of long genomic sequences using prime editing. Nat. Biotechnol. 40, 227–234. https://doi.org/10.1038/s41587-021-01026-y.

4. Anzalone, A.V., Gao, X.D., Podracky, C.J., Nelson, A.T., Koblan, L.W., Raguram, A., Levy, J.M., Mercer, J.A.M., and Liu, D.R. (2022). Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nat. Biotechnol. 40, 731–740. https://doi.org/10.1038/s41587-021-01133-w.

5. Chen, P.J., and Liu, D.R. (2023). Prime editing for precise and highly versatile genome manipulation. Nat. Rev. Genet. 24, 161–177. https://doi.org/10.1038/s41576-022-00541-1.

6. Liu, B., Dong, X., Cheng, H., Zheng, C., Chen, Z., Rodríguez, T.C., Liang, S.-Q., Xue, W., and Sontheimer, E.J. (2022). A split prime editor with

untethered reverse transcriptase and circular RNA template. Nat. Biotechnol. *40*, 1388–1393.

7. Jang, H., Jo, D.H., Cho, C.S., Shin, J.H., Seo, J.H., Yu, G., Gopalappa, R., Kim, D., Cho, S.-R., Kim, J.H., et al. (2022). Application of prime editing to the correction of mutations and phenotypes in adult mice with liver and eye diseases. Nat Biomed Eng *6*, 181–194.

8. Choi, J., Chen, W., Minkina, A., Chardon, F.M., Suiter, C.C., Regalado, S.G., Domcke, S., Hamazaki, N., Lee, C., Martin, B., et al. (2022). A time-resolved, multi-symbol molecular recorder via sequential genome editing. Nature *608*, 98–107. https://doi.org/10.1038/s41586-022-04922-8.

9. Chen, W., Choi, J., Nathans, J.F., Agarwal, V., Martin, B., Nichols, E., Leith, A., Lee, C., and Shendure, J. (2021). Multiplex genomic recording of enhancer and signal transduction activity in mammalian cells. Preprint at bioRxiv. https://doi.org/10.1101/2021.11.05.467434.

10. Loveless, T.B., Carlson, C.K., Hu, V.J., Dentzel Helmy, C.A., Liang, G., Ficht, M., Singhai, A., and Liu, C.C. (2021). Molecular recording of sequential cellular events into DNA. Preprint at bioRxiv. https://doi.org/10.1101/2021.11.05.467507.

11. Erwood, S., Bily, T.M.I., Lequyer, J., Yan, J., Gulati, N., Brewer, R.A., Zhou, L., Pelletier, L., Ivakine, E.A., and Cohn, R.D. (2022). Saturation variant interpretation using CRISPR prime editing. Nat. Biotechnol. *40*, 885–895. https://doi.org/10.1038/s41587-021-01201-1.

12. Mathis, N., Allam, A., Kissling, L., Marquart, K.F., Schmidheini, L., Solari, C., Balázs, Z., Krauthammer, M., and Schwank, G. (2023). Predicting prime editing efficiency and product purity by deep learning. Nat. Biotechnol. *41*, 1151–1159. https://doi.org/10.1038/s41587-022-01613-7.

13. Chen, P.J., Hussmann, J.A., Yan, J., Knipping, F., Ravisankar, P., Chen, P.-F., Chen, C., Nelson, J.W., Newby, G.A., Sahin, M., et al. (2021). Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell *184*, 5635–5652.e29. https://doi.org/10.1016/j.cell.2021.09.018.

14. Ferreira da Silva, J., Oliveira, G.P., Arasa-Verge, E.A., Kagiou, C., Moretton, A., Timelthaler, G., Jiricny, J., and Loizou, J.I. (2022). Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. Nat. Commun. *13*, 760. https://doi.org/10.1038/s41467-022-28442-1.

15. Doman, J.L., Pandey, S., Neugebauer, M.E., An, M., Davis, J.R., Randolph, P.B., McElroy, A., Gao, X.D., Raguram, A., Richter, M.F., et al. (2023). Phage-assisted evolution and protein engineering yield compact, efficient prime editors. Cell *186*, 3983–4002.e26. https://doi.org/10.1016/j.cell.2023.07.039.

16. Nelson, J.W., Randolph, P.B., Shen, S.P., Everette, K.A., Chen, P.J., Anzalone, A.V., An, M., Newby, G.A., Chen, J.C., Hsu, A., et al. (2022). Engineered pegRNAs improve prime editing efficiency. Nat. Biotechnol. *40*, 402–410. https://doi.org/10.1038/s41587-021-01039-7.

17. Kim, H.K., Yu, G., Park, J., Min, S., Lee, S., Yoon, S., and Kim, H.H. (2021). Predicting the efficiency of prime editing guide RNAs in human cells. Nat. Biotechnol. *39*, 198–206. https://doi.org/10.1038/s41587-020-0677-y.

18. Koeppel, J., Weller, J., Peets, E.M., Pallaseni, A., Kuzmin, I., Raudvere, U., Peterson, H., Liberante, F.G., and Parts, L. (2023). Prediction of prime editing insertion efficiencies using sequence features and DNA repair determinants. Nat. Biotechnol. *41*, 1446–1456. https://doi.org/10.1038/s41587-023-01678-y.

19. Yu, G., Kim, H.K., Park, J., Kwak, H., Cheong, Y., Kim, D., Kim, J., Kim, J., and Kim, H.H. (2023). Prediction of efficiencies for diverse prime editing systems in multiple cell types. Cell *186*, 2256–2272.e23. https://doi.org/10.1016/j.cell.2023.03.034.

20. Mathis, N., Allam, A., Tálas, A., Benvenuto, E., Schep, R., Damodharan, T., Balázs, Z., Janjuha, S., Schmidheini, L., Böck, D., et al. (2023). Predicting prime editing efficiency across diverse edit types and chromatin contexts with machine learning. Preprint at bioRxiv. https://doi.org/10.1101/2023.10.09.561414.

21. Yarrington, R.M., Verma, S., Schwartz, S., Trautman, J.K., and Carroll, D. (2018). Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. Proc. Natl. Acad. Sci. USA *115*, 9351–9358. https://doi.org/10.1073/pnas.1810062115.

22. Isaac, R.S., Jiang, F., Doudna, J.A., Lim, W.A., Narlikar, G.J., and Almeida, R. (2016). Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. eLife *5*, e13450. https://doi.org/10.7554/eLife.13450.

23. Janssen, J.M., Chen, X., Liu, J., and Gonçalves, M.A.F.V. (2019). The chromatin structure of CRISPR-Cas9 target DNA controls the balance between mutagenic and homology-directed gene-editing events. Mol. Ther. Nucleic Acids *16*, 141–154. https://doi.org/10.1016/j.omtn.2019.02.009.

24. Schep, R., Brinkman, E.K., Leemans, C., Vergara, X., van der Weide, R.H., Morris, B., van Schaik, T., Manzo, S.G., Peric-Hupkes, D., van den Berg, J., et al. (2021). Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. Mol. Cell *81*, 2216–2230.e10. https://doi.org/10.1016/j.molcel.2021.03.032.

25. Daer, R.M., Cutts, J.P., Brafman, D.A., and Haynes, K.A. (2017). The impact of chromatin dynamics on Cas9-mediated genome editing in human cells. ACS Synth. Biol. *6*, 428–438. https://doi.org/10.1021/acssynbio.5b00299.

26. Gisler, S., Gonçalves, J.P., Akhtar, W., de Jong, J., Pindyurin, A.V., Wessels, L.F.A., and van Lohuizen, M. (2019). Multiplexed Cas9 targeting reveals genomic location effects and gRNA-based staggered breaks influencing mutation efficiency. Nat. Commun. *10*, 1598. https://doi.org/10.1038/s41467-019-09551-w.

27. Ochman, H., Gerber, A.S., and Hartl, D.L. (1988). Genetic applications of an inverse polymerase chain reaction. Genetics *120*, 621–623. https://doi.org/10.1093/genetics/120.3.621.

28. Potter, C.J., and Luo, L. (2010). Splinkerette PCR for mapping transposable elements in Drosophila. PLoS One *5*, e10168. https://doi.org/10.1371/journal.pone.0010168.

29. Schmidt, M., Zickler, P., Hoffmann, G., Haas, S., Wissler, M., Muessig, A., Tisdale, J.F., Kuramoto, K., Andrews, R.G., Wu, T., et al. (2002). Polyclonal long-term repopulating stem cell clones in a primate model. Blood *100*, 2737–2743. https://doi.org/10.1182/blood-2002-02-0407.

30. Hamada, M., Nishio, N., Okuno, Y., Suzuki, S., Kawashima, N., Muramatsu, H., Tsubota, S., Wilson, M.H., Morita, D., Kataoka, S., et al. (2018). Integration mapping of piggyBac-mediated CD19 chimeric antigen receptor T cells analyzed by novel tagmentation-assisted PCR. EBioMedicine *34*, pp. 18–26.

31. Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell *122*, 473–483. https://doi.org/10.1016/j.cell.2005.07.013.

32. Wilson, M.H., Coates, C.J., and George, A.L., Jr. (2007). PiggyBac transposon-mediated gene transfer in human cells. Mol. Ther. *15*, 139–145. https://doi.org/10.1038/sj.mt.6300028.

33. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330. https://doi.org/10.1038/nature14248.

34. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

35. Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V.M., Su, L., Xu, G., and Zhang, Y. (2005). hDOT1L links histone methylation to leukemogenesis. Cell *121*, 167–178. https://doi.org/10.1016/j.cell.2005.02.020.

36. Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.-E., Chen, J., Lazar, M.A., et al. (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Mol. Cell. Biol. *28*, 2825–2839. https://doi.org/10.1128/MCB.02076-07.

37. Feng, Q., Wang, H., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of H3-lysine 79 is mediated by a new

CellPress
OPEN ACCESS

Cell
Article

family of HMTases without a SET domain. Curr. Biol. *12*, 1052–1058. https://doi.org/10.1016/s0960-9822(02)00901-6.

38. Padeken, J., Methot, S.P., and Gasser, S.M. (2022). Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. Nat. Rev. Mol. Cell Biol. *23*, 623–640. https://doi.org/10.1038/s41580-022-00483-w.

39. Peters, A.H.F.M., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A.H.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol. Cell *12*, 1577–1589. https://doi.org/10.1016/s1097-2765(03)00477-5.

40. Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135. https://doi.org/10.1126/science.1084274.

41. Brinkman, E.K., Chen, T., de Haas, M., Holland, H.A., Akhtar, W., and van Steensel, B. (2018). Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks. Mol. Cell *70*, 801–813.e6. https://doi.org/10.1016/j.molcel.2018.04.016.

42. Price, B.D., and D'Andrea, A.D. (2013). Chromatin remodeling at DNA double-strand breaks. Cell *152*, 1344–1354. https://doi.org/10.1016/j.cell.2013.02.011.

43. Vergara, X., Manjon, A.G., Morris, B., Schep, R., Leemans, C., Sanders, M.A., Beijersbergen, R.L., Medema, R.H., and van Steensel, B. (2022). Widespread chromatin context-dependencies of DNA double-strand break repair proteins. Preprint at bioRxiv. https://doi.org/10.1101/2022.10.07.511243.

44. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell *167*, 1853–1866.e17. https://doi.org/10.1016/j.cell.2016.11.038.

45. Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. Cell *167*, 1867–1882.e21. https://doi.org/10.1016/j.cell.2016.11.048.

46. Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods *14*, 297–301. https://doi.org/10.1038/nmeth.4177.

47. Yin, Y., Jiang, Y., Lam, K.-W.G., Berletch, J.B., Disteche, C.M., Noble, W.S., Steemers, F.J., Camerini-Otero, R.D., Adey, A.C., and Shendure, J. (2019). High-throughput single-cell sequencing with linear amplification. Mol. Cell *76*, 676–690.e10. https://doi.org/10.1016/j.molcel.2019.08.002.

48. Askary, A., Sanchez-Guardado, L., Linton, J.M., Chadly, D.M., Budde, M.W., Cai, L., Lois, C., and Elowitz, M.B. (2020). In situ readout of DNA barcodes and single base edits facilitated by in vitro transcription. Nat. Biotechnol. *38*, 66–75. https://doi.org/10.1038/s41587-019-0299-4.

49. Martin, B.K., Qiu, C., Nichols, E., Phung, M., Green-Gladden, R., Srivatsan, S., Blecher-Gonen, R., Beliveau, B.J., Trapnell, C., Cao, J., et al. (2023). Optimized single-nucleus transcriptional profiling by combinatorial indexing. Nat. Protoc. *18*, 188–207. https://doi.org/10.1038/s41596-022-00752-0.

50. Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D.Y., Barbosa, I.A.M., Kwon, J.S., Guan, Y., et al. (2013). An optimized microRNA backbone for effective single-copy RNAi. Cell Rep. *5*, 1704–1713. https://doi.org/10.1016/j.celrep.2013.11.020.

51. Pelossof, R., Fairchild, L., Huang, C.-H., Widmer, C., Sreedharan, V.T., Sinha, N., Lai, D.-Y., Guan, Y., Premsrirut, P.K., Tschaharganeh, D.F., et al. (2017). Prediction of potent shRNAs with a sequential classification algorithm. Nat. Biotechnol. *35*, 350–353. https://doi.org/10.1038/nbt.3807.

52. Hussmann, J.A., Ling, J., Ravisankar, P., Yan, J., Cirincione, A., Xu, A., Simpson, D., Yang, D., Bothmer, A., Cotta-Ramusino, C., et al. (2021). Mapping the genetic landscape of DNA double-strand break repair. Cell *184*, 5653–5669.e25. https://doi.org/10.1016/j.cell.2021.10.002.

53. Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. Annu. Rev. Biochem. *74*, 681–710. https://doi.org/10.1146/annurev.biochem.74.082803.133243.

54. Eckner, R., Ewen, M.E., Newsome, D., Gerdes, M., DeCaprio, J.A., Lawrence, J.B., and Livingston, D.M. (1994). Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes Dev. *8*, 869–884.

55. Kang, J.-Y., Kim, J.-Y., Kim, K.-B., Park, J.W., Cho, H., Hahm, J.Y., Chae, Y.-C., Kim, D., Kook, H., Rhee, S., et al. (2018). KDM2B is a histone H3K79 demethylase and induces transcriptional repression via sirtuin-1-mediated chromatin silencing. FASEB J. *32*, 5737–5750. https://doi.org/10.1096/fj.201800242R.

56. Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Basavapathruni, A., Jin, L., Boriack-Sjodin, P.A., Allain, C.J., Klaus, C.R., Raimondi, A., Scott, M.P., et al. (2013). Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. Blood *122*, 1017–1025. https://doi.org/10.1182/blood-2013-04-497644.

57. Unk, I., Hajdú, I., Fátyol, K., Hurwitz, J., Yoon, J.-H., Prakash, L., Prakash, S., and Haracska, L. (2008). Human HLTF functions as a ubiquitin ligase for proliferating cell nuclear antigen polyubiquitination. Proc. Natl. Acad. Sci. USA *105*, 3768–3773. https://doi.org/10.1073/pnas.0800563105.

58. Nuñez, J.K., Chen, J., Pommier, G.C., Cogan, J.Z., Replogle, J.M., Adriaens, C., Ramadoss, G.N., Shi, Q., Hung, K.L., Samelson, A.J., et al. (2021). Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. Cell *184*, 2503–2519.e17. https://doi.org/10.1016/j.cell.2021.03.025.

59. Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588. https://doi.org/10.1038/nature14136.

60. Sockolosky, J.T., Trotta, E., Parisi, G., Picton, L., Su, L.L., Le, A.C., Chhabra, A., Silveria, S.L., George, B.M., King, I.C., et al. (2018). Selective targeting of engineered T cells using orthogonal IL-2 cytokine-receptor complexes. Science *359*, 1037–1042. https://doi.org/10.1126/science.aar3246.

61. Tian, S., Choi, W.-T., Liu, D., Pesavento, J., Wang, Y., An, J., Sodroski, J.G., and Huang, Z. (2005). Distinct functional sites for human immunodeficiency virus type 1 and stromal cell-derived factor 1alpha on CXCR4 transmembrane helical domains. J. Virol. *79*, 12667–12673. https://doi.org/10.1128/JVI.79.20.12667-12673.2005.

62. Thress, K.S., Paweletz, C.P., Felip, E., Cho, B.C., Stetson, D., Dougherty, B., Lai, Z., Markovets, A., Vivancos, A., Kuang, Y., et al. (2015). Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. Nat. Med. *21*, 560–562. https://doi.org/10.1038/nm.3854.

63. Chardon, F.M., McDiarmid, T.A., Page, N.F., Martin, B.K., Domcke, S., Regalado, S.G., Lalanne, J.-B., Calderon, D., Starita, L.M., Sanders, S.J., et al. (2023). Multiplex, single-cell CRISPRa screening for cell type specific regulatory elements. Preprint at bioRxiv. https://doi.org/10.1101/2023.03.28.534017.

64. Tian, R., Abarientos, A., Hong, J., Hashemi, S.H., Yan, R., Dräger, N., Leng, K., Nalls, M.A., Singleton, A.B., Xu, K., et al. (2021). Genome-wide CRISPRi/a screens in human neurons link lysosomal failure to ferroptosis. Nat. Neurosci. *24*, 1020–1034. https://doi.org/10.1038/s41593-021-00862-0.

65. Lin, J.-R., Zeman, M.K., Chen, J.-Y., Yee, M.-C., and Cimprich, K.A. (2011). SHPRH and HLTF act in a damage-specific manner to coordinate

Cell
Article

different forms of postreplication repair and prevent mutagenesis. Mol. Cell *42*, 237–249. https://doi.org/10.1016/j.molcel.2011.02.026.

66. Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive *piggyBac* transposase for mammalian applications. Proc. Natl. Acad. Sci. USA *108*, 1531–1536. https://doi.org/10.1073/pnas.1008322108.

67. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

68. Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., et al. (2012). BEDOPS: high-performance genomic feature operations. Bioinformatics *28*, 1919–1920. https://doi.org/10.1093/bioinformatics/bts277.

69. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

70. Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. *48*, 443–453. https://doi.org/10.1016/0022-2836(70)90057-4.

71. Kruskal, J.B. (1983). An overview of sequence comparison: time warps, string edits, and macromolecules. SIAM Rev. *25*, 201–237. https://doi.org/10.1137/1025045.

72. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

73. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419. https://doi.org/10.1038/nmeth.4197.

74. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

75. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

76. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118. https://doi.org/10.1371/journal.pcbi.1003118.

77. Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383. https://doi.org/10.1093/bioinformatics/btv145.

78. Wang, Q., Li, M., Wu, T., Zhan, L., Li, L., Chen, M., Xie, W., Xie, Z., Hu, E., Xu, S., et al. (2022). Exploring epigenomic datasets by ChIPseeker. Curr. Protoc. *2*, e585. https://doi.org/10.1002/cpz1.585.

79. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

80. Akalin, A., Franke, V., Vlahoviček, K., Mason, C.E., and Schübeler, D. (2015). Genomation: a toolkit to summarize, annotate and visualize genomic intervals. Bioinformatics *31*, 1127–1129. https://doi.org/10.1093/bioinformatics/btu775.

81. Putri, G.H., Anders, S., Pyl, P.T., Pimanda, J.E., and Zanini, F. (2022). Analysing high-throughput sequencing data in Python with HTSeq 2.0. Bioinformatics *38*, 2943–2945. https://doi.org/10.1093/bioinformatics/btac166.

82. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

83. Cribari-Neto, F., and Zeileis, A. (2010). Beta regression inR. J. Stat. Softw. *34*, 1–24. https://doi.org/10.18637/jss.v034.i02.

84. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

85. Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190. https://doi.org/10.1101/gr.849004.

86. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

87. Hahne, F., and Ivanek, R. (2016). Visualizing genomic data using Gviz and bioconductor. In Statistical Genomics: Methods and Protocols, E. Mathé and S. Davis, eds. (Springer), pp. 335–351. https://doi.org/10.1007/978-1-4939-3578-9_16.

88. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

89. McKenna, A., and Shendure, J. (2018). FlashFry: a fast and flexible tool for large-scale CRISPR target design. BMC Biol. *16*, 74. https://doi.org/10.1186/s12915-018-0545-0.

90. Schmidt, H., Zhang, M., Mourelatos, H., Sánchez-Rivera, F.J., Lowe, S.W., Ventura, A., Leslie, C.S., and Pritykin, Y. (2022). Genome-wide CRISPR guide RNA design and specificity analysis with GuideScan2. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.02.490368.

91. Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784. https://doi.org/10.1038/nmeth.3047.

92. Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.-C., Ogrodowski, P., Crippa, A., Rekhtman, N., de Stanchina, E., et al. (2014). In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature *516*, 423–427. https://doi.org/10.1038/nature13902.

93. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308. https://doi.org/10.1038/nprot.2013.143.

94. Tsai, B.P., Wang, X., Huang, L., and Waterman, M.L. (2011). Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach. Mol. Cell. Proteomics *10*, M110.007385. https://doi.org/10.1074/mcp.M110.007385.

95. Escobar, M., Li, J., Patel, A., Liu, S., Xu, Q., and Hilton, I.B. (2022). Quantification of genome editing and transcriptional control capabilities reveals hierarchies among diverse CRISPR/Cas systems in human cells. ACS Synth. Biol. *11*, 3239–3250. https://doi.org/10.1021/acssynbio.2c00156.

96. Kremers, G.-J., Hazelwood, K.L., Murphy, C.S., Davidson, M.W., and Piston, D.W. (2009). Photoconversion in orange and red fluorescent proteins. Nat. Methods *6*, 355–358. https://doi.org/10.1038/nmeth.1319.

97. Chen, J.J., Nathaniel, D.L., Raghavan, P., Nelson, M., Tian, R., Tse, E., Hong, J.Y., See, S.K., Mok, S.-A., Hein, M.Y., et al. (2019). Compromised function of the ESCRT pathway promotes endolysosomal escape of tau seeds and propagation of tau aggregation. J. Biol. Chem. *294*, 18952–18966. https://doi.org/10.1074/jbc.RA119.009432.

98. ENCODE Project Consortium, Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699–710. https://doi.org/10.1038/s41586-020-2493-4.

99. Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C., Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T., et al. (2016). ENCODE data at the ENCODE portal. Nucleic Acids Res. *44*, D726–D732. https://doi.org/10.1093/nar/gkv1160.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-β-actin	Sigma-Aldrich	Cat# A2228; Clone: AC-74; RRID: AB_476697
Donkey anti-Rabbit IgG polyclonal antibody (IRDye 800CW)	LI-COR Biosciences	Cat# 926-32213; RRID: AB_621848
Donkey anti-Mouse IgG polyclonal antibody (IRDye 680CW)	LI-COR Biosciences	Cat# 926-68072; RRID: AB_10953628
Rabbit monoclonal anti-HLTF (E9H5I)	Cell Signaling Technology	Cat# 45965; RRID: AB_3094661
Rabbit monoclonal anti-PMS2 (E9U4P)	Cell Signaling Technology	Cat# 27884; RRID: AB_3094662
Mouse monoclonal anti-MLH1 (4C9C7)	Cell Signaling Technology	Cat# 3515; RRID: AB_2145615
Bacterial and virus strains		
NEB Stable Competent *E. coli*	New England Biolabs	Cat# C3040H
NEB 10-beta Electrocompetent *E. coli*	New England Biolabs	Cat# C3020K
Chemicals, peptides, and recombinant proteins		
RPMI 1640 Medium	Gibco	Cat# 11875119
DMEM, High Glucose	Gibco	Cat# 11965118
FBS	Hyclone	Cat# SH30396.03
Penicillin-Streptomycin (10,000 U/mL)	Gibco	Cat# 15140122
mTeSR™ Plus kit	STEMCELL technologies	Cat# 100-0276
Y-27632	Stemgent	Cat# 04-0012-02
Geltrex™	Gibco	Cat# A1413302
Trimethoprim	Sigma-Aldrich	Cat# 92131
DNeasy Blood & Tissue Kit	QIAGEN	Cat# 69504
HiScribe™ T7 High Yield RNA Synthesis Kit	New England Biolabs	Cat# E2040S
HiScribe® T7 Quick High Yield RNA Synthesis Kit	New England Biolabs	Cat# E2050
RNase-Free DNase Set	QIAGEN	Cat# 79254
TURBO™ DNase	Invitrogen	Cat# AM2238
TRIzol LS Reagent	Invitrogen	Cat# 10296010
RNeasy Mini Kit	QIAGEN	Cat# 74104
Glycogen (5 mg/mL)	Invitrogen	Cat# AM9510
SuperScript™ IV Reverse Transcriptase	Invitrogen	Cat# 18091050
RNaseOUT™ Recombinant Ribonuclease Inhibitor	Invitrogen	Cat# 10777019
Proteinase K	Thermo Scientific	Cat# EO0491
AMPure XP Reagent	Beckman Coulter Life Sciences	Cat# A63882
KAPA2G Robust HotStart ReadyMix PCR Kit	Roche	Cat# KK5702
Power SYBR™ Green PCR Master Mix	Applied Biosystems	Cat# 4367659
NEBuilder® HiFi DNA Assembly Cloning Kit	New England Biolabs	Cat# E5520S
NEBridge® Golden Gate Assembly Kit (BsmBI- v2)	New England Biolabs	Cat# E1602S
NEBridge® Golden Gate Assembly Kit (BsaI-HF® v2)	New England Biolabs	Cat# E1601S
BsmBI-v2	New England Biolabs	Cat# R0739S
BsaI-HF®v2	New England Biolabs	Cat# R3733S
FastDigest Esp3I (IIs class)	Thermo Scientific	Cat# FD0454
I-SceI	New England Biolabs	Cat# R0694S
XhoI	New England Biolabs	Cat# R0146S
EcoRI-HF	New England Biolabs	Cat# R3101S

(Continued on next page)

Cell
Article

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BamHI-HF	New England Biolabs	Cat# R3136S
XbaI	New England Biolabs	Cat# R0145S
NotI-HF	New England Biolabs	Cat# R3189S
T4 DNA ligase	New England Biolabs	Cat# M0202L
SF Cell Line 4D-Nucleofector™ X Kit	Lonza Bioscience	Cat# V4XC-2012
SF Cell Line 96-well Nucleofector™ Kit	Lonza Bioscience	Cat# V4SC-2096
P3 Primary Cell 96-well Nucleofector™ Kit	Lonza Bioscience	Cat# V4SP-3096
ViraPower™ Lentiviral Packaging Mix	Invitrogen	Cat# K497500
Lipofectamine™ 3000 Transfection Reagent	Invitrogen	Cat# L3000001
PEG-it™ Virus Precipitation Solution (5×)	System Biosciences	Cat# LV810A-1
Polybrene	Millipore	Cat# TR-1003-G
Geneticin	Gibco	Cat# 10131035
Blasticidin S HCl (10 mg/mL)	Gibco	Cat# A1113903
DEPC (Diethyl Pyrocarbonate)	Millipore Sigma	Cat# D5758-25ML
YOYO™-1 Iodide (491/509) - 1 mM Solution in DMSO	Invitrogen	Cat# Y3601
dNTP mix	New England Biolabs	Cat# N0447L
NEBNext® Ultra™ II Non-Directional RNA Second Strand Synthesis Module	New England Biolabs	Cat# E6111L
Protease	QIAGEN	Cat# 19157
Buffer EB	QIAGEN	Cat319086
Tn5 transposase	Diagenode	Cat# C01070010-20
BSA	New England Biolabs	Cat# B9000S
Monarch® DNA Gel Extraction Kit	New England Biolabs	Cat# T1020S
NEBNext® High-Fidelity 2X PCR Master Mix	New England Biolabs	Cat# M0541L
OneTaq® Hot Start 2X Master Mix with Standard Buffer	New England Biolabs	Cat# M0484L
SYBR™ Green I Nucleic Acid Gel Stain	Invitrogen	Cat# S7563
Alt-R™ S.p. Cas9 Nuclease V3	Integrated DNA Technologies	Cat# 1081059
Alt-R® Cas9 Electroporation Enhancer, 2 nmol	Integrated DNA Technologies	Cat# 1075915
Critical commercial assays		
TruSeq RNA Library Prep Kit v2	Illumina	Cat# RS-122-2001
TruSeq Stranded mRNA kit	Illumina	Cat# 20020594
Illumina TruSeq RNA UD Indexes	Illumina	Cat# 20023785
NextSeq 1000/2000 P2 Reagents (100 Cycles) v3	Illumina	Cat# 20046811
NextSeq 500/550 Mid Output Kit v2.5 (150 Cycles)	Illumina	Cat# 20024904
MiSeq Reagent Kit v2 (300-cycles)	Illumina	Cat# MS-102-2002
MiSeq Reagent Kit v3 (150-cycle)	Illumina	Cat# MS-102-3001
MiSeq Reagent Kit v2 (50-cycles)	Illumina	Cat# MS-102-2001
Tagment DNA TDE1 Enzyme	Illumina	Cat# 20034197
Deposited data		
Sequencing data generated in this study	This manuscript	GEO: GSE228465
K562 DHS	ENCODE Project Consortium[34]	GEO:GSM816655 and ENCFF972GVB
K562 H3K79me2 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733653 and ENCFF957YJT
K562 CTCF ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM935407 and ENCFF682MFJ
K562 POLR2A ChIP-seq	ENCODE Project Consortium[34]	GEO:GSE91721 and ENCFF806LCJ
K562 ATAC-seq	ENCODE Project Consortium[34]	GEO:GSE170378 and ENCFF093IIW
K562 H3K9ac ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733778 and ENCFF286WRJ
K562 H3K9me3 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733776 and ENCFF601JGK
K562 H3K9me1 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733777 and ENCFF654SLZ

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
K562 H4K20me1 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733675 and ENCFF605FAF
K562 BRD4 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSE101225 and ENCFF251SRH
K562 EZH2 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM1003576 and ENCFF587SWK
K562 H2AFZ ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733786 and ENCFF621DJP
K562 POLR2AS2 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM935402 and ENCFF434PYZ
K562 SMC3 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM935310 and ENCFF469OWD
K562 HDAC1 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSE10583 and ENCFF684RNO
K562 HDAC2 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSE91451 and ENCFF954LGE
K562 HDAC3 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSE127356 and ENCFF975DCO
K562 H3K4me3 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSE96303 and ENCFF253TOF
K562 H3K4me2 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733651 and ENCFF959YJV
K562 H3K4me1 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733692 and ENCFF834SEY
K562 H3K27me3 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733658 and ENCFF405HIO
K562 H3K27ac ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733656 and ENCFF849TDM
K562 H3K36me3 ChIP-seq	ENCODE Project Consortium[34]	GEO:GSM733714 and ENCFF163NTH
Experimental models: Cell lines		
K562	ATCC	CCL-243
K562 PE2-Puro	Choi et al.[2]	N/A
K562 dCas9-VP64	Chardon et al.[63]	N/A
K562 PEmax	This manuscript	N/A
WTC11 DHFR-dCas9-VPH	Tian et al.[64]	N/A
Oligonucleotides		
List of oligonucleotides	This manuscript, Table S1	N/A
Recombinant DNA		
PB-T7-HEK3-BC (synHEK3)	This manuscript	N/A
LT3-GFP-T7-miR-E-CS1-PGK-Neomycin	This manuscript	N/A
Lenti-rtTA-P2A-Blast	This manuscript	N/A
pU6-Sp-pegRNA-HEK3-ins6N	This manuscript	N/A
pU6-Sp-pegRNA-HEK3-ins3N	This manuscript	N/A
pU6-Sp-dual-gRNA	This manuscript	N/A
pU6-Sp-gRNA-2XMS2	This manuscript	N/A
PB-CMV-MCP-XTEN80-p65-Rta-3xNLS-P2A-T2A-mPlum	This manuscript	N/A
PB-CMV-PEmax-EF1a-Puro	This manuscript	N/A
PB-UCOE-EF1a-PEmax-P2A-mCherry-PGK-Blast	This manuscript	N/A
PB-CMV-MCS-EF1α-Puro	System Biosciences	Cat# PB510B-1
pCMV-HyPBase	Yusa et al.[66]	N/A
PB-CMV-PE2-EF1a-Puro	Choi et al.[2]	N/A
pU6-Sp-pegRNA-HEK3-insCTT	Anzalone et al.[1]	Addgene: 132778
pCMV-PEmax	Chen et al.[13]	Addgene: 174820
pCMV-PEmax-P2A-hMLH1dn	Chen et al.[13]	Addgene: 174828
CRISPRoff-v2.1	Nuñez et al.[58]	Addgene: 167981
LT3GEN	Fellmann et al.[50]	Addgene: 111173
Software and algorithms		
Bcl2fastq (v2.20)	Illumina	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
Samtools (v1.9)	Danecek et al.[67]	http://www.htslib.org

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bedops (v2.4.35)	Neph et al.[68]	https://bedops.readthedocs.io/en/latest/index.html
featureCounts	Liao et al.[69]	https://subread.sourceforge.net/featureCounts.html
Needleall	Needleman and Wunsch[70]; Kruskal[71]	https://emboss.sourceforge.net/apps/release/6.6/emboss/apps/needleall.html
STAR (v2.7.6a)	Dobin et al.[72]	https://github.com/alexdobin/STAR
Salmon (v1.9.0)	Patro et al.[73]	https://salmon.readthedocs.io/en/latest/
Cutadapt (v4.1)	Martin[74]	https://cutadapt.readthedocs.io/en/stable/
TrimGalore (v0.6.6)	Martin[74]	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
Bwa (v0.7.17)	Li and Durbin[75]	https://bio-bwa.sourceforge.net/
GenomicRanges	Lawrence et al.[76]	https://bioconductor.org/packages/release/bioc/html/GenomicRanges.html
ChIPseeker	Yu et al.[77]; Wang et al.[78]	https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html
Homer	Heinz et al.[79]	http://homer.ucsd.edu/homer/
Genomation	Akalin et al.[80]	https://bioconductor.org/packages/release/bioc/html/genomation.html
HTseq (v.2.0.2)	Putri et al.[81]	https://htseq.readthedocs.io/en/master/
Seurat (v4.0.0)	Hao et al.[82]	https://satijalab.org/seurat/
Hmisc (v.5.1-1)	Harrell and Dupont	https://cran.r-project.org/web/packages/Hmisc/index.html
Betareg (v3.1.4)	Cribari-Neto and Zeileis[83]	https://cran.r-project.org/web/packages/betareg/index.html
DESeq2	Love et al.[84]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
WebLogo 3	Crooks et al.[85]	https://weblogo.threeplusone.com/
ComplexHeatmap	Gu et al.[86]	https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
Gviz	Hahne and Ivanek[87]	https://bioconductor.org/packages/release/bioc/html/Gviz.html
IGV	Robinson et al.[88]	http://software.broadinstitute.org/software/igv/
FlashFry	McKenna and Shendure[89]	https://github.com/mckennalab/FlashFry
GuideScan2	Schmidt et al.[90]	https://www.guidescan.com/
DeepPrime	Yu et al.[19]	https://deepcrispr.info/DeepPrime/
Other		
UCSC trackhub for visualizing prime editing results in synHEK3 reporters and surrounding epigenetic environments	This manuscript	https://shendure-web.gs.washington.edu/content/members/xyli10/public/nobackup/hub.txt

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jay Shendure (shendure@uw.edu).

Materials availability
All unique reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

- All sequencing data have been deposited at GEO (GSE228465) and are publicly available. This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. The UCSC trackhub created for visualizing results of this work is: https://shendure-web.gs.washington.edu/content/members/xyli10/public/nobackup/hub.txt.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines and cell culture

K562 cells (CCL-243, female) were purchased from ATCC. PE2,[2] PEmax and dCas9-VP64[63] K562 lines were generated previously or in this study. All K562 cells were maintained in RPMI 1640 medium (Gibco) supplied with 10% FBS (Hyclone) and penicillin/streptomycin (Gibco, 100 U/ mL). HEK293T cells were maintained in DMEM medium (Gibco) supplied with 10% FBS and penicillin/streptomycin. The DHFR-dCas9-VPH (VP48-P65-HSF1) WTC11 (human iPSCs, male) line was generated and kindly provided by the Martin Kampmann lab.[64] iPSCs were maintained on Geltrex (Gibco) coated plates in the mTeSR Plus medium (STEMCELL Technologies) supplied with 10 μM Y-27632 (Stemgent). All cells were kept in a humidified incubator at 37 °C, 5% CO_2.

METHOD DETAILS

Molecular cloning
PB-T7-HEK3-BC (synHEK3)

First, a minimal *piggyBac* cargo construct was created by deleting all intervening sequences between the 5' and 3' terminal repeats (including core insulators) of the PB-CMV-MCS-EF1α-Puro vector (System Biosciences, PB510B-1). A gBlock (Integrated DNA Technologies, IDT) consisting of a filler sequence and flanking scaffold sequences (from GFP) was inserted to create a shuttle vector. The filler sequence contains two divergent BsmBI recognition sites and can be removed scarlessly. Second, the shuttle vector was digested with BsmBI (New England Biolabs). An 87-bp region around the HEK3 gRNA target site was synthesized from IDT and amplified with a pair of primers to introduce a T7 promoter and a 16-bp barcode to its upstream and downstream, respectively. The resulting PCR product was inserted into the linearized shuttle vector using NEB HiFi assembly. Ligated products were cleaned up and concentrated with AMPure XP beads (Beckman Coulter), and electroporated into NEB 10-beta electrocompetent *E. coli*. Electroporation was performed in a 0.1 cm electroporation cuvette using a Bio-Rad GenePulser electroporator at 2.0 kV, 200 Omega, and 25 μF.

LT3-GFP-T7-miR-E-CS1-PGK-Neomycin

The LT3GEN vector was purchased from Addgene (#111173)[50] and digested with I-SceI (New England Biolabs). A fragment containing a T7 promoter and homologous sequences was ordered from IDT and assembled into the backbone. The LT3-GFP-T7-PGK-Neomycin vector was digested with XhoI and EcoRI-HF (New England Biolabs). An shRNA targeting the Renilla luciferase (Ren713) was inserted along with a Capture Sequence 1 (CS1) after the EcoRI site. For shRNA cloning, the LT3-GFP-T7-miR-E-CS1-PGK-Neomycin construct was digested with XhoI and EcoRI-HF. shRNAs were ordered from IDT as 97-nt 4 nmole Ultramers (Table S1) or oPool (Table S5) and amplified with primers p1 (5'-ATTACTTCGACTTCTTAACCCAACAGAAGGCTCGAGAAGGTA TATTGCTGTTGACAGTGAGCG-3') and p2 (5'-AATTGCTCTTGCTAGGACCGGCCTTAAAGCGAATTCTAGCCCCTTGAAGTCCGA GGCAGTAGGCA-3'). The PCR products were assembled into the backbone using NEB HiFi assembly and transformed into NEB Stable Competent *E. coli* (single shRNA vectors) or electroporated into NEB 10-beta electrocompetent *E. coli* (library).

Lenti-rtTA-P2A-Blast

The Lenti-Cas9-P2A-Blast (Addgene: 52962) vector[91] was digested with XbaI and BamHI (New England Biolabs) to create a backbone. rtTA was amplified from the LT3GEPIR vector (Addgene: 111177)[50] and cloned into the backbone using NEB HiFi assembly.

pU6-Sp-pegRNA-HEK3-ins6N and pU6-Sp-pegRNA-HEK3-ins3N

the pU6-Sp-pegRNA-HEK3-insCTT vector (Addgene:132778)[1] was linearized by PCR with 5' phosphorylated oligos p3 (5'-TCTGCC ATCANNNNNNCGTGCTCAGTCTGTTTTTTTTAAGCTTG-3', ins6N) or p4 (5'-TCTGCCATCANNNCGTGCTCAGTCTGTTTTTTTTAAGC TTG-3', ins3N) and p5 (5'-GGACCGACTCGGTCCCACTT-3') and ligated with T4 DNA ligase. Ligation product was concentrated and electroporated into NEB 10-beta electrocompetent *E. coli*.

pU6-Sp-dual-gRNA vectors

A pU6-Sp-dual-gRNA scaffold vector was generated by replacing the pegRNA expressing cassette of pU6-Sp-pegRNA-HEK3-insCTT vector with a dual U6-gRNA cassette from the PX333 vector (Addgene: 64073).[92] The second gRNA cloning site (two BsaI sites) was modified to two BsmBI sites. Spacer sequences were cloned into this vector sequentially using the oligo annealing method.[93]

pU6-Sp-gRNA-2XMS2 vectors

A pU6-Sp-gRNA-2XMS2 scaffold vector was generated by modifying the pU6-Sp-pegRNA-HEK3-insCTT backbone. The scaffold sequence is 5'-GTTTAAGAGCTAAGCCAACATGAGGATCACCCATGTCTGCAGGGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGGCCAACATGAGGATCACCCATGTCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTTTT-3'.[58] Spacer sequences were cloned in between two BsmBI sites using the oligo annealing method.[93]

PB-CMV-MCP-XTEN80-p65-Rta-3xNLS-P2A-T2A-mPlum

The MCP(N55K) sequence[94] was synthesized as an IDT gBlock and amplified. XTEN80 and 3XNLS-P2A were amplified from TETv4 (Addgene: 167983).[58] p65-Rta was amplified from sadCas9-VPR (Addgene: 188514).[95] mPlum was amplified from mPlum-C1(Addgene: 54839).[96] All PCR products were purified and cloned into the PB-CMV-MCS-EF1α-Puro vector between the NotI site and the SV40 polyA sequence using NEB HiFi assembly. Then, a T2A sequence was inserted next to P2A by NEB HiFi assembly.

PB-CMV-PEmax-EF1a-Puro

PE2max was amplified from pCMV-PEmax-P2A-hMLH1dn (Addgene: 174828)[13] and cloned into the PB-CMV-MCS-EF1α-Puro vector.

PB-UCOE-EF1a-PEmax-P2A-mCherry-PGK-Blast

The UCOE-EF1a was amplified from pMH0006 (Addgene: 135448).[97] PE2max was amplified from pCMV-PEmax-P2A-hMLH1dn. The rest sequences were synthesized as IDT gblocks and amplified. All PCR products were cloned into an empty *piggyBac* transposon vector.

(e)pegRNA plasmids

(e)pegRNAs (pU6-Sp-(e)pegRNA) used in this study were ordered as 4nm ultramers (IDT) or long primers containing the spacer and 3' extension sequences and cloned into the backbone of pU6-Sp-pegRNA-HEK3-CTT using NEB HiFi assembly. The epegRNA libraries were ordered as individual IDT eBlocks, pooled and cloned into the same backbone by Golden Gate assembly (New England Biolabs).

General transfection strategies

Transfections of K562 cells were performed using a Lonza Bioscience 4D-Nucleofector system and the SF Cell Line 4D-Nucleofector X kits (Lonza). For single nucleocuvettes (100 μL), 1-3 x 10^6 cells were transfected with up to 9 μg DNA. For 96-well Nucleocuvette plates (20 μL), 2 x 10^5 ~ 4 x 10^5 cells were transfected with up to 1.2 μg DNA. Program code was FF-120.

Transfections of WTC11 cells were performed using a Lonza Bioscience 4D-Nucleofector system and the P3 Primary Cell 96-well Nucleofector Kit (Lonza) . For each transfection, 2 x 10^5 cells were transfected with up to 3.5 μg DNA. Program code was CB-150.

Transfections of the HEK293T cell line were performed using Lipofectamine 3000 (Invitrogen) following manufacturer's instructions.

Cell line generation

Wild-type K562 cells were transfected with PB-UCOE-EF1a-PEmax-P2A-mCherry-PGK-Blast and pCMV-HyPBase[66] at a ratio of 3:1 in a 100 μL nucleofection reaction. Cells were selected with 10 μg/mL blasticidin (Gibco) for 7 days. Monoclonal lines were isolated and the clone with the brightest mCherry fluorescent signal was used for following experiments.

Prime editing experiment in synHEK3 reporters

3 x 10^6 PE2(+) or wild-type K562 cells were transfected with synHEK3 and the pCMV-HyPBase plasmid at a 3:1 ratio (total 9 μg) in a 100 μL nucleofection reaction. Cells were cultured in a large flask to maintain complexity. After 10 days, gDNA was extracted to estimate synHEK3 copy number by qPCR. Cells were counted and 500 cells were seeded to a new dish. gDNA was extracted from a portion of the expansion of these cells and used for T7-assisted reporter mapping (described below).

For measuring prime editing efficiency, 2 x 10^6 PE2(+) or wild-type K562 cells with integrated synHEK3 reporters from the bottlenecked 500-cell pools generated above were transfected with 2 μg pU6-Sp-pegRNA-HEK3-insCTT or 1 μg pU6-Sp-pegRNA-HEK3-insCTT with 1 μg PB-CMV-PE2-EF1a-Puro,[2] respectively. Wild-type K562 cells were selected with 2 μg/mL puromycin for 2 days 24 hours after transfection. Cells were lysed 4 days after transfection and subjected to amplicon sequencing (described below).

Prime editing experiment in endogenous sites

SpCas9 spacers of the pegRNAs were searched in a set of 1-kb windows 500-bp away from a different set of randomly integrated synHEK3 reporters in K562 cells (n = 3,634) using FlashFry.[89] gRNAs were filtered by specificity (specificity score Hsu 2013 > 80) and ranked by cutting efficiency (Doench 2014), and best gRNAs were selected. This resulted in 3,157 gRNAs. DeepPrime was used to predict best pegRNAs or epegRNAs for installing 3-bp insertions (CCT) at the cut sites (the HEK293T PE2max models). The predicted scores were highly correlated for each spacer between the pegRNA and epegRNA (Pearson's r = 0.83). Spacers scored highly in both pegRNA (>40%) and epegRNAs (>65%) were selected (n = 121). Eventually, the library was synthesized in the form of epegRNAs (with the evopreQ1 hairpin; Table S1).

The epegRNAs were grouped into 2 separate pools (N = 60 and 61, respectively) with 2% spike-in of the pU6-Sp-pegRNA-HEK3-insCTT vector. The abundance of pegRNAs were measured by sequencing of the plasmid libraries. The plasmid pools (2 μg) were

then transfected into 1 x 10^6 PEmax(+) K562 cells in duplicates. Editing rates were measured at the 121 sites and normalized using the abundance of the HEK3 plasmid and editing efficiency at the HEK3 loci in the individual pools (Table S4).

Cas9 RNP editing experiment
Alt-R sgRNAs were ordered from IDT and resuspended in TE buffer to 100 μM. The RNP complex was assembled by combining 2.1 μL PBS, 1.2 μL Alt-R sgRNA (100 μM) and 1.7 μL Alt-R Cas9 Nuclease (61 μM, IDT) and incubating at room temperature for 10-20 min. The RNP mixture was added to 1 x 10^6 PE2(+) K562 500-cell pool resuspended in nucleofection solution along with 0.5 μg pmax-GFP (Lonza) and 1 μL electroporation enhancer (100 μM, IDT). gDNA was extracted at Day 1, 2 and 4 after transfection for amplicon sequencing of synHEK3 reporter.

CRISPRoff experiment
On Day 0, 2 x 10^6 clone 5 cells were transfected with CRISPRoff-v2.1 (3 μg), pU6-Sp-dual-gRNA (1 μg) and pmax-GFP (500 ng, Lonza) using the SF Cell Line 4D-Nucleofector X kit L (Lonza). On Day 2, cells were sorted based on high GFP expression (top 20%) on a flow cytometer and expanded. On Day 11, 4 x 10^5 cells were transfected with the pU6-Sp-pegRNA-HEK3-insCTT plasmid (1 μg). Cells were lysed on Day 15 for amplicon sequencing of the synHEK3 reporter. A portion of cells were collected on Day 11 for total RNA extraction and RNA sequencing (described below).

CRISPRa experiments
K562 experiments
On Day 0, 4 x 10^5 K562 dCas9-VP64 cells were transfected with PB-CMV-MCP-XTEN80-p65-Rta-3xNLS-P2A-T2A-mPlum (600 ng) and paired pU6-Sp-gRNA-2XMS2 (200 ng each) plasmids targeting the same promoter. On Day 2, 4 x 10^5 cells from the previous transfection were transfected with pCMV-PEmax (Addgene: 174820)[13] or PB-CMV-PEmax-EF1a-Puro (800 ng) and the pU6-Sp-(e)pegRNA plasmids (400 ng). Cells transfected with PB-CMV-PEmax-EF1a-Puro were selected with 2 μg/mL puromycin 24 hours after transfection for 2 days. Cells were lysed at Day 5 or 6 for amplicon sequencing.
iPSCs experiments
On Day 0, 2 x 10^5 the DHFR-dCas9-VPH WTC11 cells were transfected with PB-CMV-MCP-XTEN80-p65-Rta-3xNLS-P2A-T2A-mPlum (2.1 μg) and paired pU6-Sp-gRNA-2XMS2 (700 ng each) plasmids targeting the same promoter. 20 μM Trimethoprim (Sigma-Aldrich) was added to induce dCas9-VPH. On Day 3, 2 x 10^5 cells from the previous transfection were transfected with PB-CMV-PEmax-EF1a-Puro (2 μg) and the pU6-Sp-(e)pegRNA plasmids (1 μg). Cells were continuously treated with 20 μM Trimethoprim and were selected with 2 μg/mL puromycin 24 hours after transfection for 1 day. Cells were lysed at Day 5 or 6 for amplicon sequencing.

Design and test of IL2RB epegRNA libraries
SpCas9 spacers for the selected exons of human *IL2RB* (Exons 1, 2, 4, 5, 8, 9 and 10) were generated using GuideScan2,[90] sorted by the number of off-target and cutting efficiency, and filtered for spacers with all of A, T, C, G bases within the first 7 base pairs from the cut sites (which allowed for modeling all 12 possible substitutions in close proximity of the cut sites). For substitutions, the first occurrences of the A, T, C, G bases were converted to the rest 3 bases. For 1-bp insertions, insertions reverse complement to the first base after the cut sites were modeled. For 3- and 6-bp insertions, insertions of CCT and CGTCAT were modeled at the cut sites. For long insertions, the 34-bp loxP sequence was inserted at the cut sites. For deletions, sequences of 1, 3 and 6 bp were deleted after the cut sites. epegRNAs for substitutions, and insertions and deletions ≤ 3 bps were designed using DeepPrime. The resulting epegRNAs for each exon had a median DeepPrime score > 40%. For 6-bp insertions and deletions, and insertions of loxP, the epegRNAs were designed manually with the length of primer binding sites being 13 bp and length of template (excluding the inserted or deleted sequences) being 10 bp. See Table S1 for sequences of the epegRNAs.

19 epegRNAs derived from the same spacer were synthesized as a pool (IDT oPool) and cloned into the pU6-Sp-pegRNA-HEK3-CTT backbone as a library. The 7 epegRNAs libraries were tested following the CRISPRa experiment procedure in the dCas9-VP64 K562 cells as described above.

Quantitative PCR analysis
qPCRs were performed on purified gDNA or cDNA reverse transcribed from total RNAs using SuperScript IV reverse transcriptase (200 U/μL, Invitrogen) following manufacturer's instructions. qPCRs were performed with Power SYBR Green PCR Master Mix (Invitrogen) or KAPA2G Robust HotStart ReadyMix (Roche) supplied with SYBR Green (Invitrogen). For copy number estimation of synHEK3 reporters, Cq values of synHEK3 were normalized to those of SNRPB (3 copies per genome). See Table S1 for the list of primers used.

T7-assisted reporter mapping
gDNA of K562 cells was purified using the DNeasy Blood & Tissue Kit (QIAGEN) and *in vitro* transcribed with the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs). Briefly, each reaction (20 μL) contained 0.3∼1 μg gDNA, NTPs (10 mM each) and 2 μl T7 RNA Polymerase Mix. The reaction mixture was incubated at 37 °C for 16 hours. Then, gDNA was digested with 2.5 μL DNase

Cell
Article

(QIAGEN) in a 100 μL reaction at room temperature for 30 min. RNA was extracted with TRIzol LS Reagent (Invitrogen), and aqueous phase was precipitated with 1 volume of isopropanol and 5 μg Glycogen (Invitrogen) at -80 °C for 1 hour. RNA pellet was collected by centrifugation at 21,000 x g at 4 °C for 1 hour. The pellet was washed with 80% ice-cold ethanol and resuspended in 11.5 μL nuclease-free water.

For reverse transcription, RNA was incubated with 0.5 μL 100 μM RT primer p6 (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGA CAGNNNNNNNNN-3') and 1 μL 10 mM dNTP at 65 °C for 5 min and cooled on ice. Then, 4 μL 5X RT buffer, 1 μL 100 mM DTT, 1 μL RNaseOUT (40 U/μL) and 1 μL SuperScript IV RT (200 U/μL) were added and the reaction mixture was incubated at 23 °C for 10 min, 50 °C for 15 min and 80 °C for 10 min.

cDNA library was amplified with KAPA2G Robust HotStart polymerase, using primers p7 (5'- GTGACTGGAGTTCAGACGTGTG CTCTTCCGATCTGAAAGGAAGCCCTGCTTCCTCCAGAGGG-3', 0.5 μM) and p8 (5'- TCGTCGGCAGCGTCAGATGTGTATAAGA GACAG-3', 0.5 μM). PCR reaction was performed as follows: 95 °C 3 min; 95 °C 15 s, 65 °C 15 s, 72 °C 30 s, 16~18 cycles; and 72 °C 1 min. The PCR product was subjected to double-sided size selection (0.5X, 0.9X) and cleaned up with AMPure XP beads. The resulting product ranged from 200 to 1000 bp. To prepare Illumina sequencing libraries, 5-10 ng PCR product was re-amplified with the Nextera P5 and TruSeq P7 library index primers as shown in Table S1 for 5 cycles. The final PCR product underwent another round of double-sided size selection (0.5X, 0.9X) and clean-up with AMPure XP beads. The library was sequenced on an Illumina MiSeq in paired-end mode (Read 1: 254 bp; Read2: 55bp).

synHEK3 editing library preparation

gDNA of K562 cells was purified using the DNeasy Blood & Tissue Kit. Alternatively, cells were lysed in a lysis buffer [10 mM Tris-HCl pH8.0, 0.05% SDS and 0.04 mg/mL proteinase K(Thermo Scientific)], and incubated at 50 °C 60 min and 80 °C 30 min. 100~250 ng gDNA or cell lysates were amplified using KAPA2G Robust HotStart ReadyMix with primers p9 (5'-GTGACTGGAGTTCA GACGTGTGCTCTTCCGATCTCTACCCCGACCACATGAAGCAGC-3', 0.5 μM) and p10 (5'-TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNNNNNNGACCATGTCATCGCGCTTCTCGT-3', 0.5 μM). PCR reaction was performed as follows: 95 °C 3 min; 95 °C 15 s, 68-N °C 15 s, 72 °C 30 s, for 9 cycles (N was cycle number); 95 °C 15 s, 65 °C 15 s, 72 °C 30 s, for 11 cycles; and 72 °C 1 min. To ensure enough coverage and accurate measurement of editing efficiencies, for the K562 synHEK3 pool, we pooled products from at least 16 PCR reactions. The PCR product was purified with AMPure XP beads. 5 ng PCR product was re-amplified with the Nextera P5 and TruSeq P7 library index primers in Table S1 for 5 cycles. The final libraries were cleaned up with AMPure XP beads and sequenced on an Illumina NextSeq 500 or an Illumina NextSeq 2000 sequencer.

Lentivirus production and transduction

1.6×10^7 HEK293T cells were seeded in a 10 cm² dish the day before transfection. 5.6 μg lentiviral vector and 14.4 μg ViraPower Lentiviral Packaging Mix (Invitrogen) were mixed and transfected using Lipofectamine 3000 following manufacturer's instructions. Medium was changed at 24 hours post transfection. Viruses were collected at 48 and 72 hours post transfection, filtered using 45 μm filters and concentrated 100-fold using PEG-it Virus Precipitation Solution (System Biosciences).

In general, K562 cells were transduced with lentivirus in the presence of 8 μg/mL polybrene (Millipore). Medium was replaced after 24 hours.

Pooled shRNA screen

To prepare for the pooled shRNA screening, two monoclonal lines (clone 3 and clone 5) from the original 500-cell synHEK3 pool were isolated and genotyped. There were 22 and 28 unique synHEK3 reporters in these two clones. These cells were transduced with Lenti-rtTA-P2A-Blast viruses and selected in 10 μg/mL blasticidin for 7 days. rtTA (+) monoclonal lines were then generated for down-stream experiments.

The shRNA lentiviral library was first titrated, and mixed with clone 3 and 5 rtTA(+) cells at an MOI of 10 and a >1000X coverage. Transduced cells were selected in 800 μg/mL Geneticin (Invitrogen) for 7 days. 1×10^6 cells were treated with 1 μg/mL doxycycline for 2 days to induce shRNA expression. Then 1.5×10^6 - 2.5×10^6 cells were transfected with 4.5-6.0 μg pU6-Sp-pegRNA-HEK3-ins6N (plasmid to introduce 6-bp insertion in HEK3). After 3-4 days, cells from the two clones were collected and mixed at a 1:1 ratio for sci-RNA-seq3 library preparation.

The sci-RNA-seq3 libraries were prepared by modifying a recently published protocol from the lab.[49] Briefly, K562 cells were counted, washed with PBS and lysed in 5 mL Hypotonic lysis buffer B with DEPC (Sigma) for 10 min on ice. Nuclei were collected by centrifugation at 500 xg, 4 °C for 3 min and resuspended in 1 mL 0.3 M SPBSTM buffer with DEPC. Nuclei were fixed with 4 mL ice-cold methanol for 15 min on ice, swirled occasionally. After rehydration by adding 10 mL SPBSTM, nuclei were collected by centrifugation and washed once with 1 mL 0.3 M SPBSTM. For T7 in situ transcription, fixed nuclei resuspended in 171 μL SPBSTM were mixed with 99 μL NTP buffer and 30 μL T7 polymerase mix (New England Biolabs) and incubated in a 1.5 mL LoBind tube (Eppendorf) at 37 °C for 30 min. Afterwards, nuclei were washed with 1 mL SPBSTM, sonicated using a Diagenode sonicator for 12 seconds. At this point, nuclei were stained with YOYO-1 Iodide (Invitrogen) and counted on a Countess automated cell counter.

Counted nuclei were resuspended in SPBSTM at 4×10^6/mL. For each RT plate, 500 μL nuclei were combined with 56 μL 10 mM dNTPs (New England Biolabs) and distributed to a low-bind 96-well plate (5 μL per well). 1 μL indexed oligo-dT, HEK3 and CS1 primers (10 μM) were added to each well. The plate was incubated at 55 °C for 5 min and cooled on ice. RT mixture was prepared

by combining 240 µL 5X SuperScript IV Buffer, 60 µL SuperScript IV and 60 µL water. 3 µL RT mixture was added to each well. The RT plate was incubated at 55 °C for 10 min and cooled on ice. 5 µL ice-cold SPBSTM was added to each well and all nuclei were pooled in pre-chilled LoBind tubes. Nuclei were washed once with 1 mL SPBSTM and resuspended in 1,200 µL SPBSTM (per ligation plate).

11 µL nuclei were distributed to each well of a new 96-well plate and mixed with 2 µL 10 µM ligation primers. For each ligation plate, 195 µL 10X T4 Ligation Buffer was mixed with 65 µL T4 DNA Ligase, and 2 µL of the ligation mixture was added to each well. Ligation was performed at room temperature for 20 min on the bench. For the shRNA screen, nuclei were distributed into 4 ligation plates (384 ligation indices) to increase cell index complexity. The ligation plate was then cooled on ice and 10 µL ice-cold SPBSTM was added to pool nuclei from all wells. Nuclei were washed twice with 1 mL SPBSTM.

Nuclei were resuspended in 1X Second Strand Synthesis Buffer (New England Biolabs), counted with YOYO-1 Iodide and diluted to 1.3×10^5 - 2.5×10^5 per 400 µL. 4 µL nuclei were distributed to multiple 96-well plates based on the number of nuclei retrieved at this step. Extra plates with nuclei were stored at -80 °C. Second strand synthesis mixture was prepared by combining 10.5 µL 10X Second Strand Synthesis Buffer, 35 µL 20X Second Strand Synthesis Enzyme with 94.5 µL water. 1 µL of second strand synthesis mixture was added to each well. The plate was incubated at 16 °C for 2.5 hours. To lyse nuclei, 1 µL ~1.07 AU/mL protease (QIAGEN) was added to each well and incubated at 37 °C for 40 min. Protease was inactivated at 75 °C for 20 min. 5 µL EB buffer (QIAGEN) was added to each well and mixed. 5 µL was taken out from each well and used for enrichment PCR, while the rest was used for Tn5 tagmentation and transcriptome library preparation.

To prepare the transcriptomic library, Tn5-N7 and mosaic end oligos were resuspended to 100 µM in annealing buffer (50 mM NaCl, 40 mM Tris-HCl pH8.0), mixed at a 1:1 ratio and annealed on a thermocycler using the following program: 95°C 5 min, cool to 65°C (0.1°C/s), 65°C 5 min, cool to 4°C (0.1°C/s). 20 µL Tagmentase (Tn5 transposase - unloaded, Diagenode) was mixed with 20 µL annealed oligos and incubated on a thermomixer at 350 rpm, 23°C for 30 min. 20 µL glycerol was added to the loaded Tn5 before storage at -20 °C. For tagmentation, 13.75 µL N7-loaded Tn5 (Diagenode) was mixed with 550 µL TD buffer and 5 µL of this mixture was added to each well. The plate was incubated at 55 °C for 5 min. To remove Tn5 transposase, 50 µL 1% SDS, 50 µL BSA (New England Biolabs) and 225 µL water were mixed, and 2.6 µL of the mixture was added to each well and incubated at 55 °C for 15 min. Then, SDS was quenched by adding 2 µL 10% Tween-20 to each well. For PCR, 96 indexed P5 primers were used with constant or indexed P7 primers (see Table S1). A PCR master mixture was prepared by combining 2,200 µL 2x NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs), 22 µL common P7 primer (100 µM) and 352 µL water. 2 µL indexed TruSeq P5 primer and 23.4 µL PCR mixture were added to each well. If using indexed P7 primers, 2 µL of 10 µM primer was added to each well. PCR reaction was performed as follows: 70 °C 3 min; 98 °C 30 s; 98 °C 10 s, 63 °C 30 s, 72 °C 60 s, 16 cycles; and 72 °C 5 min. 3 µL of each well was pooled and cleaned up with 0.8X AMPure XP beads. The library was resolved on a 1% agarose gel and the smear between 300-600 bp was extracted using Monarch DNA Gel Extraction Kit (New England Biolabs).

For HEK3 and shRNA libraries, enrichment PCRs were directly performed on protease-treated cDNAs. To match PCR indices with those of the transcriptome library, the same indexed TruSeq P5 primers were used. The enrichment PCR master mixture contained 2,200 OneTaq 2X Master Mix (New England Biolabs), 16.5 µL synHEK3 P7 enrichment primer (100 µM), 16.5 µL shRNA P7 enrichment primer (100 µM), 44 µL 100X SYBR green (Invitrogen) and 1,353 µL water. 2 µL indexed P5 primer and 33 µL PCR mixture were added to each well. PCR reaction was performed as follows: 95 °C 3 min; 95 °C 15 s, 68-N °C 15 s, 72 °C 30 s, for 9 cycles (N was cycle number); 95 °C 15 s, 65 °C 15 s, 72 °C 30 s, for M cycles (decided by qPCR); and 72 °C 1 min. We monitored the reaction on a real-time qPCR machine and terminated the reaction at the log phase. All PCR products were pooled and concentrated using 0.9X AMPure XP beads. The library was resolved on a 1% agarose gel and the two discrete bands corresponding to HEK3 and shRNA constructs were extracted using Monarch DNA Gel Extraction Kit.

Eventually, all libraries were sequenced on an Illumina NextSeq 500 or an Illumina NextSeq 2000 sequencer. For HEK3 and shRNA libraries, custom Index 1 primers and Read 2 primers were used. Usually, 34 cycles were allocated to Read 1 for reading the RT index, ligation index, and unique molecular identifier (UMI). See Table S1 for a list of oligos and sequencing primers used in this experiment.

Amplicon sequencing

Cells were lysed in a lysis buffer (10 mM Tris-HCl pH8.0, 0.05% SDS and 0.04 mg/mL proteinase K), and incubated at 50 °C 60 min and 80 °C 30 min. Lysates were directly used for PCR with the KAPA2G Robust HotStart ReadyMix with primers (0.5 µM each) designed for the endogenous targets. PCR reaction was performed as follows: 95 °C 3 min; 95 °C 15 s, 66-N °C 15 s, 72 °C 40 s, for 8 cycles (N was cycle number); 95 °C 15 s, 60 °C 15 s, 72 °C 40 s, for M cycles (decided by qPCR); and 72 °C 1 min. The PCR product was purified with AMPure XP beads. 5 ng PCR product was re-amplified with the Nextera P5 and TruSeq P7 library index primers in Table S1 for 5 cycles. The final libraries were cleaned up with AMPure XP beads and sequenced on an Illumina Miseq, an Illumina NextSeq 500 or an Illumina NextSeq 2000 sequencer.

Sequencing reads were demultiplexed using the bcl2fastq software (Illumina). A custom script was used to determine editing efficiency.

RNA sequencing

For bulk RNA sequencing of K562 PE2(+) cells and EPZ-5676 treated cells, total RNAs were purified from cells using TRIzol LS Reagent following manufacturer's instructions, treated with RNase-free DNase (QIAGEN) and cleaned up using RNeasy Mini kit (QIAGEN). 1 µg total RNA (RNA integrity >= 9.7) were used for unstranded mRNA library preparation using the Illumina TruSeq

RNA Library Prep Kit v2. For the CRISPRoff experiment and the HLTF knockdown experiment, total RNAs were purified from cells using TRIzol LS Reagent, treated with Turbo DNase (Invitrogen) and cleaned up using the RNeasy Mini kit. 1 μg total RNA (RNA integrity >= 9.9) were used for stranded mRNA library preparation using the Illumina TruSeq Stranded mRNA Library Prep kit. The RNA libraries were indexed using the Illumina TruSeq RNA UD Indexes. All libraries were sequenced on an Illumina NextSeq 500 or an Illumina NextSeq 2000 sequencer in a paired end mode.

Sequencing reads were demultiplexed using the bcl2fastq software. For K562 PE2(+) and EPZ-5676 treated cells, an average of 34 million 75-bp paired-end reads were obtained. For the CRISPRoff experiment, on average, 55 million 50-bp paired-end reads were obtained per sample. For the HLTF knockdown experiment, 20 million 59-bp paired-end reads were obtained per sample. For calculating TPM of genes, sequencing reads from the unstranded RNA libraries were aligned to the GRCh38 reference genome (Gencode V43) using Salmon (v1.9.0).[73] Reads were aligned to the GRCh38 reference genome and counted against all Ensembl genes using STAR (2.7.6a).[72] Raw counts were analyzed with DESeq2.[84]

ATAC sequencing

1×10^5 cells were collected, washed with PBS, and pelleted by centrifugation at 500 xg, 4 °C for 5 min. Cells were then lysed with 50 μL lysis buffer (10 mM Tris-HCl pH7.4, 10 mM NaCl, 3 mM MgCl$_2$, 0.1% NP-40, 0.1% Tween-20, 0.01% Digitonin) on ice for 3 min and neutralize with 250 μL RSB buffer with Tween-20 (10 mM Tris-HCl pH7.4, 10 mM NaCl, 3 mM MgCl$_2$, 0.1% Tween-20). Nuclei were counted on a hemocytometer and 5×10^4 nuclei underwent tagmentation. For tagmentation, per 50 μL contained 25 μL 2x TD Buffer, 8.25 μL PBS, 0.5 μL 1% Digitonin, 0.5 μL 10% Tween-20, 2.5 μL Tn5 enzyme (Illumina, 2.5 μM) and 13.25 μL water. The reaction was incubated at 37°C for 30 min. DNA was purified using the Clean and Concentrate kit (Zymo) and eluted in 10–20 μL EB Buffer (QIAGEN). All or half of the eluted DNA was used for qPCR. In each reaction, 10 μL tagmented DNA was mixed with 25 μL 2x NEBNext High-Fidelity 2X PCR Master Mix (NEB), 2.5 μL i5 primer, 2.5 μL i7 primer (Table S1), 0.25 μL 100X SYBR Green (Invitrogen) and 9.75 μL water. The qPCR reaction was performed as follows: 72°C 5 min, 98 °C 30 s; 98 °C 10 s, 63 °C 30 s, 72 °C 1 min. The reaction was stopped after 7 cycles. Libraries were pooled, purified with AMPure XP beads, and sequenced on an Illumina NextSeq 2000 sequencer.

On average, 34 million 59-bp paired-end reads were obtained per sample. Reads were processed using the ENCODE[98] ATAC-seq pipeline. Filtered peaks were sorted and merged across samples using bedops (v2.4.35)[68] and were used to generate a peak count table using featureCounts (v2.0.2).[69] Raw counts were analyzed with DESeq2 (Table S6).

T7-assisted reporter mapping library analysis

Sequencing reads were first demultiplexed using the bcl2fastq software. Under the current library design and sequencing scheme, Read1 started from genomic sequence and extended into the integrated synHEK3 construct, while Read2 contained reporter barcode information. Processing was as follows: 1) for each sequencing read pair, the 16-bp reporter barcode was extracted from Read 2 and attached to the read name of Read 1. 2) Read1 was trimmed using cutadapt (v4.1)[74] with following parameters: –cores=4— discard-untrimmed -e 0.2 -m 10 -O 8 -a CCCTAGAAAGATAGTCTGCGTAAAATTGACGCATG. The adapter corresponded to the 3'ITR of *piggyBac* transposon. Since parameter "–discard-untrimmed" was used, only reads spanning insertion junctions were kept and trimmed. 3) Trimmed sequences were aligned to the GRCh38 reference genome using bwa mem (v0.7.17).[75] The "-Y" option was used to enable soft clipping for supplementary alignments. 4) Reads uniquely (without XA:Z tag) and contiguously mapped near putative *piggyBac* landing pads (TTAA motifs) were kept using samtools (v1.9)[67] and a custom script. 5) Aligned reads were converted to BED format using the sam2bed function in bedops (v2.4.35). Reads aligned to standard chromosomes were kept. 6) Insertion points were calculated for all reads based on strand of alignment. And reads were sorted by the insertion coordinates. 7) The first 8-bp of reads were used as UMIs. A custom script was used to collapse reads at a per-location, per-barcode, per UMI basis. A barcode-location-UMI count table was generated. 8) synHEK3 barcodes <3 Levenshtein Distances at each location were collapsed and barcodes with > 3 UMIs were kept. 9) The count table was converted to a GenomicRanges[76] object in R. Coordinates of the last 4 base pairs of aligned reads were designated as genomic locations of the inserted synHEK3 reporters. 10) "Landing pads" of the mapped synHEK3 reporters were retrieved using the getSeq() function in the BSgenome package. Reporters that didn't have a TTAA sequence (which could be due to PCR error, mapping error, or the use of non-canonical landing pads) were removed. 11) SynHEK3 barcodes mapped to more than one location were removed.

Aligned reads were visualized on the Integrated Genome Browser (IGV; Figure 1B).[88] Motif enrichment and visualization (Figure 1C) was performed using WebLogo 3.[85] Genomic coverage analysis and annotation of the synHEK3 reporters (Figures 1D and 1E) were performed using the ChIPseeker package[77,78] in R. Enrichment of synHEK3 reporters across genomic features (Figures S1G and S1H) were performed using the annotatePeaks.pl function in Homer.[79] A custom script was used to find overlapping genes and calculate distance to closest TSSs.

ENCODE dataset analysis

ENCODE datasets were downloaded from the ENCODE portal.[99] Accessions of the datasets are listed in key resources table and Table S2. The GenomicRanges object containing all unique synHEK3 reporters was resized to 2 kb in width. The R package Genomation[80] was used to extract values within the 2-kb windows from the bigwig files of the chromatin features. For a specific dataset, scores of the synHEK3 reporters were calculated as $ln(\sum_1^{2000} x_i + c)$, where x_i was the value at base i, and c was the minimum

non-zero value of the chromatin feature score in a base within all the windows surveyed. For the generation of heatmap in Figure 6B, chromatin feature scores were scaled using the scale() function in R. Clustering and visualization of synHEK3 reporters was performed using the ComplexHeatmap package[86] in R. Visualization of the epigenetic tracks was done by Gviz[87] in R (Figures 2F and S2I).

synHEK3 editing library analysis
Sequencing reads from the Illumina NextSeq platforms were first demultiplexed using the bcl2fastq software. The 16-bp reporter barcodes were extracted from read and attached to its read name. Sequences around the CRISPR cut site were extracted for all reads. A barcode-editing outcome table was generated. For prime editing experiments, a custom script was used to align sequences to a reference sequence and count mutation frequency for every barcode. For Cas9 mutagenesis analysis, all sequences were aggregated and editing outcomes (alleles) with the highest number of counts were selected. These most frequent alleles were then aligned to reference sequences using needleall[70,71] with the following parameters: -gapopen 20 -gapextend 0.5 -endopen 20. A custom script was used to annotate the mutational events. MMEJ alleles were selected based on the following criteria: 1) microhomology sequences being at least 2 bp; 2) observed allele frequencies being 6-fold higher than expected frequencies. The 9 most frequent non-wild-type alleles were used for calculating the MMEJ/(MMEJ+NHEJ) ratio in Figures S3G–S3I.

sci-RNA-seq3 transcriptome library analysis
Sequencing reads from the Illumina NextSeq platforms were first demultiplexed based on P5 PCR index using the bcl2fastq software. Reads were filtered based on RT and ligation index, by allowing <2 Hamming distance from reference. Filtered reads were trimmed with TrimGalore (v0.6.6)[74] with parameters: -a AAAAAAAA –three_prime_clip_R1 1. Reads were then aligned to the GRCh38 reference genome using STAR (v2.7.6a). PCR duplicates were collapsed using the RT index, ligation index, UMI sequence and end co-ordinate of reads. Reads were further demultiplexed based on the combination of the RT, ligation and PCR index and split into files for individual cells. To generate gene expression count matrices, reads were assigned to the exonic and intronic region of closest genes with HTseq (v.2.0.2).[81] Reads with ambiguous assignments were discarded. Cells were further filtered based on total UMI (> 100) and mitochondrial reads percentage (<10%). Cells with the number of features detected between 10% and 90% percentile of all cells were kept and considered high-quality cells. The single cell analysis was performed using the Seurat (v4.0.0) package[82] in R.

Pooled shRNA screen analysis
Sequencing reads from the Illumina NextSeq platforms were first demultiplexed based on P5 PCR index using the bcl2fastq software. A custom script was used to further demultiplex and filter the synHEK3 and shRNA libraries based on the combination of RT, ligation and PCR index (cell ID), by allowing <2 Hamming distance from reference. For the synHEK3 library, reporter barcodes (<2 Hamming distance to the reference barcode set), prime editing outcomes and read UMIs were extracted. For the shRNA library, shRNA sequences (<2 Hamming distance to the reference set) and read UMIs were extracted. UMIs with <3 Hamming distances were collapsed.
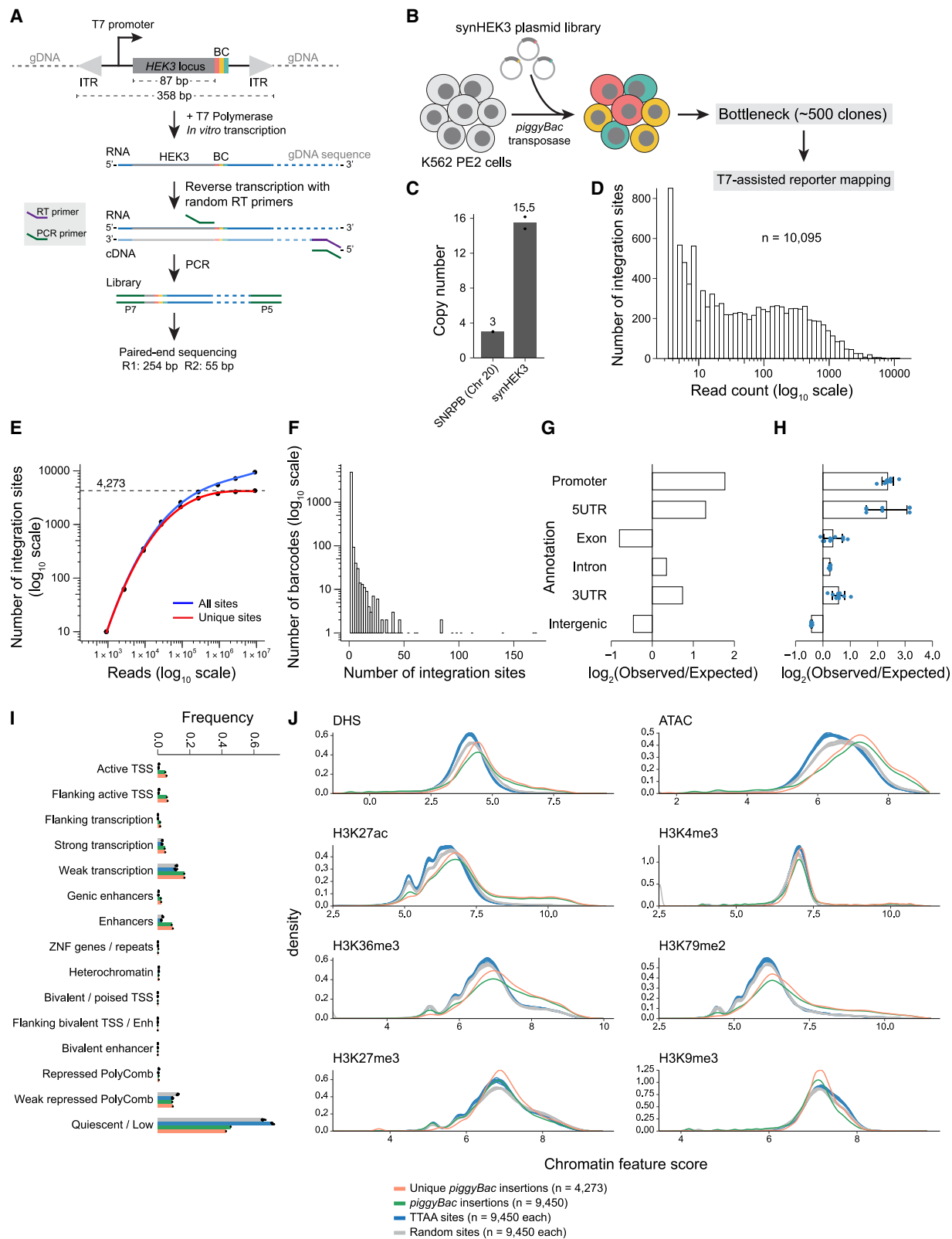
A series of pre-processing steps were applied to the data. 1) We matched cell IDs between the sci-RNA-seq3 transcriptome libraries and the capture libraries, and only kept high-quality cells (number of features between 10% and 90% percentile). 2) For the shRNA library, cells with <3 or >200 UMIs or >20 shRNA were removed. 3) For the synHEK3 library, cells with <4 UMIs were first removed. We counted UMIs for synHEK3 reporters belonging to the two clones and calculated a clone UMI/total UMI ratio. If the ratio was >80% for a specific clone, the cell is assigned to the corresponding clone. Cells with extremely high UMIs (cutoff between 100 and 250 UMIs, depending on library complexity and sequencing depth of the plate) were also removed before downstream analysis. 4) For a cell-synHEK3 barcode combination, if multiple insertion sequences were observed, sequences with <3 Hamming distance were collapsed. Cell-synHEK3 barcode pairs having conflicting editing outcomes were discarded. 5) A cell-synHEK3 barcode-editing outcome-shRNA matrix was generated for a common set of cells identified by the shRNA and synHEK3 libraries.

In general, the prime editing efficiency of a specific synHEK3 reporter was estimated in all cells containing the reporter by collapsing reads on a per-cell, per-barcode basis. To assess the effect of an shRNA on a specific synHEK3 reporter, editing efficiency of this reporter was calculated in cells without the specific shRNA, which was defined as the hypothesized probability of editing. Then, by treating prime editing outcomes as binary events (unedited vs. insertion), the total number of cells containing the shRNA and number of cells with 6-bp insertions were counted. A p value was computed using the function binom.test() in R. For each clone, all synHEK3-shRNA pairs were assessed by binomial tests. Q-Q plots in Figure 5A were generated using uncorrected p values. The p values were further corrected using the function p.adjust() in R and the Benjamini-Hochberg method was used. Empirical p values of candidate synHEK3-shRNA pairs were calculated as $(n\_lower + 1)/(N + 1)$, where n_low was the number of control tests with a raw p value lower than the candidate test's raw p value, and N was the total number of control tests. Empirical p values were then Benjamini-Hochberg corrected, and those with eFDR < 5% were reported. Processed screen results for clone 3 and clone 5 are provided in Table S5.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical tests for each experiment are described in the text, figure legends or STAR Methods. In Figure 2C, the beta regression model was performed using the function betareg() in R package betareg[83] using default parameters. In Figures 2B–2D, 2G–2I, 3A, S2A, S2C, S2D, S4A, and S5C, Spearman's and Pearson's correlation coefficient and *p* values were generated using the cor() function or the rcorr() function in the package Hmisc (Harrell and Dupont) in R. In Figures S3I, S3J, and S5J, the fit with confidence intervals was produced using function geom_smooth() in the R package ggplot using parameter 'method = ''lm''' and all other parameters being default. In Figure 5, binomial tests were performed to measure the effects of shRNAs and the resulting *p* values were corrected for multiple hypothesis testing using the Benjamini-Hochberg method. Empirical *p* values of candidate synHEK3-shRNA pairs were calculated as (n_lower + 1)/(N + 1), where n_low was the number of control tests with a raw *p* value lower than the candidate test's raw *p* value, and N was the total number of control tests. Empirical *p* values were then Benjamini-Hochberg corrected, and those with eFDR < 5% were reported. In Figures S5I, S5J, S6F, S6G, and S7B, differential gene expression or accessibility analyses were performed using the DESeq2 package in R (Wald tests). In Figure 6C, Fisher's exact test was used to compare the differential enrichment of gene-overlapping synHEK3 reporters in the two groups. In Figures 3E–3G, 6D, S2H, S3F, S3G, S6D, and S6E, *p* values were calculated using two-sided Kolmogorov–Smirnov tests. In Figure 7B, scatter plots report mean editing efficiencies of two biological replicates and error bars correspond to standard deviations. In Figure 7C, control editing efficiencies were predicted using linear model trained in synHEK3 reporters that are not in the corresponding CRISPRoff target genes using the lm() function in R. In Figure S7C, welch's two sample t-tests were used to compare prime editing efficiency of genes before and after CRISPR activation.

A T7 promoter — gDNA, ITR, HEK3 locus, BC, gDNA, 87 bp, 358 bp, + T7 Polymerase *In vitro* transcription, RNA, HEK3, BC, gDNA sequence, Reverse transcription with random RT primers, RT primer, PCR primer, RNA, cDNA, PCR, Library, P7, P5, Paired-end sequencing R1: 254 bp R2: 55 bp

B synHEK3 plasmid library, K562 PE2 cells, *piggyBac* transposase, Bottleneck (~500 clones), T7-assisted reporter mapping

C Copy number — SNRPB (Chr 20): 3, synHEK3: 15.5

D Number of integration sites vs Read count (log₁₀ scale), n = 10,095

E Number of integration sites (log₁₀ scale) vs Reads (log₁₀ scale), 4,273, All sites, Unique sites

F Number of barcodes (log₁₀ scale) vs Number of integration sites

G Annotation: Promoter, 5UTR, Exon, Intron, 3UTR, Intergenic vs log₂(Observed/Expected)

H log₂(Observed/Expected)

I Frequency — Active TSS, Flanking active TSS, Flanking transcription, Strong transcription, Weak transcription, Genic enhancers, Enhancers, ZNF genes / repeats, Heterochromatin, Bivalent / poised TSS, Flanking bivalent TSS / Enh, Bivalent enhancer, Repressed PolyComb, Weak repressed PolyComb, Quiescent / Low

J DHS, ATAC, H3K27ac, H3K4me3, H3K36me3, H3K79me2, H3K27me3, H3K9me3 — density vs Chromatin feature score

Unique *piggyBac* insertions (n = 4,273)
piggyBac insertions (n = 9,450)
TTAA sites (n = 9,450 each)
Random sites (n = 9,450 each)

(legend on next page)

Figure S1. Characterization of synHEK3 insertion sites determined by T7-assisted reporter mapping assay, related to Figure 1

(A) Schematic of the synHEK3 reporter construct and the T7-assisted reporter mapping method.

(B) Experimental workflow. A complex library of synHEK3 reporters was transfected into PE2(+) K562 cells along with the *piggyBac* transposase. After transposons were stably integrated, the population of cells was bottlenecked to ~500 clones. Cells derived from the expansion of this bottlenecked pool were then subjected to T7-assisted reporter mapping.

(C) Copy number of synHEK3 reporters was estimated by qPCR using that of *SNRPB* (3 copies) as a reference.

(D) Histogram of T7 mapping counts of all synHEK3 reporters (n = 10,095). x axis on $\log_{10}$ scale.

(E) Scatter plot of the number of integration sites recovered at different sequencing depths by subsampling the original dataset. Overlaid polynomial regression lines fit to all sites (blue) or unique (red) sites. x and y axes on $\log_{10}$ scale.

(F) Histogram of number of distinct genomic integration sites per synHEK3 barcode. y axis on $\log_{10}$ scale.

(G) $\log_2$ enrichment of the number of synHEK3 reporters across major genomic features relative to the expected numbers based on feature sizes.

(H) $\log_2$ enrichment of the number of synHEK3 reporters across major genomic features relative to the expected number of TTAA motifs in each feature. Blue points: $\log_2$ enrichment calculated using 10 sets of randomly sampled TTAA motifs (n = 4,273). Error bar: standard deviation.

(I) Barplot showing fractions of genomic sites overlapping with the Roadmap annotations. For *piggyBac* integrations, the overlapping status of all integrations (n = 9,450, green) and integrations marked by unique barcodes (n = 4,273, orange) are shown. For background, the mean fraction of 10 sets of TTAA sites (n = 9,450 each, blue) and 10 sets of random 4-bp sites (n = 9,450 each, gray) are shown.

(J) Density plots of chromatin feature scores of the selected genomic sites. For *piggyBac* integrations, distributions of all integrations (n = 9,450, green) and integrations marked by unique barcodes (n = 4,273, orange) are plotted. For background, 10 sets of TTAA sites (n = 9,450 each, blue) and 10 sets of random 4-bp sites (n = 9,450 each, gray) are plotted.
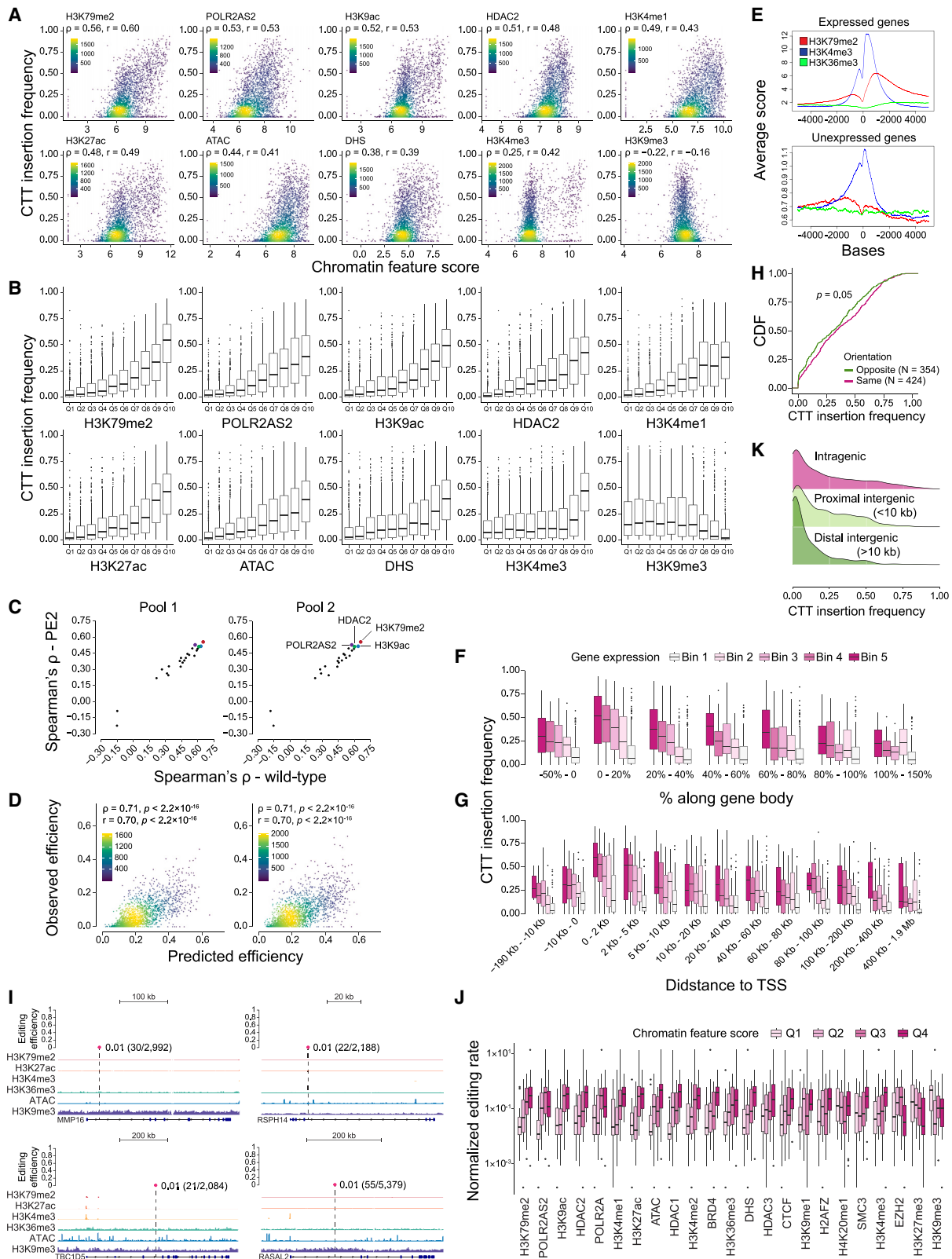
(legend on next page)

Figure S2. Chromatin context has a major impact on prime editing efficiency, related to Figure 2

(A) Scatter plots of chromatin feature scores vs. prime editing efficiencies for individual synHEK3 reporters. Points are colored by the number of neighboring points. The Spearman's (ρ) and Pearson's (r) correlation coefficients between the chromatin feature score and prime editing efficiency are annotated.

(B) Boxplots of prime editing efficiency in synHEK3 sites, binned by chromatin feature scores. Q1–Q10 correspond to 10 equally sized bins of synHEK3 reporters with increasing chromatin feature scores.

(C) Scatter plots of Spearman's ρ of chromatin feature scores vs. prime editing efficiency observed in two independent pools of synHEK3 reporters in wild-type (x axis) or PE2(+) K562 cells (y axis).

(D) Scatter plots of predicted (beta regression model) vs. observed prime editing efficiencies in two independent pools of synHEK3 reporters in wild-type K562 cells. Points are colored by the number of neighboring points. p values determined by the Spearman correlation test or by the Pearson correlation test.

(E) Histogram of ChIP-seq signal of H3K79me2 (red), H3K4me3 (blue), and H3K36me3 (green) in an 10-kb window surrounding the TSSs of genes that are expressed (n = 11,514, TPM > 3, top) or unexpressed (n = 10,506, TPM < 3, bottom) in K562 cells.

(F) Boxplot of prime editing efficiencies of gene-proximal synHEK3 reporters. Distances were calculated relative to the closest TSS, scaled by gene length, and binned. Negative values refer to synHEK3 sites located upstream of TSSs. Values >100% refer to synHEK3 sites located downstream of TTSs. The reporters were binned based on the TPM of the overlapping/nearest genes. Bin 1 contains unexpressed genes, while bins 2–5 are equally sized in terms of the number of assigned genes (though not necessarily in terms of the number of assigned synHEK3 reporters), sorted by increasing expression levels.

(G) Boxplot of prime editing efficiencies of gene-proximal reporters. Same as (F) except that distance groupings are in terms of absolute rather than scaled distances from TSSs.

(H) Cumulative distribution function (CDF) plot of prime editing efficiencies of intragenic synHEK3 reporters of different orientations with respect to the direction of transcription. "Opposite" means synHEK3 reporter on the opposite strand of the coding strand, and "same" means synHEK3 reporter on the same strand as the coding strand. Genes with detectable expression (TPM > 3) and synHEK3 reporters within 50 kb from the TSSs were selected for this analysis. p value: two-sided Kolmogorov-Smirnov test.

(I) Genome browser views of 4 very poorly editable sites. Sites of integration and measured editing efficiencies are shown as a dot plot at the top and aligned with selected epigenetic tracks. For each synHEK3 insertion, editing efficiency, number of reads with edit (numerator), and total number of reads (denominator) are annotated. The dashed vertical lines mark locations of the insertions.

(J) Boxplots of normalized editing rates (y axis) for epegRNAs designed for prime editing at endogenous genomic sites, stratified by chromatin feature scores. The epegRNAs are grouped into quartiles based on each chromatin feature score (Q1 = lowest; Q4 = highest). The y axis is in $\log_{10}$ scale.

(K) Density plot of CTT insertion frequency of synHEK3 reporters in intragenic, proximal intergenic (<10 kb), and distal intergenic (>10 kb) regions.
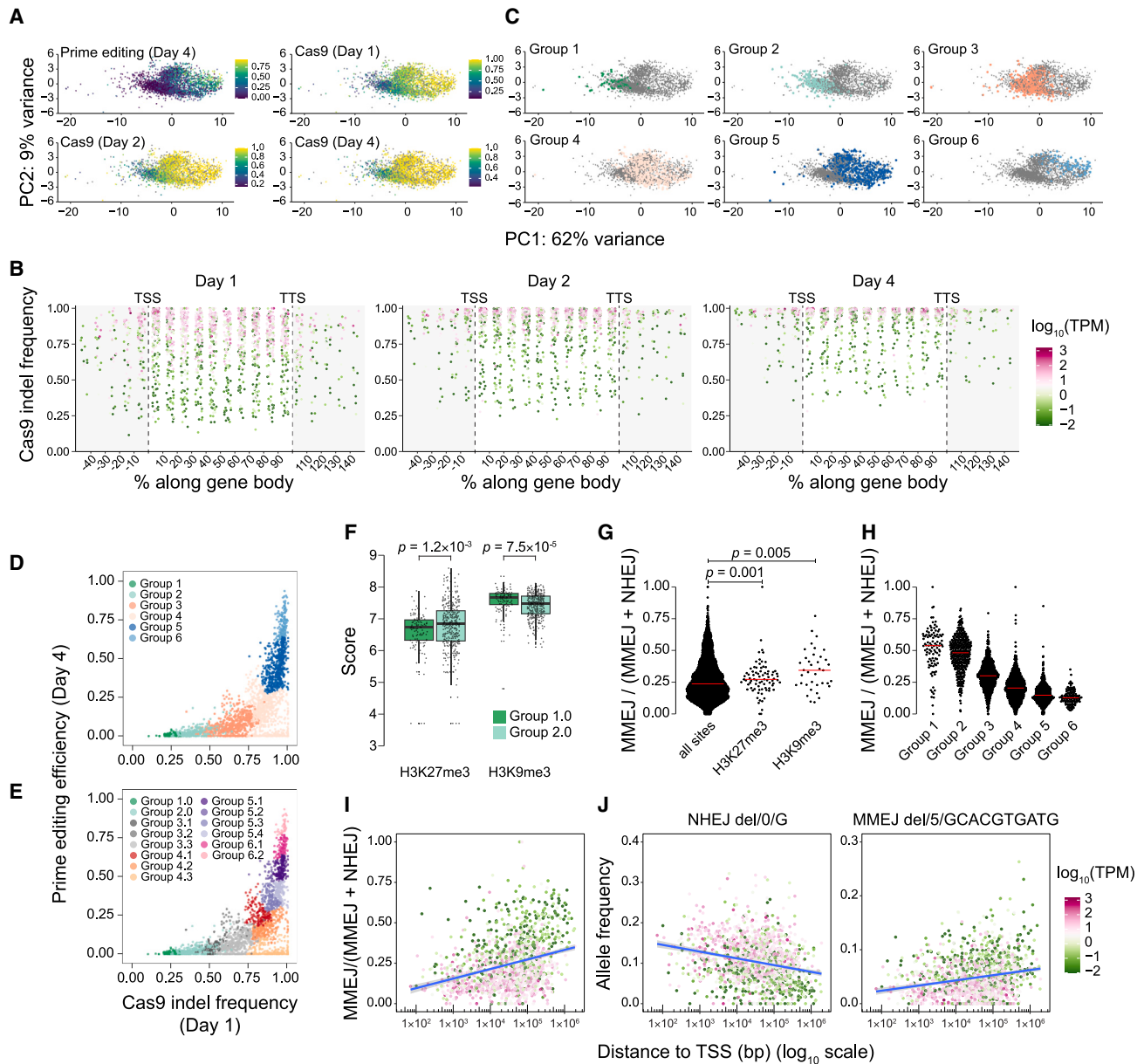
Figure S3. Comparison between prime editing and Cas9 editing, leveraging a common set of integrated editing reporters, related to Figure 3

(A) PCA plots of synHEK3 reporters generated using chromatin scores as features. The first two PCs are plotted (PC1: variance 62%; PC2: variance 9%). Points are colored by prime editing efficiency at day 4 and Cas9 indel frequency measured at days 1, 2, and 4.

(B) Cas9 indel frequency measured at days 1, 2, and 4 for gene-proximal reporters. Distance was calculated relative to the closest TSS, scaled by gene length, and binned. Negative values refer to synHEK3 sites located upstream of TSSs. Values >100% refer to synHEK3 sites located downstream of TTSs. Points are colored based on the expression levels (log_{10}) of the genes.

(C) PCA plots of synHEK3 reporters generated using chromatin scores as features. The first two PCs are plotted (PC1: variance 62%; PC2: variance 9%). Points are colored by the 6 main groups as in Figure 3B.

(D) Scatter plot of Cas9 editing (day 1) and prime editing (day 4) efficiency. Points are colored by the 6 main groups as in (C).

(E) Scatter plot of Cas9 editing (day 1) and prime editing (day 4) efficiency, colored by the 14 subgroups.

(F) Boxplot of H3K27me3 and H3K9me3 scores of synHEK3 reporters in groups 1.0 and 2.0. *p* value: two-sided Kolmogorov-Smirnov test.

(G) The MMEJ/(MMEJ + NHEJ) ratio in all synHEK3 sites and in sites overlapping with H3K27me3 and H3K9me3 peaks. Red line: the median MMEJ/(MMEJ + NHEJ) ratio. *p* value: two-sided Kolmogorov-Smirnov test.

(H) The MMEJ/(MMEJ + NHEJ) ratio in the 6 groups of synHEK3 sites as in (C). Red line: the median MMEJ/(MMEJ + NHEJ) ratio.

(legend continued on next page)

(I and J) Scatter plots of the MMEJ/(MMEJ + NHEJ) ratio or allele frequencies of intragenic synHEK3 reporters and their distance to corresponding TSSs (x axis; $\log_{10}$ scale). Points are colored based on the expression levels ($\log_{10}$) of the genes. Blue line: linear regression line with confidence interval in gray.

(I) The MMEJ/(MMEJ + NHEJ) ratio is plotted.

(J) Allele frequencies of the most frequent NHEJ (left) and MMEJ (right) alleles are plotted. The "NHEJ del/0/G" allele contains a single-base deletion of G at the Cas9 cut site. The "MMEJ del/5/GCACGTGATG" allele contains a bidirectional deletion of a 10-bp sequence (GCACGTGATG) around the cutsite.
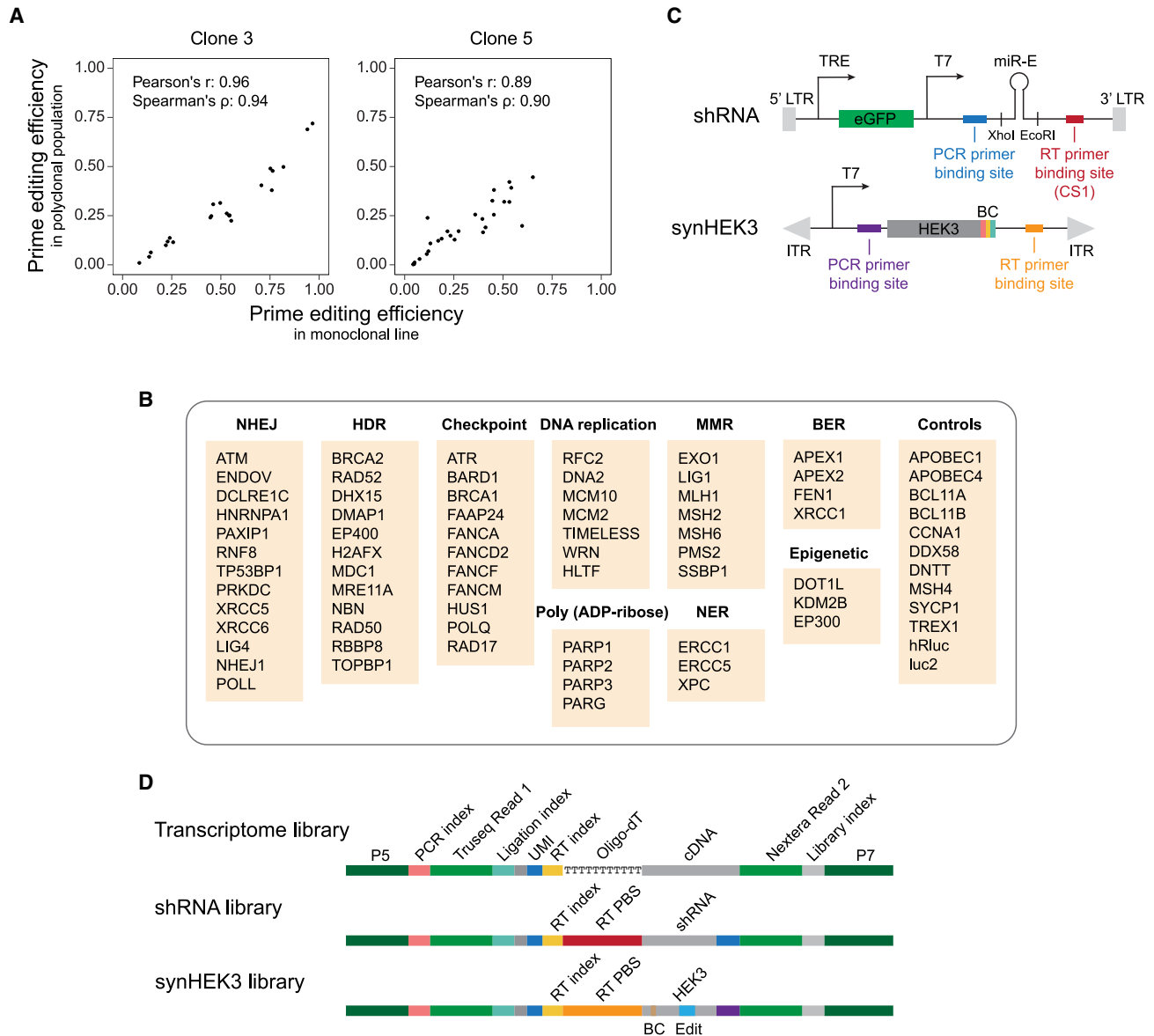
Figure S4. Experimental setup of the pooled shRNA screen and its readout via T7 IST-assisted sci-RNA-seq3, related to Figure 4

(A) Correlation between the efficiencies of random 3-bp insertions measured in the monoclonal lines vs. efficiencies measured for the same reporters in the original polyclonal population.

(B) List of genes targeted by the shRNA library. Genes are grouped by pathways. Control genes are shown separately.

(C) Schematic of the lentiviral shRNA construct and the synHEK3 reporter, with key features relevant to the sci-RNA-seq3 workflow highlighted. TRE, tetracycline response element.

(D) Schematic structures of the sci-RNA-seq3, shRNA, and synHEK3 libraries. UMI, unique molecular identifier; RT PBS, reverse transcription primer-binding site.
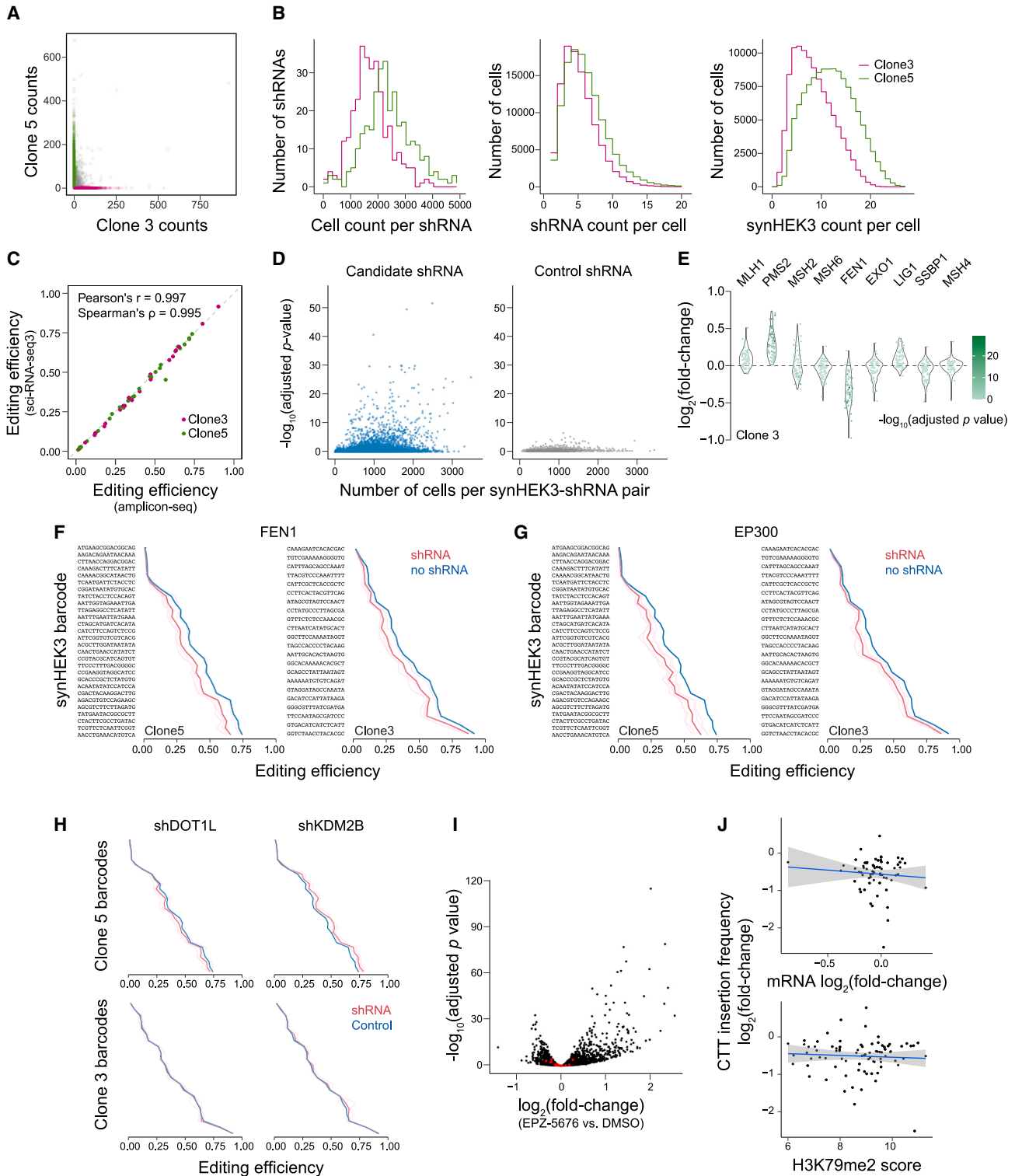
Figure S5. A pooled shRNA screen with sci-RNA-seq3 and effects of perturbing the MMR pathway on prime editing, related to Figure 5

(A) Scatter plot of synHEK3 UMIs detected in single cells in sci-RNA-seq3 data. Cells assigned to the two clones are colored (green: clone 5; pink: clone 3). Mixed cells are in gray.

(B) Histograms of cell count per shRNA (left), number of shRNAs captured per cell (middle), and number of synHEK3 reporters captured per cell (right) in the two clones.

(C) Scatter plot of prime editing frequencies of synHEK3 reporters estimated with sci-RNA-seq3 vs. bulk amplicon sequencing.

(D) Scatter plots of the number of cells per synHEK3-shRNA pair and corresponding adjusted p values ($-\log_{10}$). Candidate shRNAs (left) and control shRNAs (right) are plotted separately. Raw p values from the binomial tests were corrected for multiple hypothesis testing using the Benjamini-Hochberg method.

(E) Effects of shRNAs targeting MMR-related genes in clone 3. Log$_2$ fold-changes of prime editing efficiencies of synHEK3-shRNA pairs are plotted and colored by their corresponding adjusted p values ($-\log_{10}$).

(F and G) Effects of shRNAs against FEN1 and EP300. Pink lines: editing frequencies in cells with individual shRNAs; red line: mean editing frequencies of the gene-targeting shRNAs; light blue lines: control editing frequencies for individual shRNAs (not visible because low variance relative to mean line, shown in blue); blue line: mean control editing frequencies.

(H) Effects of shRNAs against DOT1L and KDM2B. Barcode sequences omitted but in the same order as (F) and (G). Colored as in (F) and (G).

(I) Volcano plots of gene expression changes in EPZ-5676-treated cells (5 μM for 6 days). Genes with synHEK3 insertions in clone 3 or clone 5 of the single-cell screen are colored in red. p value: Wald test.

(J) Scatter plots of log$_2$ fold-change of prime editing efficiency vs. gene expression change (log$_2$ fold; top) and H3K79me2 score (bottom) for a set of intragenic synHEK3 reporters with baseline efficiency >20%. Blue line: linear regression line with confidence interval in gray.
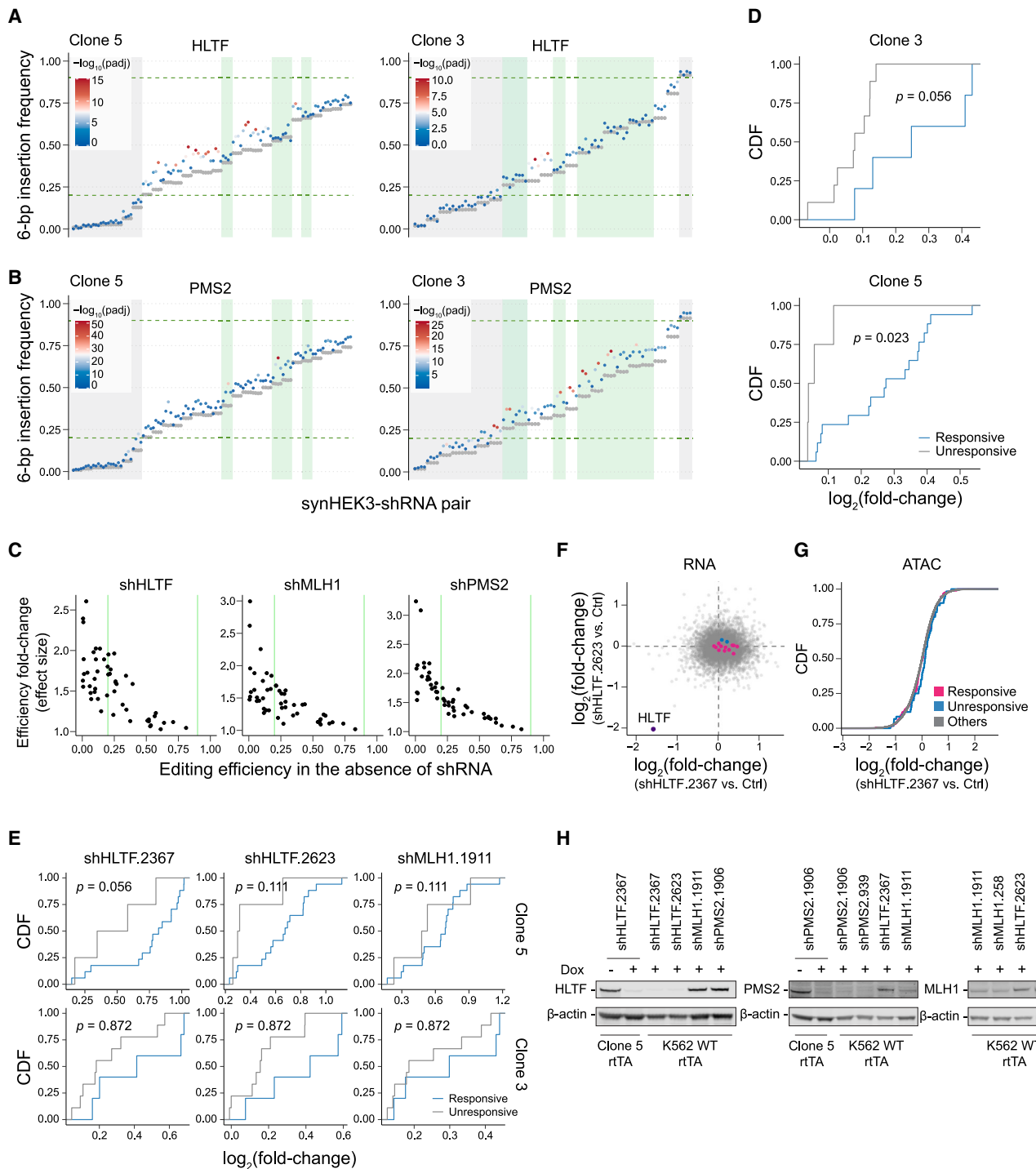
Figure S6. Chromatin context-specific response to HLTF inhibition, related to Figure 6

(A) Effects of shRNAs against HLTF on all synHEK3 reporters in clone 5 (left) and clone 3 (right). For every synHEK3-shRNA pair, editing frequencies in cells with candidate shRNA are plotted and colored by their statistical significance. Control editing frequencies are shown in gray. Regions of the plot containing synHEK3 reporters that are significantly less responsive to HLTF inhibition are shaded green. Regions of the plot containing synHEK3 reporters with low (<0.2) or very high (>0.9) editing frequencies are shaded gray. Dashed horizontal lines indicate the editing frequencies of 0.2 and 0.9. Raw p values from the binomial tests were corrected for multiple hypothesis testing using the Benjamini-Hochberg method.

(B) Effects of shRNAs against PMS2 on all synHEK3 reporters in clone 5 (left) and clone 3 (right). Layout as in (A).

C) Scatter plots of baseline editing efficiencies (x axes) vs. fold-changes induced by shRNAs against HLTF, MLH1, and PMS2 (y axes) across 50 synHEK3 reporters. Green vertical lines correspond to the baseline editing efficiencies of 0.2 and 0.9.
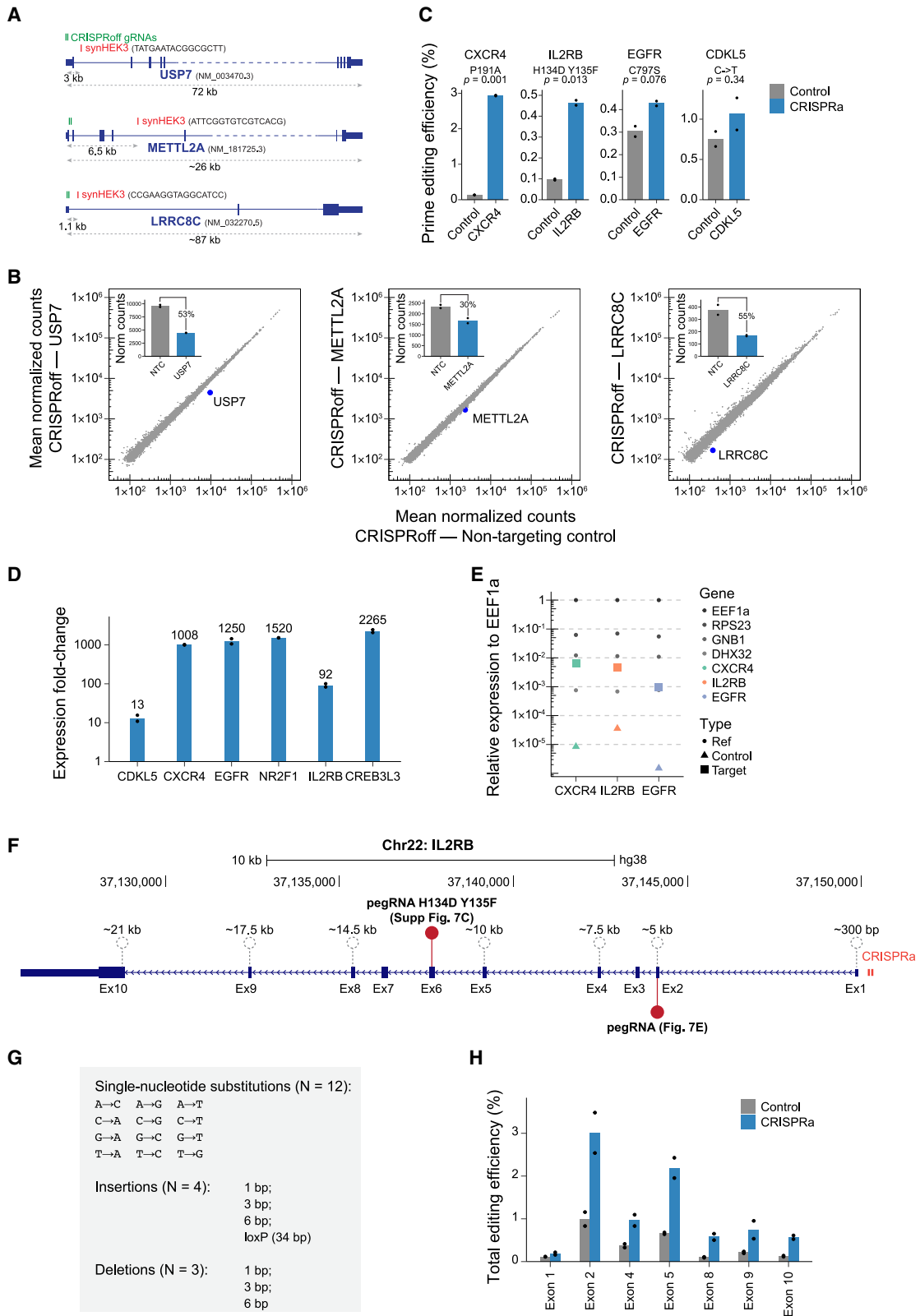
(D) CDF plots of $\log_2$ fold-change of mean editing frequency induced by shRNAs against HLTF. synHEK3 reporters are colored by their responsiveness to HLTF inhibition based on the shRNA screen. p value: two-sided Kolmogorov-Smirnov test.

(E) Validation of differential responsiveness to HLTF inhibition in cells transduced with individual shRNAs. CDF plots of $\log_2$ fold-change of editing frequency induced by shRNAs against HLTF (shHLTF.2367: used in the shRNA screen; shHLTF.2623: an orthogonal shRNA) or MLH1 (shMLH1.1911). synHEK3 reporters are colored by their responsiveness to HLTF inhibition based on the shRNA screen. p value: two-sided Kolmogorov-Smirnov test.

(F) Scatter plot of RNA $\log_2$ fold-changes induced by two shRNAs against HLTF. Genes overlapping or near shHLTF-responsive sites are in pink; genes overlapping or near shHLTF-unresponsive sites are in blue; other sites are in gray; HLTF is shown in purple (at the bottom left corner; −1.5 and −2 $\log_2$ fold-change).

(G) CDF plot of $\log_2$ fold-changes of ATAC-seq peak counts induced by shHLTF.2367. Peaks near synHEK3 sites were defined as those (1) within 5 kb of a synHEK3 site and/or (2) in the promoter or body of a gene overlapping or proximal (within 10 kb) to a synHEK3 insertion.

(H) Western blot analysis of HLTF, PMS2, and MLH1. Clone 5 or a wild-type (WT) K562 line were transduced with rtTA and shRNAs targeting the candidate genes. Protein expression was compared with cells without doxycycline (Dox) or cells transduced with an shRNA targeting a different gene.

A

II CRISPRoff gRNAs
I synHEK3 (TATGAATACGGCGCTT)

USP7 (NM_003470.3)
3 kb
72 kb

II
I synHEK3 (ATTCGGTGTCGTCACG)
METTL2A (NM_181725.3)
6.5 kb
~26 kb

II I synHEK3 (CCGAAGGTAGGCATCC)
LRRC8C (NM_032270.5)
1.1 kb
~87 kb

C

Prime editing efficiency (%)

CXCR4
P191A
p = 0.001

IL2RB
H134D Y135F
p = 0.013

EGFR
C797S
p = 0.076

CDKL5
C->T
p = 0.34

Control
CRISPRa

B

Mean normalized counts
CRISPRoff — USP7

53%
NTC USP7

USP7

CRISPRoff — METTL2A

30%
NTC METTL2A

METTL2A

CRISPRoff — LRRC8C

55%
NTC LRRC8C

LRRC8C

Mean normalized counts
CRISPRoff — Non-targeting control

D

Expression fold-change

13   1008   1250   1520   92   2265

CDKL5 CXCR4 EGFR NR2F1 IL2RB CREB3L3

E

Relative expression to EEF1a

Gene
EEF1a
RPS23
GNB1
DHX32
CXCR4
IL2RB
EGFR

Type
Ref
Control
Target

CXCR4 IL2RB EGFR

F

Chr22: IL2RB

10 kb
hg38

37,130,000   37,135,000   37,140,000   37,145,000   37,150,000

pegRNA H134D Y135F
(Supp Fig. 7C)

~21 kb   ~17.5 kb   ~14.5 kb   ~10 kb   ~7.5 kb   ~5 kb   ~300 bp

CRISPRa
II

Ex10   Ex9   Ex8 Ex7 Ex6 Ex5   Ex4 Ex3 Ex2   Ex1

pegRNA (Fig. 7E)

G

Single-nucleotide substitutions (N = 12):
A→C   A→G   A→T
C→A   C→G   C→T
G→A   G→C   G→T
T→A   T→C   T→G

Insertions (N = 4):   1 bp;
                        3 bp;
                        6 bp;
                        loxP (34 bp)

Deletions (N = 3):   1 bp;
                        3 bp;
                        6 bp

H

Total editing efficiency (%)

Control
CRISPRa

Exon 1 Exon 2 Exon 4 Exon 5 Exon 8 Exon 9 Exon 10

(legend on next page)

Figure S7. Modulating prime editing outcomes by epigenetic reprogramming, related to Figure 7

(A) Schematic diagrams of target genes in the CRISPRoff experiment, with the locations of the CRISPRoff gRNAs (green) and synHEK3 reporters (red) annotated.

(B) Scatter plots of mean normalized counts of gene expression in cells transfected with CRISPRoff gRNAs at day 11 (2 replicates each). x and y axes on $\log_{10}$ scale. Insets are bar plots of normalized counts of the CRISPRoff target genes. NTC, non-targeting control.

(C) Prime editing efficiency (%) at endogenous gene targets in K562 cells, with or without epigenetic editing via CRISPRa. Gray bars show mean prime editing efficiencies when control promoters were activated via CRISPRa. Blue bars show mean prime editing efficiencies when target gene promoters were activated via CRISPRa. p value: Welch's two-sample t test.

(D) mRNA fold-change of CRIPSRa target genes quantified by qPCR.

(E) Relative expression levels of selected CRISPRa target genes compared with a set of reference genes. Circle: reference genes; triangle: endogenous expression levels of the target genes; square: expression levels of the target genes after CRISPRa. y axis on $\log_{10}$ scale.

(F) Schematic of the IL2RB gene with locations of gRNA (for CRISPRa) or (e)pegRNA (for prime editing) targets annotated. Red lollipops indicate pegRNAs that were separately tested in experiments shown in Figures 7E and S7C. Dashed lollipops indicate locations of the new epegRNA pools.

(G) List of epegRNA edit types tested, including all possible single-nucleotide substitutions, small insertions of 1, 3, and 6 bp, insertion of a *loxP* site (34 bp), and small deletions of 1, 3, and 6 bp.

(H) Total editing efficiencies measured at the individual *IL2RB* exons. Gray bars show mean prime editing efficiencies in control groups, and blue bars show mean prime editing efficiencies when the *IL2RB* promoter was first activated via CRISPRa.