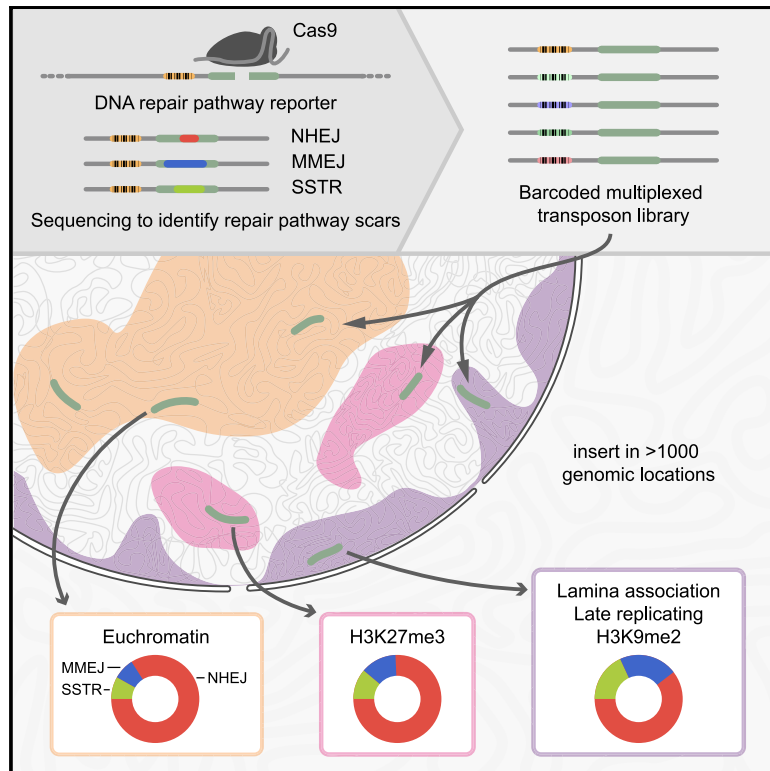


Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance

Graphical abstract



Authors

Ruben Schep, Eva K. Brinkman, Christ Leemans, ..., Roderick L. Beijersbergen, René H. Medema, Bas van Steensel

Correspondence

b.v.steensel@nki.nl

In brief

Schep et al. designed a reporter to probe the activities of three DNA double-strand break repair pathways. Integration of this reporter in >1,000 genomic locations in a human cell line provided a detailed view of the impact of local chromatin context on the balance between the repair pathways.

Highlights

- Sequencing-based reporter detects activities of multiple DSB repair pathways
- Multiplexing approach to integrate reporter in >1,000 genomic locations
- Overlay with epigenome data reveals effects of chromatin context on pathway balance
- Pathway balance differs between heterochromatin and euchromatin



Article

Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance

Ruben Schep,^{1,2} Eva K. Brinkman,^{1,2,7} Christ Leemans,^{1,2} Xabier Vergara,^{1,2,3} Robin H. van der Weide,^{1,2} Ben Morris,^{4,5} Tom van Schaik,^{1,2} Stefano G. Manzo,^{1,2} Daniel Peric-Hupkes,^{1,2} Jeroen van den Berg,^{1,3} Roderick L. Beijersbergen,^{4,5} René H. Medema,^{1,3} and Bas van Steensel^{1,2,6,8,*}

¹Oncode Institute, Netherlands Cancer Institute, 1066 CX, Amsterdam, the Netherlands

²Division of Gene Regulation, Netherlands Cancer Institute, 1066 CX, Amsterdam, the Netherlands

³Division of Cell Biology, Netherlands Cancer Institute, 1066 CX, Amsterdam, the Netherlands

⁴Division of Molecular Carcinogenesis, Netherlands Cancer Institute, 1066 CX, Amsterdam, the Netherlands

⁵Robotics Screening Center, Netherlands Cancer Institute, 1066 CX, Amsterdam, the Netherlands

⁶Department of Cell Biology, Erasmus University Medical Centre, 3015 CN, Rotterdam, the Netherlands

⁷Present address: SciLifeLab, Department of Microbiology, Tumor and Cell Biology, Karolinska Institute, Box 1031, 171 21 Solna, Stockholm, Sweden

⁸Lead contact

*Correspondence: b.v.steensel@nki.nl

<https://doi.org/10.1016/j.molcel.2021.03.032>

SUMMARY

DNA double-strand break (DSB) repair is mediated by multiple pathways. It is thought that the local chromatin context affects the pathway choice, but the underlying principles are poorly understood. Using a multiplexed reporter assay in combination with Cas9 cutting, we systematically measure the relative activities of three DSB repair pathways as a function of chromatin context in >1,000 genomic locations. This reveals that non-homologous end-joining (NHEJ) is broadly biased toward euchromatin, while the contribution of microhomology-mediated end-joining (MMEJ) is higher in specific heterochromatin contexts. In H3K27me3-marked heterochromatin, inhibition of the H3K27 methyltransferase EZH2 reverts the balance toward NHEJ. Single-stranded template repair (SSTR), often used for precise CRISPR editing, competes with MMEJ and is moderately linked to chromatin context. These results provide insight into the impact of chromatin on DSB repair pathway balance and guidance for the design of Cas9-mediated genome editing experiments.

INTRODUCTION

The repair of DNA double-strand breaks (DSBs) is crucial for genetic stability. In addition, it is a key step in CRISPR-Cas9-mediated genome editing (Jasin and Haber, 2016; Yeh et al., 2019). Several pathways can repair DSBs, including classical non-homologous end-joining (NHEJ), homologous recombination (HR), and microhomology-mediated end-joining (MMEJ) (Chang et al., 2017; Iliakis et al., 2015; McVey and Lee, 2008; Scully et al., 2019; Yeh et al., 2019). NHEJ directly re-joins blunt-ended DSBs, while HR typically uses the intact sister chromatid in G2 phase as a template to mend the break. In contrast, MMEJ recombines short homologous sequences that are close to either end of the DSB and consequently results in a small deletion. An additional variant is single-stranded template repair (SSTR), a type of homology-directed repair that requires a single-stranded oligodeoxynucleotide (ssODN) donor sequence (Lin et al., 2014; Richardson et al., 2016). SSTR is highly relevant because it is leveraged in CRISPR-Cas9-mediated genome editing to generate precisely designed small mutations, such as point mutations or small insertions or deletions (indels) (DeWitt et al., 2016; Okamoto et al., 2019; Riesenberger et al., 2019).

Which pathway repairs a particular DSB depends in part on the local DNA sequence (Allen et al., 2018; Chakrabarti et al., 2019; Chen et al., 2019; Shen et al., 2018) and on the stage of the cell cycle (reviewed in Chapman et al., 2012; Hustedt and Durocher, 2016; Mladenov et al., 2016). In addition, local chromatin packaging can affect the choice of repair pathway (Clouaire and Legube, 2015; Jeggo and Downs, 2014; Kalousi and Soutoglou, 2016; Scully et al., 2019). Most studies of chromatin effects so far have focused on the balance between HR and NHEJ. For example, the histone modification H3K36me3, which is present along active transcription units, is thought to promote HR (Aymard et al., 2014; Carvalho et al., 2014; Clouaire et al., 2018; Daugaard et al., 2012; Pfister et al., 2014). Paradoxically, H3K9 di- or trimethylated (H3K9me2/3) heterochromatin, which packages transcriptionally inactive regions of the genome, has also been implicated in promoting HR (Alagoz et al., 2015; Baldeyron et al., 2011; Lee et al., 2013; Soria and Almouzni, 2013; Sun et al., 2009), although some single-locus studies in mouse and fruit fly found no major change in the balance between NHEJ and HR when a sequence was shifted between heterochromatin and euchromatin states (Janssen et al., 2016; Kallimasioti-Pazi et al., 2018). Furthermore, reduced binding of HR proteins was



observed at a locus that was artificially tethered to the nuclear lamina (Lemaître et al., 2014), suggesting that spatial positioning of the DSB inside the nucleus may also play a role.

Much less is known about the impact of chromatin on MMEJ and SSTR. Like HR, these pathways require resection of the DNA ends to produce single-stranded DNA overhangs, but downstream of this step, the mechanisms and responsible proteins diverge (Chang et al., 2017; Scully et al., 2019; Yeh et al., 2019). It is thus possible that the local chromatin environment also modulates MMEJ and SSTR in unique ways, but this has remained largely unexplored (Clouaire and Legube, 2019; Mitrentsi et al., 2020).

One strategy to investigate the impact of local chromatin context on repair pathway balance is to generate DSBs at various genomic locations with known chromatin states and compare pathway use across these locations (Chakrabarti et al., 2019; Clouaire et al., 2018; van Overbeek et al., 2016). However, with such an approach it is difficult to separate the effects of chromatin context from the effects of sequence context because both vary simultaneously along the genome. Ideally, different chromatin contexts are compared while the sequence context is kept fixed.

Here, we report a strategy that effectively tackles these challenges in human cells. The strategy consists of two parts. First, we used a reporter that, when cut with Cas9, produces distinct “scars” when repaired by either NHEJ, MMEJ, or SSTR; high-throughput sequencing of these scars provides highly accurate measurements of the relative contributions of the three pathways. Second, we used a modification of our TRIP (thousands of reporters integrated in parallel) method (Akhtar et al., 2013) to insert this reporter into >1,000 random genomic locations, tracking each individual reporter in parallel by molecular barcoding. We thus systematically measured the relative contributions of NHEJ, MMEJ, and SSTR as a function of chromatin context in >1,000 genomic locations. This yielded datasets that (1) comprehensively sample the broad diversity of chromatin “flavors” across the entire genome, (2) bypass the confounding effects of varying sequence context, and (3) probe the three pathways with high accuracy and sensitivity. The results provide a detailed view of the impact of chromatin context on the relative activities of the three repair pathways.

RESULTS

Multiplexed DSB repair pathway assay: Principle

We developed a strategy to measure the relative contribution of several DSB repair pathways in more than 1,000 genomic locations that include all known common chromatin states. The strategy involves a pathway-specific reporter construct that contains a short DNA sequence (derived from the human *LBR* gene) that predominantly produces a +1 insertion or a –7 deletion when cut at a specific base pair position by Cas9 (Figure 1A). We previously found that these two indels are primarily the result of NHEJ (+1) and MMEJ (–7), respectively (Brinkman et al., 2018), and we provide additional support below. The relative abundance of these signature indels can therefore be interpreted as a measure of the relative contribution of these two pathways. Furthermore, this readout can be extended to include SSTR and

translocations (see below). We note that HR cannot be detected with this assay, because HR generally repairs DSBs perfectly, and perfectly repaired DNA cannot be distinguished from uncut DNA. However, we previously estimated that perfect repair of this reporter sequence is rare and by inference that the contribution of HR is likely to be very minor (Brinkman et al., 2018). Further evidence supporting this inference is presented below.

With this reporter, we implemented a variant of the TRIP technology (Akhtar et al., 2013) to systematically probe the effects of many chromatin environments on the repair pathway use. We inserted the reporter sequence into a PiggyBac transposon vector, together with a 16 bp random barcode sequence that was located 56 bp from the DSB site (Figure 1B). We then randomly integrated this construct into the genomes of pools of K562 cells (Figure 1C). We chose K562 cells because the chromatin landscape has been extensively characterized. From one of these pools, we also generated several clones for smaller scale experiments. Each copy of the integrated reporter carried a different barcode. We mapped the genomic locations of these integrated pathway reporters (IPRs) together with their barcode sequences by inverse PCR (Akhtar et al., 2013). Next, after Cas9-mediated DSB induction and the ensuing repair, we determined the accumulated spectrum of indels of each individual IPR in a multiplexed fashion, by PCR amplification (see primer locations in Figure 1C) followed by high-throughput sequencing. Because each barcode is linked to its genomic location, the sequence information enabled us to infer the relative DSB repair pathway use at each location. Comparison of the resulting data with the local chromatin state of the IPRs then provides insight into the impact of chromatin context on DSB repair pathway use.

Implementation and validation of the multiplexed reporter assay

For these experiments we used a human K562 cell line that expresses Cas9 protein in an inducible manner (Brinkman et al., 2018). We generated two cell pools with 979 and 1,099 (total 2,078) uniquely mapped IPRs (Figures 2A and S1A). In addition, we established 14 clonal cell lines, for which the barcodes and locations were also mapped by inverse PCR and tagmentation (Stern, 2017). On average these clones carried 6.8 integrations, which we take as an estimate of the numbers of integrations per cell in the cell pools. For some additional analyses described below, we used one clone (clone 5) with 19 mapped IPRs that are located across most major chromatin types (Figures S1A and S1B).

Next, we induced Cas9 in the cell pools by ligand-dependent stabilization (Banaszynski et al., 2006) and transfection with the single guide RNA (sgRNA) (named sgRNA-LBR2). We collected genomic DNA after 64 h and determined the indel spectra of all IPRs. At this time point, indel accumulation in the *LBR* gene has reached near saturation (Brinkman et al., 2018). After applying stringent quality criteria (see STAR Methods), we obtained robust indel spectra of 1,229 IPRs (Figure S1A); the other IPRs were mostly discarded because they were insufficiently represented in the cell pools. Overall, the IPRs of both pools showed a similar pattern of indels as the endogenous *LBR* sequence in the same cells, dominated by +1 and –7 indels (Figures 2B, 2C, S1C, and S1D). This supports previous findings that

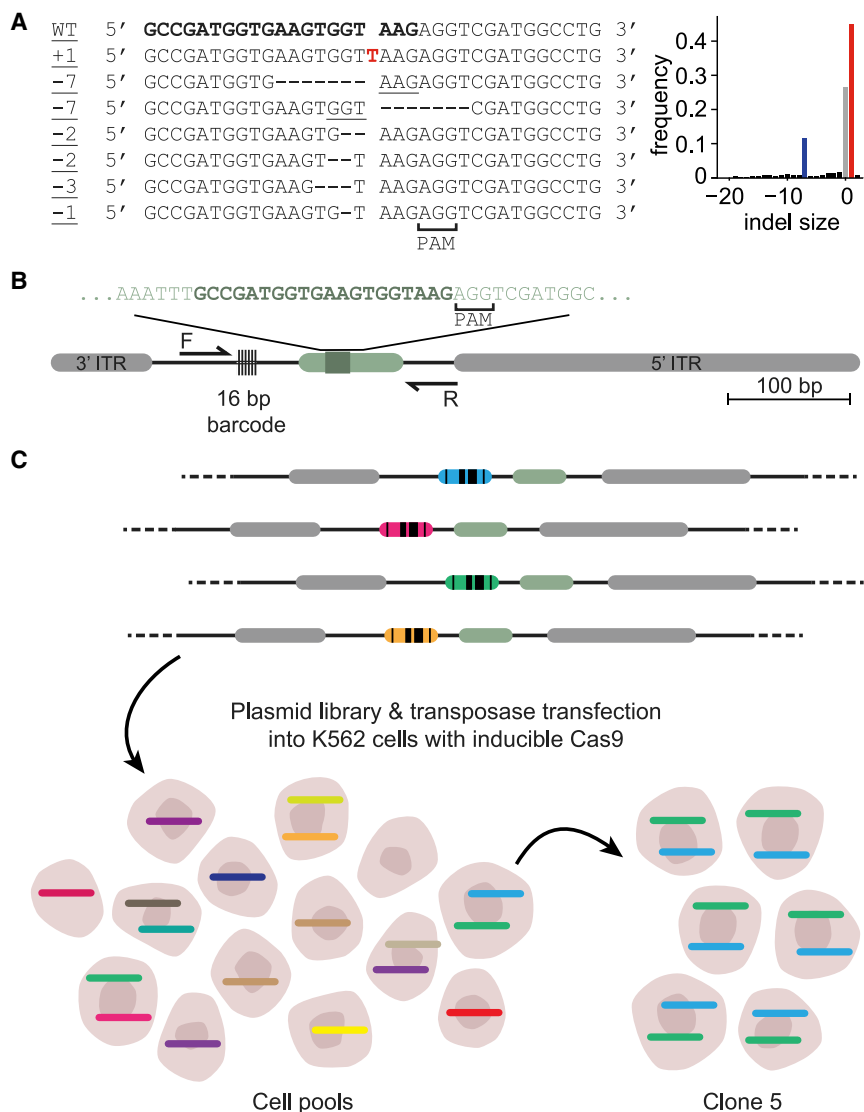


Figure 1. Principle of multiplexed DSB repair pathway reporter assay

(A) Left panel: sequences of the most common insertions and deletions produced after break and repair induced with sgRNA LBR2. Inserted nucleotide in red, microhomologies are underlined. Right panel: indel frequency distribution after cutting the *LBR* gene. Negative indel sizes refer to deletions, positive sizes refer to insertions. The +1 and -7 are marked in red and blue, respectively.

(B) Schematic of the TRIP construct. ITR, inverted terminal repeat of PiggyBac transposable element; the *LBR* gene-derived sequence is shown in light green, with the sgRNA target sequence in dark green. PCR primers are indicated by the arrows (F and R).

(C) Schematic of the TRIP experimental setup. See main text.

the sequence determines the overall indel pattern (Allen et al., 2018; Chakrabarti et al., 2019; Chen et al., 2019; Shen et al., 2018; van Overbeek et al., 2016), but we also observe clear variations in indel frequencies.

As noted before (Brinkman et al., 2018), the -7 deletions come in two variants that both involve 3 nt microhomologies (Figure 1A), consistent with MMEJ. To further verify that the +1 and -7 indels indeed represent NHEJ and MMEJ, respectively, we depleted or inhibited several pathway-specific proteins (Chang et al., 2017; Scully et al., 2019) in either the pools or clone 5 (Figure S2). The +1 insertion was strongly reduced, and the -7 deletion was increased by inhibition of DNA-PKcs by the compounds NU7441 (Figures S2A and S2B) or M3814 (Figure S2C). DNA-PKcs is a key component of NHEJ (Gottlieb and Jackson, 1993). In contrast, the -7 deletions but not the +1 insertion were selectively reduced upon depletion of DNA polymerase theta (POLQ) and CTIP (also known as RBBP8), which are pro-

NHEJ and MMEJ, respectively, and that the contribution of HR is minor at Cas9-induced DSBs.

Detection of large genomic rearrangements

The sequencing of the repair “scars” indicated that large indels are very rare (Figure S1C). However, because of the length of the sequence reads, we could not capture any deletions >119 bp, nor translocations and other rearrangements that might occur. We therefore modified a tagmentation-based approach (Gianoukos et al., 2018; Stern, 2017) to identify distal genomic sequences that became ligated to the IPRs after Cas9 activation (Figure S2J). We applied this assay to clone 5. The results indicate that large indels, inversions, and translocations occur with a relative frequency of approximately 4% (Figure S2K). As may be expected (Frock et al., 2015), these rearrangements preferentially occur in *cis*, with a bias for shorter distances (Figures S2L and S2M). The *LBR* gene, which is also cut by sgRNA-LBR2, is

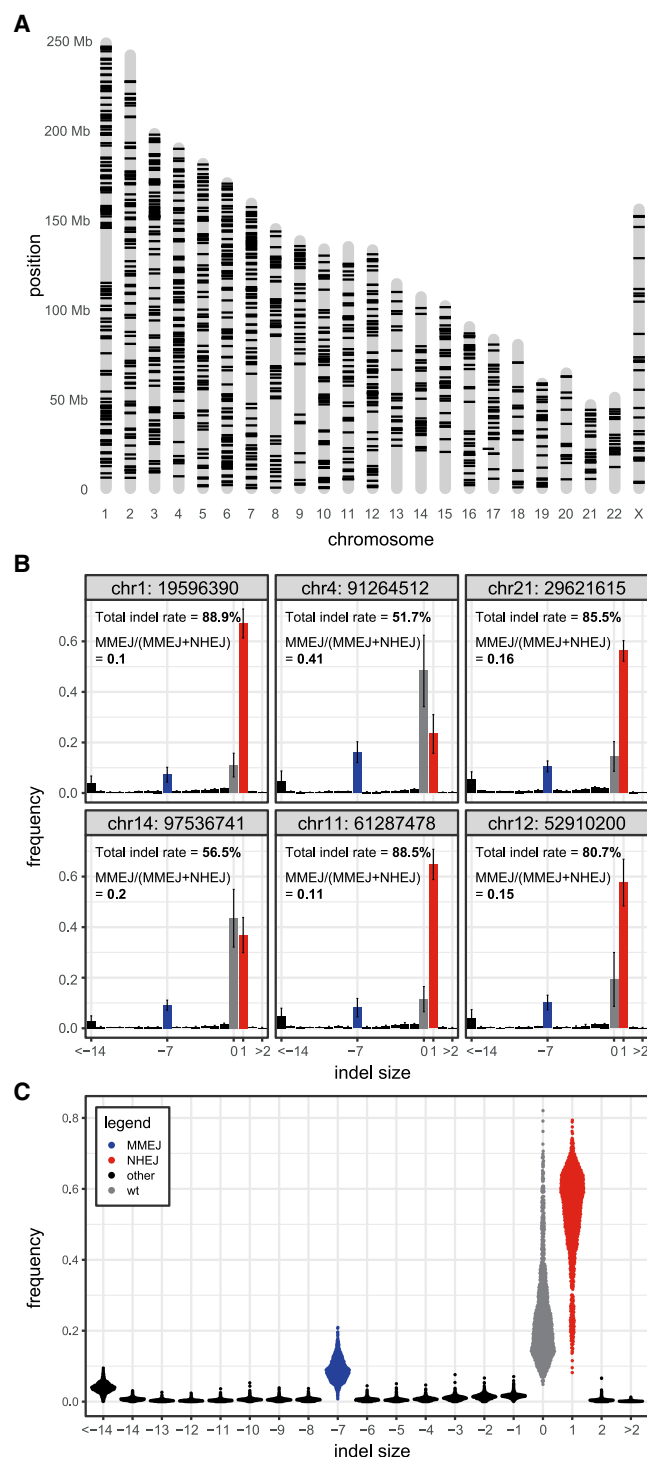


Figure 2. Multiplexed detection of DSB repair pathway use

(A) Genomic integration coordinates of 1,229 uniquely mapped IPRs (both cell pools combined) that passed filtering as described in STAR Methods and are used in this work.

(B) Indel frequency distributions of six randomly selected IPRs, 64 h after Cas9 induction. Data are average of six independent replicates. Error bars are \pm SD. Gray, wild-type sequence; red, +1 insertion, diagnostic of NHEJ; blue, -7 deletion, diagnostic of MMEJ; black, other indels.

involved in about 22.2% and 16.4% of the long-range rearrangements with IPRs in *cis* and *trans* respectively (Figures S2L and S2M). This is particularly frequent for an IPR located ~0.5 Mb away on the same chromosome. Unfortunately, because of a technical limitation (see STAR Methods) we were unable to reliably detect junctions between two IPRs. However, considering that there are four alleles of the *LBR* gene in K562 cells (Brinkman et al., 2018) and an estimated 6.8 IPRs/cell in our cell pools, it seems unlikely that IPR-IPR junctions occur with a total frequency of more than 1% (see STAR Methods). Although the detection of rearrangements by this approach may not be fully quantitative because of possible technical biases, we conclude that IPRs are involved in inversions, large deletions, and translocations. Yet the frequency of these rearrangements is generally too low to substantially skew our estimates of MMEJ and NHEJ.

Effects of chromatin context on overall indel frequencies

Using the indel spectra from the two cell pools, we first investigated the impact of chromatin context on total indel frequencies (TIFs; i.e., the proportion of reporter sequences carrying any type of indel). Across the IPRs these frequencies varied from ~25% to ~100% (Figures 3A and 3B). This variation most likely reflects differences in either the cutting efficiency by Cas9 or the DSB repair rate, or both. The fact that in some IPRs the TIFs approached 100% indicates that virtually all cells received the sgRNA in our transfection protocol. We repeated these experiments with three different sgRNAs targeting different sequences in the same reporter (Figures 3A, 3B, and S3A). This yielded overall indel frequencies that strongly correlated with sgRNA-LBR2 and with each other, although with sgRNA-LBR2 we may have approximated saturation of indels more than with the other sgRNAs (Figures 3B, 3C, and S3B).

Because the sequences of the IPRs are identical (except for the short barcodes located 56 bp from the cut site), the differences in indel frequencies across integration sites are presumably due to variation in the local chromatin environment. To investigate this, we correlated the indel data for each sgRNA with a curated set of 24 genome-wide maps of chromatin features that represent most of the known main chromatin types, including a multitude of markers of transcription and active regulatory elements; chromatin accessibility; DNA methylation; and heterochromatic features such as the histone modifications H3K27me3, H3K9me2, late replication, and nuclear lamina interactions (Tables S1 and S2) (Chen et al., 2018; Dekker et al., 2017; ENCODE Project Consortium, 2012; Leemans et al., 2019; Ott et al., 2018; Salzberg et al., 2017; Schmidl et al., 2015; Schwalb et al., 2016; Shah et al., 2018). The IPRs lack gene regulatory elements and are only 640 bp long and may thus be expected to adopt the local chromatin state. Chromatin immunoprecipitation (ChIP) experiments for three histone modifications generally confirmed this (Figure S3C). This is consistent with previous studies showing that integrated reporters adopt and strongly

(C) Indel frequencies of all IPRs shown in (A), 64 h after Cas9 induction. Data are average of two to six independent replicates.

See also Figures S1 and S2.

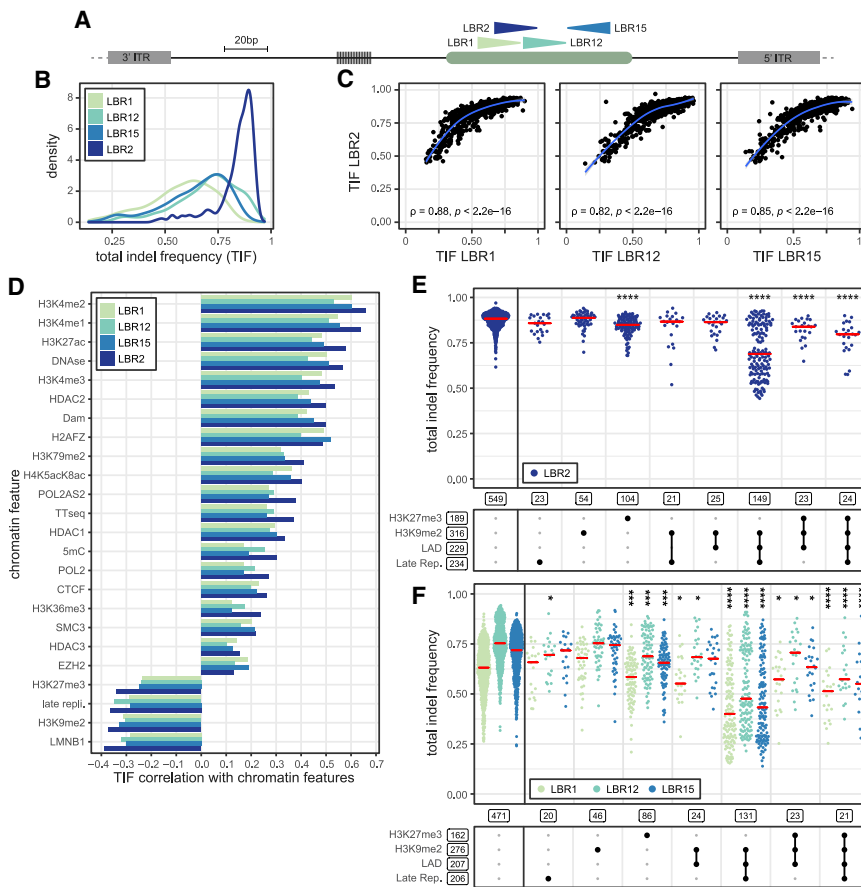


Figure 3. IPR total indel frequency varies as a function of chromatin context

(A) Schematic of the IPR with four different gRNA target sites indicated by arrowheads, oriented toward the PAM.

(B) Total indel frequency distributions of IPRs after cutting with each sgRNA, shown as density plots, which can be interpreted as smoothed histograms. For each sgRNA, IPRs were included only if they yielded reliable data in at least two independent experiments (LBR2: 1,010 IPRs [n = 2–8]; LBR1: 956 IPRs [n = 2]; LBR12: 942 IPRs [n = 2]; LBR15: 932 IPRs [n = 2]). Indel frequencies were not corrected for transfection efficiency, which we infer to be close to 100% because some IPRs have indel frequencies near 100%.

(C) Scatterplots of total indel frequencies obtained with LBR2 versus the three other sgRNAs. ρ is Spearman's rank correlation coefficient.

(D) Spearman's correlations between total indel frequencies in IPRs and the local intensities of 24 chromatin features at the IPR integration sites, for each sgRNA. $p < 0.01$ for all correlations. Chromatin features are ordered by the LBR2 correlation coefficients.

(E) Total indel frequency at each IPR obtained with LBR2 sgRNA, split into different combinations of heterochromatin features present, as indicated by black dots in the scheme below the graph. Red lines show median values. Boxed numbers indicate the number of IPRs in each group; only groups with >20 IPRs are shown. Asterisks mark p values according to Wilcoxon test, compared with euchromatin IPRs (leftmost column): * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

(F) Same as (E) but for the other three sgRNAs. See also Figure S3.

respond to the local chromatin state (Akhtar et al., 2013; Corrales et al., 2017; Leemans et al., 2019). We therefore assume that the chromatin state of the integration positions is a reasonable approximation of the chromatin state of the IPRs themselves.

Indel frequencies at the IPRs generally correlated positively with various markers of euchromatin and negatively with markers of heterochromatin. These correlations were highly consistent between the four sgRNAs that we tested (Figure 3D); minor differences may reflect differences in statistical power or subtle modulating effects of the DNA sequence at the broken ends. On average, IPRs integrated in heterochromatin regions showed lower indel frequencies than in euchromatic regions. However, within heterochromatin, the magnitude of this effect varied depending on the specific combination of features (Figures 3E and 3F). The most pronounced effect was observed in regions marked by the combination of H3K9me2, lamina-associated domains (LADs), and late replication (further referred to as triple heterochromatin). In these regions the distribution of indel frequencies is bimodal. This is similar to TRIP reporters of promoter activity, which also tend to show a bimodal distribution of activities across LADs, correlating with local chromatin features (Leemans et al., 2019). Differences in lamina interactions and replication time could explain part of the bimodal distribution of indel frequencies (Figure S3D). Remarkably, when H3K27me3 is additionally present, the reduction of indel frequencies is less

pronounced, and regions marked by H3K27me3 alone only slightly affect indel frequencies compared with euchromatin. Thus, H3K27me3 only mildly impedes Cas9 editing and may even counteract the effects of other heterochromatin features. Regions marked by H3K9me2 together with either late replication or lamina interactions (but not both) show only marginally reduced indel frequencies, compared with the triple-marked regions (Figures 3E and 3F). For euchromatic regions, we did not survey combinatorial effects, because there are too many possible combinations and hence statistical power is insufficient.

Together, these results indicate that the overall indel frequency depends on the local chromatin context and that heterochromatin features are correlated with the efficiency of indel accumulation in a combinatorial manner. These effects may be through modulation of Cas9 cutting efficiency, modulation of indel-forming repair rates, or both.

Impact of chromatin context on MMEJ:NHEJ balance

Next, we analyzed the variation in the balance between MMEJ and NHEJ, throughout this paper referred to as “MMEJ:NHEJ balance” and defined for each IPR as the number of -7 reads over the sum of -7 and $+1$ reads. Importantly, this metric intrinsically corrects for any differences in cutting efficiencies, because it scores only sequences that were broken and

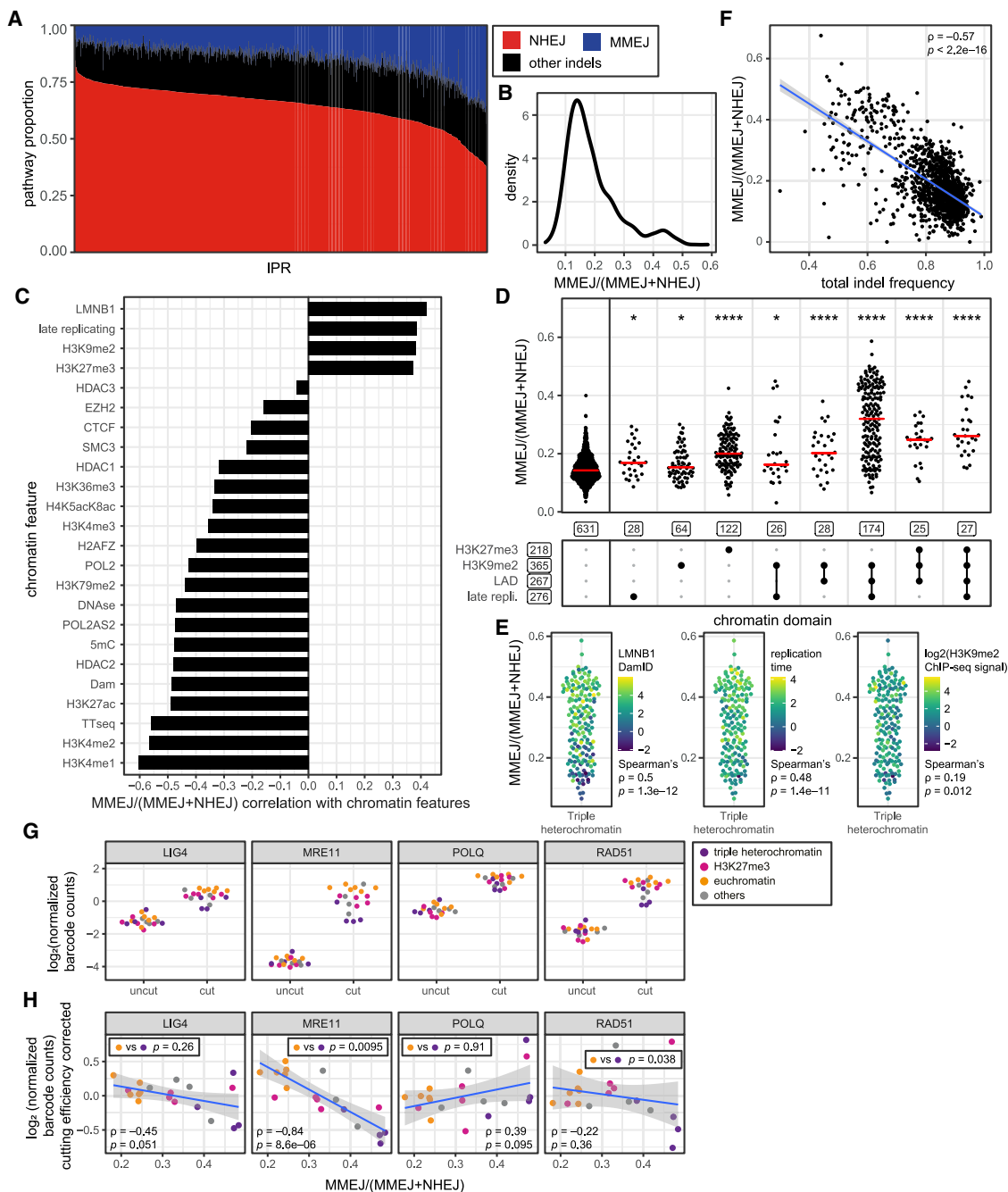


Figure 4. Chromatin context effects on MMEJ:NHEJ balance and protein binding

(A) Variation in indel composition across IPRs. Red, +1 (NHEJ); blue, -7 (MMEJ); black, other indels (unknown pathway). IPRs are ordered by +1 insertion frequency (1,171 IPRs, two to eight independent experiments).

(B) MMEJ:NHEJ balance distribution across all IPRs.

(C) Spearman's correlation coefficients of MMEJ:NHEJ balance versus the local intensities of 24 chromatin features. $p < 0.001$ for all correlations.

(D) MMEJ:NHEJ balance per IPR, split into different combinations of heterochromatin features similar to Figure 3E; see Figure S4D for all groups. Asterisks mark p values according to the Wilcoxon test, compared with euchromatin IPRs (leftmost column): * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

(E) MMEJ:NHEJ balance of IPRs in triple heterochromatin, colored by levels of H3K9me2, LMNB1, or replication timing.

(F) Correlation between total indel frequency and MMEJ:NHEJ balance across all IPRs.

(G) Barcode counts (normalized \log_2 values) of IPRs in clone 5 after ChIP of indicated proteins, 16 h after Cas9 induction with (cut) or without (uncut) sgRNA LBR2.

(legend continued on next page)

repaired. The MMEJ:NHEJ balance varies profoundly depending on the integration site (Figures 4A, 4B, and S4A). Globally, it correlates negatively with markers of euchromatin and positively with markers of heterochromatin (Figures 4C and S4B). The strongest negative correlations were observed for H3K4me1, H3K4me2, and H3K27ac, which are histone modifications that primarily mark enhancers and to a lesser extent promoters (Gasperini et al., 2020), and TT-seq, which measures transcription activity (Schwalb et al., 2016). Positive correlations occur with multiple markers of heterochromatin such as H3K27me3, H3K9me2, and LADs. Thus, in heterochromatin the MMEJ:NHEJ balance is generally higher than in euchromatin.

Cutting of the IPRs by Cas9 in combination with each of the other three sgRNAs yields rather complex indel spectra (Figure S3A). When we provisionally assigned each of these indels to either MMEJ or NHEJ (on the basis of their sensitivity to M3814 treatment; see STAR Methods), we observed consistently that the MMEJ:NHEJ balance is higher in heterochromatin than in euchromatin (Figure S4C). Thus, the results are consistent among all sgRNAs tested, although sgRNA LBR2 is more suited to measure the two pathways.

Within heterochromatin, we further explored whether certain combinations of chromatin features are more predictive than others. The strongest effect on the balance was observed in IPRs located in regions marked by triple heterochromatin (Figures 4D and S4D). As for the TIF, the bimodal distribution of the triple heterochromatin group can be partly explained by local differences in levels of lamina interactions and replication timing (Figure 4E). Regions marked by two of these three features showed less pronounced but significant increases in MMEJ:NHEJ balance compared with euchromatin regions, as did regions marked by H3K27me3. Altogether, these data show that the balance between MMEJ and NHEJ is broadly linked to the global heterochromatin/euchromatin dichotomy, and within heterochromatin depends on the local combination of heterochromatin features.

Overall, the patterns of MMEJ:NHEJ balance (Figure 4C) and indel frequencies (Figure 3D) appeared to mirror each other. Indeed, across all IPRs the two variables correlate, although imperfectly (Spearman's correlation coefficient = -0.57) (Figure 4F). This correlation does not necessarily imply a causal relationship. But it seems improbable that the cutting rate itself determines the pathway balance.

Repair protein binding is linked to pathway balance and chromatin context

To test whether the variation in pathway balances across the IPRs may be explained by differences in presence of pathway-specific repair proteins, we conducted ChIP experiments on clone 5 cells. We probed the occupancy of several proteins at the IPR barcodes (i.e., close to the break sites) within the resolution of ChIP. We focused on time point 16 h after Cas9 induction, at which

we previously found the largest amount of broken DNA (Brinkman et al., 2018). The results show that MRE11 (involved in early DSB detection and processing), LIG4 (specific to NHEJ), POLQ (specific to MMEJ), and RAD51 (specific to HR) can all be detected at each IPR when Cas9 cutting is induced (Figure 4G). Thus, the binding of these proteins appears to be universal across different chromatin contexts. However, correction of the ChIP signals for the approximate cutting frequency (see STAR Methods) revealed quantitative trends. As expected, the levels of LIG4 and POLQ at the IPRs are negatively and positively correlated with the MMEJ:NHEJ balance, respectively, although these trends are of borderline statistical significance (Figure 4H). Surprisingly, binding of MRE11 shows a strong negative correlation with the MMEJ:NHEJ balance and a significant preference for euchromatin compared with triple heterochromatin. Finally, binding of RAD51 shows no significant correlation with the MMEJ:NHEJ balance, although it shows an overall preference for euchromatin over triple heterochromatin. We conclude that proteins of all major pathways are present at each DSB, with quantitative differences that may in part explain the observed pathway preferences.

Different kinetics of MMEJ engagement between chromatin types

To explore how the difference in pathway balance between heterochromatin and euchromatin develops over time after DSB induction, we conducted time-series experiments. We used a robotics setup to collect DNA samples every 3 h over a period of 3 days following Cas9 activation. For these experiments we focused on clone 5; because all 19 IPRs in this clone are in the same cell, their repair kinetics can be directly compared.

As expected, Cas9 activation resulted in a gradual accumulation of +1 and -7 indels in all IPRs, concomitant with a loss of wild-type sequence (Figures 5A and S5). These kinetics were generally slower in regions in triple heterochromatin (e.g., IPRs 7, 14, and 16; Figure S5). Remarkably, the MMEJ:NHEJ balance was not constant over time but was strongly skewed toward NHEJ at the early time points and gradually shifted toward MMEJ for all IPRs, culminating in a plateau approximately 50 h after Cas9 activation (Figure 5B). This points to a delayed use of the MMEJ pathway, as we had observed previously for a single locus (Brinkman et al., 2018). This buildup of MMEJ use may occur eventually at most DSBs. However, over time the pathway balance diverged between IPRs in different chromatin environments, with nearly all heterochromatic IPRs developing a higher MMEJ:NHEJ balance than the euchromatin IPRs (Figure 5B).

Overall robustness of pathway balance in heterochromatin

We then investigated the role of several heterochromatin features in pathway balance by perturbation experiments. To distinguish direct from indirect effects, we compared the MMEJ:NHEJ

(H) Same ChIP data as (G) under cut conditions, but now corrected for estimated cutting efficiencies at $t = 72$ h and plotted against MMEJ:NHEJ balance. Bottom p values refer to the Pearson's correlation with MMEJ:NHEJ balance; top p values refer to the difference between IPRs in euchromatin and triple heterochromatin (Wilcoxon test); blue line and gray shading show linear regression fit with 95% confidence interval. Triple heterochromatin is the combination of H3K9me2, lamina association, and late replication; euchromatin is here defined as the absence of any heterochromatin features. See also Figure S4.

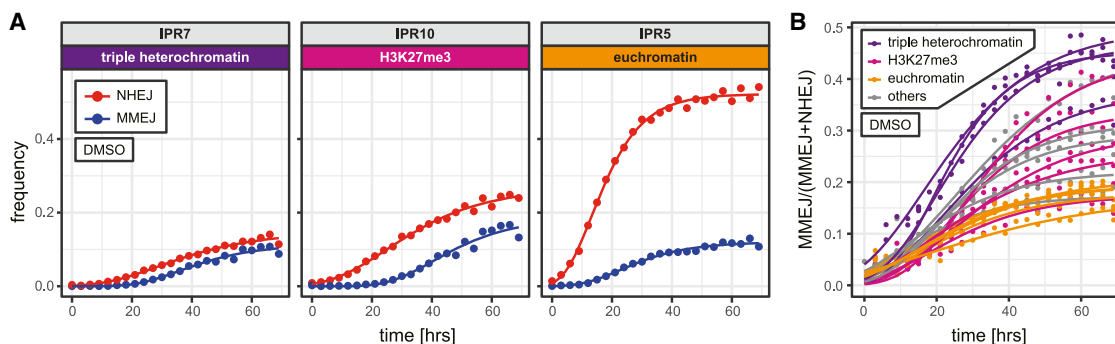


Figure 5. Accumulation of indels over time

(A) Time curves of the +1 insertion (red) and -7 deletion (blue) in single IPRs located in three different types of chromatin. See Figure S5 for plots of all 19 IPRs in clone 5. Dots are measured values; lines are fitted sigmoid curves.

(B) Shifting MMEJ:NHEJ balance over time in 19 IPRs of clone 5, colored by chromatin type. Data in (A) and (B) are averages of two independent experiments. See also Figure S5.

balance of IPRs in regions with the to-be-perturbed feature to that of IPRs in regions already lacking the feature prior to treatment. Direct effects should primarily alter the MMEJ:NHEJ balance in regions originally marked by the feature.

We first reduced the levels of H3K9me2 by treatment with the G9a inhibitor BIX01294 (Figures S6A–S6C). This did not alter the MMEJ:NHEJ balance in H3K9me2 domains, except when in combination with LADs and late replication, where the balance increased slightly ($p = 0.02$; Figure S6A).

We then tested the effect of GSK126, a compound that inhibits the H3K27me3 methyltransferase EZH2 and causes a global loss of H3K27me3 (Figures S6D and S6E). This inhibitor caused a significant reduction of the MMEJ:NHEJ balance in H3K27me3-only domains compared with euchromatin regions ($p = 2e-11$), as well as in virtually all domain combinations that include H3K27me3 (Figure 6A). Because the IPRs in the H3K27me3 domains are in the same cell pools and receive exactly the same drug treatment as the euchromatic IPRs, the effect of GSK126 must be local in the genome; as GSK126 reduces H3K27me3 levels, we conclude that its effect in the H3K27me3 domains is direct. Unexpectedly, GSK126 treatment also reduced levels of H3K9me2 (Figures S6B and S6C) and also slightly lowered the MMEJ:NHEJ balance in the triple heterochromatin domains ($p = 5.6e-05$), but not in the single H3K9me2 domains. The most prominent shift in balance was, however, in H3K27me3 domains, pointing to a local effect of this histone modification on MMEJ:NHEJ balance.

Finally, because IPRs in LADs often show a high MMEJ:NHEJ balance, we used CRISPR-Cas9 editing to derive cell lines from clone 5 that lacked Lamin A/C (LMNA) or Lamin B receptor (LBR) (Figures S6F–S6I). These two lamina proteins are important for the peripheral positioning of heterochromatin (Clowney et al., 2012; Solovei et al., 2013), and LMNA has been implicated in the control of NHEJ by sequestering 53BP1 (Redwood et al., 2011). Using the pA-DamID method, we mapped genome-wide changes in lamina interactions in four knockout (KO) clones each of LMNA and LBR. LMNA-KO cells showed very few changes in lamina interactions, while the LBR-KO clones showed many regions with either gains or losses in lamina inter-

actions. A detailed analysis of these changes will be reported elsewhere. Here, we investigated whether changes in lamina interactions of the IPRs coincide with changes MMEJ:NHEJ balance.

The majority of the IPRs did not undergo substantial changes in lamina interactions in either the LMNA- or LBR-KO clones compared with the parental clone 5, and they also did not show significant changes in MMEJ:NHEJ balance (Figure S6J). An exception was IPR 2, in which the lamina interactions became stronger in all four LBR-KO clones (Figures 6B and 6C). However, the MMEJ:NHEJ balance in IPR 2 was not detectably altered in these clones (Figures 6C and S6K). This suggests that lamina contacts do not modulate this balance, but we cannot rule out that effects on this balance emerge only when lamina contacts are stronger than those of IPR 2 in the LBR-KO clones (note that the lamina interaction Z score in these clones is about 0, which corresponds to a moderate level of lamina interactions). Interestingly, IPR 17 showed a marked increase in the MMEJ:NHEJ balance in two of the four LBR-KO clones (Figures 6D, 6E, and S6J). However, for this IPR the lamina interactions did not change (Figure S6L), and we do not understand why only two of the four clones show this behavior. Nevertheless, this result underscores that it is possible to shift the MMEJ:NHEJ balance in an IPR markedly without any change in its sequence. Presumably, an unknown change in the local chromatin state in the two clones is responsible for this.

Together, these data indicate that the MMEJ:NHEJ balance in specific heterochromatin types is not easily shifted by targeting individual key markers of the respective heterochromatin types. Pathway balance may be redundantly controlled by multiple factors in each heterochromatin type. Nevertheless, depletion of H3K27me3 did cause a detectable reduction in MMEJ:NHEJ balance in heterochromatin domains that normally carry this mark.

Impact of chromatin context on SSTR

Finally, we investigated a third repair pathway, SSTR, which is commonly used to create specific mutations by CRISPR-Cas9 editing. We hypothesized that this pathway may also be modulated by the local chromatin environment. To test this, we

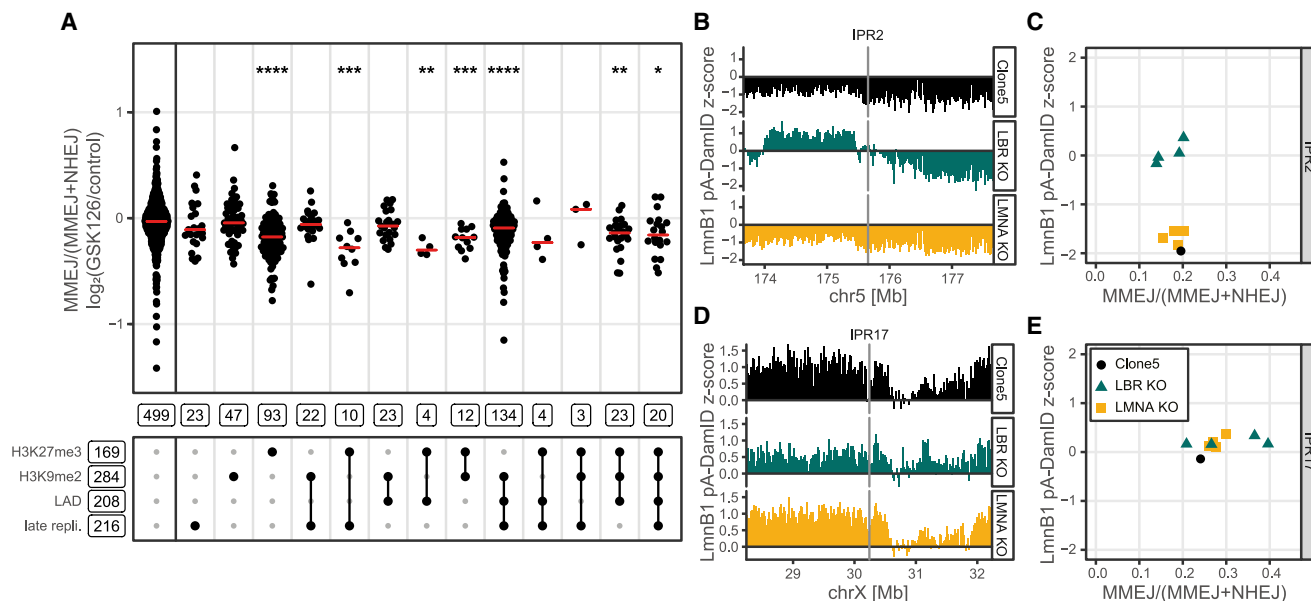


Figure 6. Effects of heterochromatin perturbations on pathway balance

(A) \log_2 fold change of MMEJ:NHEJ balance in GSK126 treated cells compared with control cells, for 19 IPRs divided by heterochromatin type. Data are average of two independent biological replicates. Wilcoxon test compared with euchromatin IPRs (leftmost column): * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$.

(B and D) Nuclear lamina interaction tracks around IPR2 (B) and IPR17 (D). The tracks for the KO clones are the average of four separate clones (individual tracks are shown in Figures S7K and S7L). All data are average of two independent biological replicates.

(C and E) Comparison of MMEJ:NHEJ balance ($n = 3$) and average lamina interaction score in a 20 kb window centered on the IPR ($n = 2$), for IPR2 (C) and IPR17 (E).

See also Figure S6.

triggered DSB formation in our reporter sequence in the presence of a template ssODN containing a specific +2 insertion (ssODN insertion) (Figure S7A). We designed this insertion within the PAM site, so that a successful editing event destroyed the PAM site and prevented further cutting by Cas9. We then transfected the IPR cell pools with this ssODN (together with the sgRNA) to probe the impact of chromatin context on the relative contribution of SSTR, NHEJ, and MMEJ in parallel. In these experiments, we found that a median 6% of the indels consisted of the SSTR insertion (Figure S7B). Accumulation of this insertion was mostly at the expense of the -7 deletion but not the +1 insertion (Figures 7A, S7C, and S7D), suggesting competition between SSTR and MMEJ. Indeed, depletion of POLQ (a key protein of MMEJ) caused an increase in ssODN insertions (Figures S2D and S2E). Consistent with an earlier study (Richardson et al., 2018), we find that knockdown of RBBP8, a factor that promotes end-resection (Sartori et al., 2007), strongly reduces SSTR and MMEJ usage (Figures S2D and S2E). This suggests that DSB end-resection is an early step in SSTR. In agreement with previous work (Richardson et al., 2018), we conclude that SSTR is distinct from NHEJ and MMEJ.

We then investigated the balance of the three simultaneously probed pathways as a function of chromatin type. In the presence of ssODN, the proportion of indels created by SSTR is inversely correlated with the overall indel frequency (Figures 7B and 7C) and is higher in various types of heterochromatin compared with euchromatin (Figure 7D). Under this condition,

the proportions of indels produced by MMEJ (Figure 7E) and NHEJ (Figure 7F) show similar effects of various types of heterochromatin as we observed in the absence of ssODN (cf. Figure 4D). We conclude that the proportion of SSTR is higher in several types of heterochromatin compared with euchromatin.

DISCUSSION

Genome-wide survey of DSB repair

Here, we present a powerful reporter system to query effects of chromatin on DSB pathway use. It consists of (1) a simple short DNA sequence that, when cut with Cas9, produces a signature indel for three repair pathways and (2) an adaptation of the TRIP multiplexed reporter assay (Akhtar et al., 2013). In combination, these tools offer precise measurements of the relative contribution of NHEJ, MMEJ, and SSTR, combined with the throughput that is needed to query the impact of a wide diversity of chromatin contexts. The sequencing-based “scar-counting” readout renders the assay highly quantitative. Furthermore, because the same reporter sequence is integrated throughout the genome, differences in cutting and repair must be due to the local chromatin context of the integration sites.

The PiggyBac transposable element is not known to use host DNA repair factors (Mitra et al., 2008) that could bias its integration across the genome, but it shows a ~ 3 -fold preference for transcriptionally active regions (Akhtar et al., 2013). Otherwise, integration of this element is thought to be largely random, and

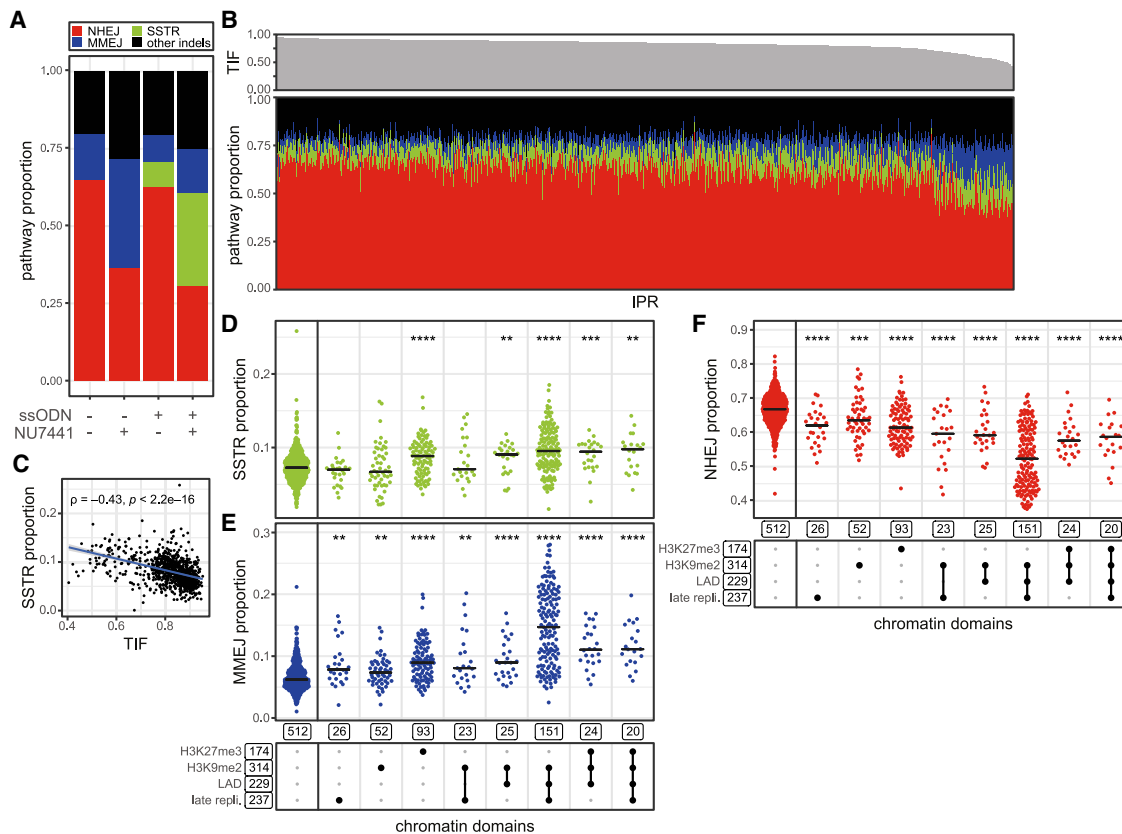


Figure 7. Balance between NHEJ, MMEJ, and SSTR in euchromatin and heterochromatin

(A) Average pathway contribution across all IPRs in the cell pools, in the absence or presence of an ssODN donor, and with or without NU7441 treatment (DMSO: 1,219 IPRs [n = 2–8, same data as in Figures 3 and 4]; DMSO + ssODN: 1,032 IPRs [n = 2 or 3]; NU7441: 1,186 IPRs [n = 2–4, same data as in Figures S2A and S2B]; NU7441 + ssODN: 1,047 IPRs [n = 2 or 3]). Red, +1 insertion (NHEJ); blue, –7 deletion (MMEJ); green, +2 insertion due to SSTR; black, other indels. (B) Pathway contributions in all individual IPRs in the presence of ssODN donor but without NU7441 (bottom panel), sorted by overall indel frequency (top panel). (C) Correlation between total indel frequency and SSTR proportion across all IPRs. Blue line and gray shading show linear regression fit with 95% confidence interval. ρ is Pearson's rank correlation coefficient. (D–F) Proportions of total indels generated by SSTR (D), MMEJ (E), and NHEJ (F) in the presence of the ssODN donor, split according to heterochromatin features as indicated in the bottom panel, similar to Figure 3E. Asterisks mark p values according to the Wilcoxon test, compared with euchromatin IPRs (most left column): *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001. See also Figure S7.

the large number of IPRs provides enough statistical power to compare all major chromatin types.

A key aspect of our analysis is the comparison of the relative pathway activities across chromatin types. For this we used pathway balance metrics that are intrinsically corrected for cutting efficiencies. Because these metrics are based on specific, well-characterized indels (–7, +1, and +2) and disregard the (relatively low abundant) other indels with uncertain pathway origin, these metrics should not be interpreted as definitive measures of relative pathway activities; however, they can be used to compare pathway balances between different chromatin contexts in the same cell pools.

Effects of chromatin context on DNA repair pathway preference

We found that the balance between MMEJ and NHEJ varies >5-fold across chromatin contexts. Generally, in heterochromatin

we observe a higher MMEJ:NHEJ balance than in euchromatin, but this shift depends on the precise heterochromatin features that are present. A previous study of smaller scale (Kallimasioti-Pazi et al., 2018) found that the indel spectrum after a Cas9-induced cut was not affected by imprinted heterochromatin, a type of heterochromatin that we did not probe. This underscores the notion that different types of heterochromatin may have distinct effects on repair pathway balance.

Two classes of models may explain the chromatin effects that we observed. These models are not mutually exclusive. First, it is possible that euchromatin carries one or more features that activate or bind the NHEJ machinery, or conversely, that certain heterochromatin features promote MMEJ. If the latter is true, then it should be considered that multiple heterochromatin features can play such a role, as H3K27me3 modulates the MMEJ:NHEJ balance in H3K27me3-marked heterochromatin, but this does not explain the high MMEJ:NHEJ balance in triple heterochromatin.

By ChIP we found that MRE11 is enriched at euchromatic DSBs relative to heterochromatic DSBs. MRE11 has been implicated in both MMEJ and NHEJ (reviewed in [Reginato and Cejka, 2020](#)), but perhaps its role is quantitatively more important for NHEJ than for MMEJ, which would cause a reduced MMEJ:NHEJ balance at euchromatic sites where it is more abundant.

A second class of models involves the differential accessibility of heterochromatin and euchromatin. By default (particularly in G1 phase), NHEJ may be globally more active than MMEJ. Hence, at a DSB in “open” euchromatin, NHEJ may mostly outcompete MMEJ. In contrast, in heterochromatin a DSB may be inaccessible to either pathway until the heterochromatin is de-compacted. This remodeling of heterochromatin may be a slow process, which would allow time for MMEJ to be activated and give both MMEJ and NHEJ a more similar chance to repair the DSB. Indeed, DSB-induced unfolding of heterochromatin has been reported ([Chiolo et al., 2011](#); [Goodarzi et al., 2008](#); [Jakob et al., 2011](#); [Janssen et al., 2016](#); [Ryu et al., 2015](#); [Tsouroula et al., 2016](#)). However, as seen by microscopy, this remodeling process occurs within ~20 min, while we found that the MMEJ indels accumulate only after several hours. It is possible that additional biochemical or structural changes in heterochromatin are involved that take place over a timescale of hours. We also considered that one early cutting event in euchromatin may trigger slow upregulation of MMEJ activity globally throughout the nucleus, which would then increase the probability of MMEJ repairing a DSB that is formed later in heterochromatin (which may be cut more slowly). However, this explanation seems unlikely, because early breaks caused by ionizing radiation do not boost MMEJ repair at a Cas9 cut ~16 h later ([Brinkman et al., 2018](#)). This result suggests that the slow MMEJ activation does not occur globally throughout the nucleus but rather locally at the DSB. Further studies are needed to understand the different kinetics of MMEJ and NHEJ.

DSB repair pathways are known to vary in activity depending on the cell cycle stage ([Chapman et al., 2012](#); [Hustedt and Duracher, 2016](#); [Mladenov et al., 2016](#)). Our cell cultures were unsynchronized, and we assumed that all IPRs in the pools were subject to the same cell cycle stage distribution; thus, cell cycle effects should be averaged out, and the differences in indel spectra must be due to effects of chromatin context. In the future it will be interesting to explore whether any cross-talk exists between cell cycle and chromatin effects on pathway balance.

Comparison with previous studies

As summarized in the [Introduction](#), previous studies have addressed the impact of specific chromatin contexts or proteins on NHEJ and HR but generally did not monitor MMEJ and SSTR. Although the latter two pathways share components with HR, they are mechanistically distinct from HR, and thus the effects of chromatin context may differ. For example, several previous studies have indicated that H3K36me3, which is generally present along active transcription units, promotes HR ([Aymerich et al., 2014](#); [Carvalho et al., 2014](#); [Clouaire and Legube, 2015](#); [Pfister et al., 2014](#)). We find H3K36me3 to correlate negatively with MMEJ. Possibly, HR and MMEJ respond differently to H3K36me3, or the repair of Cas9-induced breaks differs from breaks induced by other means.

The activities of MMEJ and SSTR that we detect indicate that end-resection is generally not impeded by various types of heterochromatin. Previous work has pointed to a role of HP1 α and HP1 β in tethering proteins involved in end-resection ([Soria and Almouzni, 2013](#)), but not all types of heterochromatin are marked by these proteins. Our data indicate that multiple heterochromatin types can create an environment that is more prone to be repaired by MMEJ. This includes LADs, which is in agreement with a previous study that implicated various MMEJ-specific proteins in the repair of DSBs near the nuclear lamina ([Lemaître et al., 2014](#)).

Practical implications for genome editing

The results obtained in this study have practical implications for genome editing by means of Cas9. First, the efficiency of Cas9 editing is generally lower in most types of heterochromatin compared with euchromatin. This has been noted before but on the basis of data that covered only a small number of loci that did not compare all heterochromatin types ([Chen et al., 2016](#); [Daer et al., 2017](#); [Jensen et al., 2017](#); [Kallimasioti-Pazi et al., 2018](#)). Our data indicate that Cas9 editing is primarily suppressed in triple heterochromatin. Most likely the relatively low accessibility of the DNA in these loci is preventing efficient cutting by Cas9. Regions that carry only one of these marks, or H3K27me3, show only a modestly reduced editing efficiency.

From a genome editing perspective, the skew toward the MMEJ and SSTR pathways in heterochromatin is a convenient partial compensation for the lower overall editing efficiency, because MMEJ and SSTR are generally more useful than NHEJ to generate specific types of mutations. MMEJ is better suited to generate frameshifts and deletions that can result in functional KO of genes, while SSTR is particularly useful to generate specifically designed mutations. Maps of heterochromatin features are thus useful resources to choose the optimal target loci for CRISPR-Cas9 editing, particularly when combined with algorithms that predict editing outcomes on the basis of sequence ([Allen et al., 2018](#); [Chakrabarti et al., 2019](#); [Chen et al., 2019](#); [Shen et al., 2018](#); [van Overbeek et al., 2016](#)).

Multiplexed DSB repair reporters: outlook

This work complements a recent study that used a multiplexed reporter for DNA mismatch repair, which did not reveal significant effects of chromatin context on the repair outcome ([Pokusaeva et al., 2019](#)). Another multiplexed integrated reporter study also found evidence that genomic location can affect Cas9 editing efficiency ([Gisler et al., 2019](#)), but these results were more difficult to interpret because the reporter sequence itself was not transcriptionally inert. Importantly, neither of these studies addressed the impact of chromatin context on the balance between specific DSB repair pathways. Multiplexed reporter assays provide new opportunities to systematically investigate the role of local chromatin context in DSB repair by multiple pathways. Moreover, our time-series experiments demonstrate that the assay can be performed in 96-well format, making it scalable for applications such as drug screens and CRISPR screens. In the future, the assay may also be modified to include the detection of DSB resection or other intermediates, computational

modeling of time series to infer perfect repair (Brinkman et al., 2018), and perhaps measurements of HR activity.

Limitations of study

By using the repair scar as a readout, the method is limited to detecting mutagenic repair. Perfect repair (e.g., by HR or NHEJ) is therefore untraceable. HR might never be measurable with the current setup. However, perfect repair by NHEJ might be if the time-series approach quantifiable by mathematical modeling of time-series data (Figure 5; Brinkman et al., 2018) is improved. Related to this, our degraon control of Cas9 activity had a slow response. Given the repeated cutting and repair, it is very difficult to know what the actual rates are in our TRIP setting. Emerging Cas9 inhibitors (Kundert et al., 2019; Maji et al., 2019) and photo-cleavable (Carlson-Stevermer et al., 2020; Zou et al., 2021) and photoactivatable (Liu et al., 2020) guide RNAs (gRNAs) might provide help overcome this limitation by having a tighter control of Cas9 activity.

Although our data contain much more information in the various other indels, we limited our analysis to the main mutations produced by our gRNA for clarity. A more thorough analysis is feasible to better dissect the smaller nuances in DNA repair. We have also only looked at spacing for the indel size and ignored more complex mutations. The Tn5 approach (Figures S2J–S2M) can be optimized and used for better understanding of the balance between more complex mutations and their local chromatin environment. Finally, it is possible that the repair of Cas9-induced DSBs is different from the repair of DSBs generated by other agents. It will therefore be interesting to apply our approach to other sequence-specific endonucleases.

STAR★METHODS

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Material availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Cell culture
- **METHOD DETAILS**
 - Constructs
 - Generation of IPR cell pools
 - Cloning, expression and purification of Tn5
 - Tagmentation-based mapping of IPRs in the clones
 - Tagmentation-based indel and rearrangement detection
 - Transfection of sgRNA plasmids and ssODN
 - LMNA & LBR knock out generation
 - TIDE method
 - siRNA lipofection
 - RT-qPCR
 - Western blots
 - Cell viability
 - Inhibitor treatments

- pA-DamID
- Generation of indel sequencing libraries
- Time series
- Chromatin immunoprecipitation analysis of IPRs
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Mapping of IPR integration sites by inverse PCR
 - TagMeppR
 - Indel scoring
 - Tagmentation to identify rearrangements
 - Preprocessing of previously published chromatin data
 - Correction of ChIP signals at IPRs for differences in cutting frequencies
 - Time series analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2021.03.032>.

ACKNOWLEDGMENTS

We thank the NKI Genomics, Flow Cytometry, and Research High Performance Computing core facilities for excellent support and members from our laboratories for inspiring and helpful discussions. We thank Luca Braccioli and the NKI Protein core facility for Tn5 protein. The NKI Protein Facility is an Instruct-ERIC center. This work was supported by ZonMW TOP grant 91215067 (to R.H.M. and B.v.S.), European Research Council (ERC) Advanced Grant 694466 (to B.v.S.), and NIH Common Fund “4D Nucleome” Program grant U54DK107965 (B.v.S.). S.G.M. is funded by Marie Curie/AIRC iCARE2.0 fellowship 800924. The Oncode Institute is partly supported by KWF Dutch Cancer Society.

AUTHOR CONTRIBUTIONS

E.K.B., R.S., R.H.M., and B.v.S. conceived and designed the study. R.S., E.K.B., X.V., T.v.S., S.G.M., D.P.-H., B.M., and J.v.d.B. performed experiments. C.L., R.S., E.K.B., R.H.v.d.W., X.V., T.v.S., and S.G.M. performed data processing, bioinformatics, and figure preparation. B.v.S., R.H.M., and R.L.B. performed supervision and project management. R.S., E.K.B., and B.v.S. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

B.v.S. is a member of the Advisory Board of *Molecular Cell*.

Received: May 4, 2020

Revised: December 23, 2020

Accepted: March 19, 2021

Published: April 12, 2021

REFERENCES

- Akhtar, W., de Jong, J., Pindyurin, A.V., Pagie, L., Meuleman, W., de Ridder, J., Berns, A., Wessels, L.F., van Lohuizen, M., and van Steensel, B. (2013). Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* 154, 914–927.
- Akhtar, W., Pindyurin, A.V., de Jong, J., Pagie, L., Ten Hoeve, J., Berns, A., Wessels, L.F., van Steensel, B., and van Lohuizen, M. (2014). Using TRIP for genome-wide position effect analysis in cultured cells. *Nat. Protoc.* 9, 1255–1281.
- Alagoz, M., Katsuki, Y., Ogiwara, H., Ogi, T., Shibata, A., Kakarougkas, A., and Jeggo, P. (2015). SETDB1, HP1 and SUV39 promote repositioning of 53BP1 to extend resection during homologous recombination in G2 cells. *Nucleic Acids Res.* 43, 7931–7944.

- Aho, A.V., Kernighan, B.W., and Weinberger, P.J. (1992). The awk programming language. *Comput. Hum. 26*, 293–297.
- Allen, F., Crepaldi, L., Alsinet, C., Strong, A.J., Kleshchevnikov, V., De Angeli, P., Palenikova, P., Khodak, A., Kiselev, V., Kosicki, M., et al. (2018). Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* Published online November 27, 2018. <https://doi.org/10.1038/nbt.4317>.
- Aymard, F., Bugler, B., Schmidt, C.K., Guillou, E., Caron, P., Briois, S., Iacovoni, J.S., Daburon, V., Miller, K.M., Jackson, S.P., and Legube, G. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat. Struct. Mol. Biol.* *21*, 366–374.
- Baldehyron, C., Soria, G., Roche, D., Cook, A.J., and Almouzni, G. (2011). HP1 α recruitment to DNA damage by p150CAF-1 promotes homologous recombination repair. *J. Cell Biol.* *193*, 81–95.
- Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G., and Wandless, T.J. (2006). A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* *126*, 995–1004.
- Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* *42*, e168.
- 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.
- Brown, G. (2020). GreyListChIP: Grey Lists – Mask Artefact Regions Based on ChIP Inputs (R package version).
- Bushnell, B., Rood, J., and Singer, E. (2017). BBMerge—accurate paired shotgun read merging via overlap. *PLoS ONE* *12*, e0185056.
- Carlson-Stevermer, J., Kelso, R., Kadina, A., Joshi, S., Rossi, N., Walker, J., Stoner, R., and Maures, T. (2020). CRISPRoff enables spatio-temporal control of CRISPR editing. *Nat. Commun.* *11*, 5041.
- Carvalho, S., Vitor, A.C., Sridhara, S.C., Martins, F.B., Raposo, A.C., Desterro, J.M., Ferreira, J., and de Almeida, S.F. (2014). SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. *eLife* *3*, e02482.
- Chakrabarti, A.M., Henser-Brownhill, T., Monserrat, J., Poetsch, A.R., Luscombe, N.M., and Scaffidi, P. (2019). Target-specific precision of CRISPR-mediated genome editing. *Mol. Cell* *73*, 699–713.e6.
- Chan, S.H., Yu, A.M., and McVey, M. (2010). Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in *Drosophila*. *PLoS Genet.* *6*, e1001005.
- Chang, H.H.Y., Pannunzio, N.R., Adachi, N., and Lieber, M.R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* *18*, 495–506.
- Chapman, J.R., Taylor, M.R., and Boulton, S.J. (2012). Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell* *47*, 497–510.
- Chen, X., Rinsma, M., Janssen, J.M., Liu, J., Maggio, I., and Gonçalves, M.A. (2016). Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res.* *44*, 6482–6492.
- Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., van Steensel, B., Ma, J., and Belmont, A.S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* *217*, 4025–4048.
- Chen, W., McKenna, A., Schreiber, J., Haeussler, M., Yin, Y., Agarwal, V., Noble, W.S., and Shendure, J. (2019). Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Res.* *47*, 7989–8003.
- Chiolo, I., Minoda, A., Colmenares, S.U., Polyzos, A., Costes, S.V., and Karpen, G.H. (2011). Double-strand breaks in heterochromatin move outside of a dynamic HP1 α domain to complete recombinational repair. *Cell* *144*, 732–744.
- Clouaire, T., and Legube, G. (2015). DNA double strand break repair pathway choice: a chromatin based decision? *Nucleus* *6*, 107–113.
- Clouaire, T., and Legube, G. (2019). A snapshot on the *cis* chromatin response to DNA double-strand breaks. *Trends Genet.* *35*, 330–345.
- Clouaire, T., Rocher, V., Lashgari, A., Arnould, C., Aguirrebengoa, M., Biernacka, A., Skrzypczak, M., Aymard, F., Fongang, B., Dojer, N., et al. (2018). Comprehensive mapping of histone modifications at DNA double-strand breaks deciphers repair pathway chromatin signatures. *Mol. Cell* *72*, 250–262.e6.
- Clowney, E.J., LeGros, M.A., Mosley, C.P., Clowney, F.G., Markenscoff-Papadimitriou, E.C., Myllys, M., Barnea, G., Larabell, C.A., and Lomvardas, S. (2012). Nuclear aggregation of olfactory receptor genes governs their mono-genic expression. *Cell* *151*, 724–737.
- Corrales, M., Rosado, A., Cortini, R., van Arensbergen, J., van Steensel, B., and Filion, G.J. (2017). Clustering of *Drosophila* housekeeping promoters facilitates their expression. *Genome Res.* *27*, 1153–1161.
- 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.
- Daugaard, M., Baude, A., Fugger, K., Povlsen, L.K., Beck, H., Sørensen, C.S., Petersen, N.H., Sorensen, P.H., Lukas, C., Bartek, J., et al. (2012). LEDGF (p75) promotes DNA-end resection and homologous recombination. *Nat. Struct. Mol. Biol.* *19*, 803–810.
- Dekker, J., Belmont, A.S., Guttman, M., Leshyk, V.O., Lis, J.T., Lomvardas, S., Mirny, L.A., O’Shea, C.C., Park, P.J., Ren, B., et al.; 4D Nucleome Network (2017). The 4D nucleome project. *Nature* *549*, 219–226.
- DeWitt, M.A., Magis, W., Bray, N.L., Wang, T., Berman, J.R., Urbinati, F., Heo, S.J., Mitros, T., Muñoz, D.P., Boffelli, D., et al. (2016). Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci. Transl. Med.* *8*, 360ra134.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57–74.
- Frock, R.L., Hu, J., Meyers, R.M., Ho, Y.J., Kii, E., and Alt, F.W. (2015). Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* *33*, 179–186.
- Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat. Rev. Genet.* *21*, 292–310.
- Giannoukos, G., Ciulla, D.M., Marco, E., Abdulkarim, H.S., Barrera, L.A., Bothmer, A., Dhanapal, V., Gloskowski, S.W., Jayaram, H., Maeder, M.L., et al. (2018). UDI α ™, a genome editing detection method for indels and genome rearrangements. *BMC Genomics* *19*, 212.
- 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.
- Goodarzi, A.A., Noon, A.T., Deckbar, D., Ziv, Y., Shiloh, Y., Löbrich, M., and Jeggo, P.A. (2008). ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Mol. Cell* *31*, 167–177.
- Gottlieb, T.M., and Jackson, S.P. (1993). The DNA-dependent protein kinase: requirement for DNA ends and association with Ku antigen. *Cell* *72*, 131–142.
- Hendel, A., Kildebeck, E.J., Fine, E.J., Clark, J., Punjya, N., Sebastiano, V., Bao, G., and Porteus, M.H. (2014). Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell Rep.* *7*, 293–305.
- Hustedt, N., and Durocher, D. (2016). The control of DNA repair by the cell cycle. *Nat. Cell Biol.* *19*, 1–9.
- Iliakis, G., Murmann, T., and Soni, A. (2015). Alternative end-joining repair pathways are the ultimate backup for abrogated classical non-homologous end-joining and homologous recombination repair: Implications for the formation of chromosome translocations. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* *793*, 166–175.
- Jakob, B., Splinter, J., Conrad, S., Voss, K.O., Zink, D., Durante, M., Löbrich, M., and Taucher-Scholz, G. (2011). DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin. *Nucleic Acids Res.* *39*, 6489–6499.

- Janssen, A., Breuer, G.A., Brinkman, E.K., van der Meulen, A.I., Borden, S.V., van Steensel, B., Bindra, R.S., LaRocque, J.R., and Karpen, G.H. (2016). A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev.* **30**, 1645–1657.
- Jasin, M., and Haber, J.E. (2016). The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. *DNA Repair (Amst.)* **44**, 6–16.
- Jeggo, P.A., and Downs, J.A. (2014). Roles of chromatin remodellers in DNA double strand break repair. *Exp. Cell Res.* **329**, 69–77.
- Jensen, K.T., Fløe, L., Petersen, T.S., Huang, J., Xu, F., Bolund, L., Luo, Y., and Lin, L. (2017). Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett.* **591**, 1892–1901.
- Kallimasioti-Pazi, E.M., Thelakkad Chathoth, K., Taylor, G.C., Meynert, A., Ballinger, T., Kelder, M.J.E., Lalevée, S., Sanli, I., Feil, R., and Wood, A.J. (2018). Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair. *PLoS Biol.* **16**, e2005595.
- Kalousi, A., and Soutoglou, E. (2016). Nuclear compartmentalization of DNA repair. *Curr. Opin. Genet. Dev.* **37**, 148–157.
- Kundert, K., Lucas, J.E., Watters, K.E., Fellmann, C., Ng, A.H., Heineke, B.M., Fitzsimmons, C.M., Oakes, B.L., Qu, J., Prasad, N., et al. (2019). Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. *Nat. Commun.* **10**, 2127.
- Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* **47** (W1), W171–W174.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.
- Lee, Y.H., Kuo, C.Y., Stark, J.M., Shih, H.M., and Ann, D.K. (2013). HP1 promotes tumor suppressor BRCA1 functions during the DNA damage response. *Nucleic Acids Res.* **41**, 5784–5798.
- Leemans, C., van der Zwalm, M.C.H., Brueckner, L., Comoglio, F., van Schaik, T., Pagie, L., van Arensbergen, J., and van Steensel, B. (2019). Promoter-intrinsic and local chromatin features determine gene repression in LADs. *Cell* **177**, 852–864.e14.
- Lemaître, C., Grabarz, A., Tsouroula, K., Andronov, L., Furst, A., Pankotai, T., Heyer, V., Rogier, M., Attwood, K.M., Kessler, P., et al. (2014). Nuclear position dictates DNA repair pathway choice. *Genes Dev.* **28**, 2450–2463.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Lin, S., Staahl, B.T., Alla, R.K., and Doudna, J.A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766.
- Liu, Y., Zou, R.S., He, S., Nihongaki, Y., Li, X., Razavi, S., Wu, B., and Ha, T. (2020). Very fast CRISPR on demand. *Science* **368**, 1265–1269.
- Luna-Vargas, M.P., Christodoulou, E., Alfieri, A., van Dijk, W.J., Stadnik, M., Hibbert, R.G., Sahtoe, D.D., Clerici, M., Marco, V.D., Littler, D., et al. (2011). Enabling high-throughput ligation-independent cloning and protein expression for the family of ubiquitin specific proteases. *J. Struct. Biol.* **175**, 113–119.
- Maji, B., Gangopadhyay, S.A., Lee, M., Shi, M., Wu, P., Heler, R., Mok, B., Lim, D., Siriwardena, S.U., Paul, B., et al. (2019). A high-throughput platform to identify small-molecule inhibitors of CRISPR-Cas9. *Cell* **177**, 1067–1079.e19.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12.
- Mateos-Gomez, P.A., Gong, F., Nair, N., Miller, K.M., Lazzarini-Denchi, E., and Sfeir, A. (2015). Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* **518**, 254–257.
- McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* **24**, 529–538.
- Mitra, R., Fain-Thornton, J., and Craig, N.L. (2008). piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J.* **27**, 1097–1109.
- Mitrentsi, I., Yilmaz, D., and Soutoglou, E. (2020). How to maintain the genome in nuclear space. *Curr. Opin. Cell Biol.* **64**, 58–66.
- Mladenov, E., Magin, S., Soni, A., and Iliakis, G. (2016). DNA double-strand-break repair in higher eukaryotes and its role in genomic instability and cancer: cell cycle and proliferation-dependent regulation. *Semin. Cancer Biol.* **37–38**, 51–64.
- Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **42**, W401–W407.
- Okamoto, S., Amaishi, Y., Maki, I., Enoki, T., and Mineno, J. (2019). Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. *Sci. Rep.* **9**, 4811.
- Ott, C.J., Federation, A.J., Schwartz, L.S., Kasar, S., Klitgaard, J.L., Lenci, R., Li, Q., Lawlor, M., Fernandes, S.M., Souza, A., et al. (2018). Enhancer architecture and essential core regulatory circuitry of chronic lymphocytic leukemia. *Cancer Cell* **34**, 982–995.e7.
- Pfister, S.X., Ahrabi, S., Zalmas, L.P., Sarkar, S., Aymard, F., Bachrati, C.Z., Helleday, T., Legube, G., La Thangue, N.B., Porter, A.C., and Humphrey, T.C. (2014). SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. *Cell Rep.* **7**, 2006–2018.
- Pokusaeva, V.O., Diez, A.R., Espinar, L., and Filion, G.J. (2019). Strand asymmetry influences mismatch repair during single-strand annealing. *bioRxiv*. <https://doi.org/10.1101/847160>.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
- R Core Team (2020). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing). <https://www.R-project.org/>.
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44** (W1), W160–W165.
- Redwood, A.B., Perkins, S.M., Vanderwaal, R.P., Feng, Z., Biehl, K.J., Gonzalez-Suarez, I., Morgado-Palacin, L., Shi, W., Sage, J., Roti-Roti, J.L., et al. (2011). A dual role for A-type lamins in DNA double-strand break repair. *Cell Cycle* **10**, 2549–2560.
- Reginato, G., and Cejka, P. (2020). The MRE11 complex: a versatile toolkit for the repair of broken DNA. *DNA Repair (Amst.)* **91–92**, 102869.
- Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L., and Corn, J.E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344.
- Richardson, C.D., Kazane, K.R., Feng, S.J., Zelin, E., Bray, N.L., Schäfer, A.J., Floor, S.N., and Corn, J.E. (2018). CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* **50**, 1132–1139.
- Riesenberg, S., Chintalapati, M., Macac, D., Kanis, P., Maricic, T., and Pääbo, S. (2019). Simultaneous precise editing of multiple genes in human cells. *Nucleic Acids Res.* **47**, e116.
- RStudio Team (2020). RStudio: Integrated Development for R. RStudio (Boston, MA: PBC). <http://www.rstudio.com/>.
- Ryu, T., Spatola, B., Delabaere, L., Bowlin, K., Hopp, H., Kunitake, R., Karpen, G.H., and Chiolo, I. (2015). Heterochromatic breaks move to the nuclear periphery to continue recombinational repair. *Nat. Cell Biol.* **17**, 1401–1411.
- Salzberg, A.C., Harris-Becker, A., Popova, E.Y., Keasey, N., Loughran, T.P., Claxton, D.F., and Grigoryev, S.A. (2017). Genome-wide mapping of histone H3K9me2 in acute myeloid leukemia reveals large chromosomal domains associated with massive gene silencing and sites of genome instability. *PLoS ONE* **12**, e0173723.

- Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S.P. (2007). Human CtIP promotes DNA end resection. *Nature* **450**, 509–514.
- Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* **12**, 963–965.
- Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J., and Odom, D.T. (2009). ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240–248.
- Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science* **352**, 1225–1228.
- Scully, R., Panday, A., Elango, R., and Willis, N.A. (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* **20**, 698–714.
- Shah, R.N., Grzybowski, A.T., Cornett, E.M., Johnstone, A.L., Dickson, B.M., Boone, B.A., Cheek, M.A., Cowles, M.W., Maryanski, D., Meiners, M.J., et al. (2018). Examining the roles of H3K4 methylation states with systematically characterized antibodies. *Mol. Cell* **72**, 162–177.e7.
- Shen, M.W., Arbab, M., Hsu, J.Y., Worstell, D., Culbertson, S.J., Krabbe, O., Cassa, C.A., Liu, D.R., Gifford, D.K., and Sherwood, R.I. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651.
- Solovei, I., Wang, A.S., Thanisch, K., Schmidt, C.S., Krebs, S., Zwerger, M., Cohen, T.V., Devys, D., Foisner, R., Peichl, L., et al. (2013). LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell* **152**, 584–598.
- Soria, G., and Almouzni, G. (2013). Differential contribution of HP1 proteins to DNA end resection and homology-directed repair. *Cell Cycle* **12**, 422–429.
- Stern, D.L. (2017). Tagmentation-based mapping (TagMap) of mobile DNA genomic insertion sites. *bioRxiv*. <https://doi.org/10.1101/037762>.
- Sun, Y., Jiang, X., Xu, Y., Ayrapetov, M.K., Moreau, L.A., Whetstone, J.R., and Price, B.D. (2009). Histone H3 methylation links DNA damage detection to activation of the tumour suppressor Tip60. *Nat. Cell Biol.* **11**, 1376–1382.
- Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034.
- Tsouroula, K., Furst, A., Rogier, M., Heyer, V., Maglott-Roth, A., Ferrand, A., Reina-San-Martin, B., and Soutoglou, E. (2016). Temporal and spatial uncoupling of DNA double strand break repair pathways within mammalian heterochromatin. *Mol. Cell* **63**, 293–305.
- van Overbeek, M., Capurso, D., Carter, M.M., Thompson, M.S., Frias, E., Russ, C., Reece-Hoyes, J.S., Nye, C., Gradia, S., Vidal, B., et al. (2016). DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol. Cell* **63**, 633–646.
- van Schaik, T., Vos, M., Peric-Hupkes, D., Hn Celie, P., and van Steensel, B. (2020). Cell cycle dynamics of lamina-associated DNA. *EMBO Rep.* **21**, e50636.
- Vogel, M.J., Peric-Hupkes, D., and van Steensel, B. (2007). Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat. Protoc.* **2**, 1467–1478.
- Wang, X., Spandidos, A., Wang, H., and Seed, B. (2012). PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.* **40**, D1144–D1149.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag).
- Yeh, C.D., Richardson, C.D., and Corn, J.E. (2019). Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* **21**, 1468–1478.
- Zorita, E., Cuscó, P., and Filion, G.J. (2015). Starcode: sequence clustering based on all-pairs search. *Bioinformatics* **31**, 1913–1919.
- Zou, R.S., Liu, Y., Wu, B., and Ha, T. (2021). Cas9 deactivation with photo-cleavable guide RNAs. *Mol. Cell*. Published online February 23, 2021. <https://doi.org/10.1016/j.molcel.2021.02.007>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-H3K4me1 (ChIP)	Abcam	Cat# ab8895; RRID: AB_306847
Anti-H3K27me3 (ChIP)	Active Motif	Cat# 39155; RRID: AB_2561020
Anti-H3K27ac (ChIP)	Active Motif	Cat# 39133; RRID: AB_2561016
Anti-POLQ (ChIP)	Sigma-Aldrich	Cat# SAB1402530.; RRID: AB_10639636
Anti-LIG4 (ChIP)	GeneTex	Cat# GTX55592; RRID: AB_2887931
Anti-RAD51 (ChIP)	Santa Cruz Biotechnology	Cat# sc-8349; RRID: AB_2253533
Anti-MRE11 (ChIP)	Novus	Cat# NB100-142; RRID: AB_10077796
Anti-LMNB2 (pA-DamID)	Abcam	Cat# ab8983; RRID: AB_306912
Anti-Mouse (pA-DamID)	Abcam	Cat# ab6709; RRID: AB_956006
Anti-H3K27me3 (Western Blot)	Cell Signaling	Cat# C36B11; RRID: AB_11220433
Anti-H3K9me2 (Western Blot)	Millipore	Cat# 07-441; RRID: AB_11212297
Anti-LBR (Western Blot)	Abcam	Cat# ab122919; RRID: AB_10902156
Anti-LMNA (Western Blot)	Santa Cruz Biotechnology	Cat# sc-376248; RRID: AB_10991536
Bacterial and virus strains		
CloneCatcher DH5 α electrocompetent <i>E. coli</i>	Genlantis	Cat# C810111
JM109 Competent Cells	Promega	Cat# L2001
Chemicals, peptides, and recombinant proteins		
RPMI 1640	GIBCO	Cat#: 21875034
Fetal Bovine Serum	Sigma	Cat#: F7524
Penicillin-Streptomycin (##10,000 U/mL)	GIBCO	Cat#: 15070063
Lipofectamine 2000	Invitrogen	Cat#: 11668019
Tn5 enzyme	Luca Braccioli & this study	N/A
PEG 8000x	Sigma	Cat#: P1458
TAPS-NAOH	Sigma	Cat#: T5130
dimethylformamide	Sigma	Cat#: D4551
Shield-1	Aeobius	Cat#: AOB1848
Phusion® Hot Start Flex DNA Polymerase	New England BioLabs	Cat#: M0535L
Phusion® HF DNA Polymerase	New England BioLabs	Cat#: M0530L
MyTac Red Mix 2x	Bioline	Cat#: BIO-25044
CleanPCR	CleanNA	Cat#: CPCR-0500
Shrimp Alkaline Phosphatase (1U/ μ l)	New England Biolabs	Cat#: M0371S
Exonuclease I	New England Biolabs	Cat#: M0293S
RNase-Free DNase Set	QIAGEN	Cat#: 79254
SensiFast no-ROX	Bioline	Cat#: BIO-86050
ATP Solution (100 mM)	Thermo Scientific	Cat#: R0441
DMSO	Sigma	Cat#: D4540
NU7441	Cayman	Cat#: 14881
M3814	MCE	Cat#: HY-101570
GSK126	Cayman	Cat#: 15415
BIX01294	Sigma	Cat#: B9311
Lipofectamine RNAiMAX Transfection Reagent	Invitrogen	Cat#: 13778150
Opti-MEM	GIBCO	Cat#: 31985047

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DirectPCR® Lysis	Viagen	Cat#: 302-C
Proteinase K	Bioline	Cat#: BIO-37084
MyTaq HS Red mix	Bioline	Cat#: BIO-25048
CellTiter-Blue® Cell Viability Assay	Promega	Cat#: G8080
spermidine	Sigma	Cat#: S0266
digitonin	Millipore	Cat#: 300410
cOMplete Protease Inhibitor Cocktail	Roche	Cat#: 11873580001
SAM	New England BioLabs	Cat#: B9003S
Dam	New England BioLabs	Cat#: #M0222L
Critical commercial assays		
ISOLATE II Genomic DNA kit	Bioline	Cat#: BIO-52067
PCR Isolate II PCR and Gel Kit	Bioline	Cat#: BIO-52060
Tetro Reverse Transcriptase	Bioline	Cat#: BIO-65050
PureLink HiPure Plasmid Midiprep Kit	Invitrogen	Cat#: K210004
Qubit dsDNA HS Assay Kit	Invitrogen	Cat#: Q32854
RNeasy Mini Kit	QIAGEN	Cat#: 74104
Deposited data		
Raw data	This study	SRA: PRJNA686952
Processed data	This study	https://osf.io/cywxd/
Code	This study	GitHub: https://github.com/vansteensellab/DSB_repair_TRIP
Unprocessed images	This study	https://osf.io/cywxd/
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/grc/human
Experimental models: cell lines		
K562#17 ddCas9	Brinkman et al., 2018	N/A
Oligonucleotides		
Tables S3 and S4	This study	N/A
Rad51 siRNA ON-TARGETplus Smart Pool	Dharmacon/Horizon	Cat#: L-003530-00-0005
PoIQ siRNA ON-TARGETplus Smart Pool	Dharmacon/Horizon	Cat#: L-015180-01-0005
BRCA1 siRNA ON-TARGETplus Smart Pool	Dharmacon/Horizon	Cat#: L-003461-00-0005
BRCA2 siRNA ON-TARGETplus Smart Pool	Dharmacon/Horizon	Cat#: L-003462-00-0005
Lig4 siRNA ON-TARGETplus Smart Pool	Dharmacon/Horizon	Cat#: L-004254-00-0005
RBBP8 siRNA ON-TARGETplus Smart Pool	Dharmacon/Horizon	Cat#: L-011376-00-0005
Oligo(dT)20	Invitrogen	Cat#: 18418020
Recombinant DNA		
pPTK-Gal4-tet-Off-Puro-IRES-eGFP-sNRP-pA-trim1	Akhtar et al., 2014	GenBank: KC710229
pPTK-P.CMV.584-eGFP-trim1-PI04	Alexey Pindyurin and Waseem Akhtar	N/A
pBlue-sgRNA	Brinkman et al., 2018	N/A
pPTK-BC-IPR	This study	GenBank: MW408732
mPB-L3-ERT2-mCherry	Akhtar et al., 2014	N/A
Software and algorithms		
Bowtie2 v2.3.4	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools v1.5	Li et al., 2009	RRID:SCR_002105; https://www.htslib.org/
Cutadapt v1.9.1	Martin, 2011	RRID:SCR_011841; https://cutadapt.readthedocs.io/en/stable/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Starcode v1.1	Zorita et al., 2015	https://github.com/gui11aume/starcode
TagMeppr	This paper	https://github.com/robinweide/tagmeppr
Sambamba v0.6.6	Tarasov et al., 2015	https://lomereiter.github.io/sambamba/
deeptools v3.3.1	Ramírez et al., 2016	RRID:SCR_016366; https://deeptools.readthedocs.io/en/develop/
HMMt	Guillaume Filion	https://github.com/gui11aume/HMMt
BBTools v38.86	Bushnell et al., 2017	RRID:SCR_016968; https://sourceforge.net/projects/bbmap/
FASTX-Toolkit v0.0.14	Hannon Lab	RRID:SCR_005534; http://hannonlab.cshl.edu/fastx_toolkit/
Bedtools v2.26.0	Quinlan and Hall, 2010	RRID:SCR_006646; https://github.com/arq5x/bedtools2
BWA MEM	Heng Li arXiv:1303.3997v2 [q-bio.GN]	https://github.com/lh3/bwa
GreyListChIP	Brown, 2020	https://doi.org/doi:10.18129/B9.bioc.GreyListChIP
chipseq-greylis	Rory Kirchner	https://github.com/roryk/chipseq-greylis
Custom code for this study	This paper	https://github.com/vansteensellab/DSB_repair_TRIP
R code for this study	This paper	https://github.com/vansteensellab/DSB_repair_TRIP
RStudio Server Version 1.3.1073	RStudio Team, 2020	https://rstudio.com/
R version 3.6.3 (2020-02-29)	R Core Team, 2020	https://www.r-project.org/
ggplot2	Wickham, 2016	https://ggplot2.tidyverse.org
CHOPCHOP	Labun et al., 2019	https://chopchop.cbu.uib.no/
inDelphi	Shen et al., 2018	https://indelphi.giffordlab.mit.edu/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Bas van Steensel (b.v.steensel@nki.nl).

Material availability

Plasmids and cell lines generated in this study are available without restriction upon request from the Lead Contact.

Data and code availability

Processed data, script outputs, and original images are available at <https://osf.io/cywxd/>. Raw sequencing data are available at the Sequence Read Archive: PRJNA686952. Code: https://github.com/vansteensellab/DSB_repair_TRIP; <https://github.com/robinweide/tagmeppr>

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

We used clonal cell line K562#17, which is a female human K562 cell line (ATCC) stably expressing DD-Cas9 (Brinkman et al., 2018). K562#17 cells were cultured in RPMI 1640 (GIBCO) supplemented with 10% fetal bovine serum (FBS, Sigma), 1% penicillin/streptomycin. Cells were free of mycoplasma according to tests performed every 1-2 months.

METHOD DETAILS

Constructs

The pPTK-BC-IPR (PiggyBac) construct was derived from the pPTK-Gal4-tet-Off-Puro-IRES-eGFP-sNRP-pA-trim1 plasmid (GenBank: KC710229). The enhanced green fluorescent protein (eGFP) expression transcription unit including promoter, the puromycin

resistance cassette (PuroR) and the internal ribosome entry site (IRES) were replaced with the sgRNA-LBR2 target sequence and its flanking region. This plasmid did contain a point mutation which was removed by restriction cloning using a derivative of plasmid with a shortened 3' ITR of 67 bp (pPTK-P.CMV.584-eGFP-trim1-PI04 – kindly provided by Alexey Pindyurin and Waseem Akhtar). The target sequence was obtained by annealing ODS001 and ODS002 (400 pM each) (for primer sequences see [Table S4](#)) in 50 μ l MyTaq Red mix followed by 5 cycles of PCR. This PCR product was then further amplified with TAC0001 and TAC0002 (50 pM each). This sequence was then inserted in the PB backbone by restriction cloning with NheI and KpnI. This construct (IPR-PB) was then used to make the barcoded plasmid libraries, the 3' ITR of PiggyBac was amplified with primers TAC0003 (containing a 16 nucleotide random barcode) and TAC0004. The PCR product was digested with KpnI and BssHII, ligated into the KpnI and MluI sites of the IPR-PB plasmid and transformed into CloneCatcher DH5 α electrocompetent *E. coli* (Genlantis). A pool of ~500,000 transformed bacterial cells were grown, and plasmids were purified, resulting in the pPTK-BC-IPR (GenBank: MW408732) plasmid library. The PB transposase expression vector (mPB-L3-ERT2-mCherry) is described in ([Akhtar et al., 2014](#)). The sgRNAs were designed using CHOP-CHOP ([Labun et al., 2019](#); [Montague et al., 2014](#)) and cloned into expression vector pBlue-sgRNA ([Brinkman et al., 2018](#)) (see [Table S3](#)). All the plasmids were extracted and purified using the PureLink HiPure Plasmid Midiprep Kit (Invitrogen). The vectors for the sgRNA sequences are listed in [Table S3](#).

Generation of IPR cell pools

Cell pools carrying IPRs were produced as described ([Akhtar et al., 2014](#)). Briefly, K562#17 cells were transfected with 32 μ g of barcoded pPTK-BC-IPR plasmid library and 6 μ g of PB transposase plasmid using Lipofectamine 2000 (ThermoFisher). Mock-transfected (without PB transposase) and GFP plasmid controls were included. After 24 h, the cells were sorted by fluorescence-activated cell sorting (FACS) based on mCherry signals. We discarded cells without any detectable mCherry signal because they most likely failed to take up any plasmid. 0.5 μ M of 4-hydroxytamoxifen (4-OHT) was added to the samples to activate the transposase. Sixteen hours later the cells were washed to remove 4-OHT. After sorting, the population was grown for 8 days to clear the cells from free plasmid. Then, the mCherry negative cells were FACS sorted in aliquots of ~2000 cells, which were expanded to establish two cell pools, each with a different collection of IPRs. We also isolated single cells to make clonal TRIP lines originating from pool B, including clone 5.

Cloning, expression and purification of Tn5

The gene encoding Tn5 was cloned into the pETNKL-his-SUMO3-LIC vector, containing a N-terminal 6xHis-SUMO3-tag, using Ligation Independent Cloning (LIC) ([Luna-Vargas et al., 2011](#)). The recombinant Tn5 protein was expressed in Rosetta2(DE3) cells in 3 l of LB medium, supplemented with 30 μ g/ml kanamycin and 40 μ g/ml chloramphenicol. Cells were grown at 37°C to OD600 = 0.6–0.8. Cells were cooled to 18°C before 0.4 mM IPTG was added and protein was expressed overnight. Cells were harvested by centrifugation (3,000 g, 15 min, 20°C) and pellet was stored at –20°C until further use.

For protein purification, the pellet was resuspended in lysis buffer (20 mM HEPES pH 7.5, 800 mM NaCl, 1 mM EDTA, 5% glycerol, 1 mM TCEP) containing 0.2% Triton X-100. Cells were lysed by sonication and the lysate was clarified by centrifugation (50,000 g, 30 min, 4°C). Polyethylenimine (0.1% w/v) was added dropwise to the supernatant, incubated for 30 min and the precipitate was removed by centrifugation (50,000 g, 30 min, 4°C). The soluble fraction was used for affinity purification (1 mL nickel Sepharose Excel, GE healthcare). Beads were washed with lysis buffer, containing 20 mM imidazole, and protein was eluted by 200 mM imidazole in lysis buffer. Fractions were analyzed by SDS-PAGE. To remove the 6xHis-SUMO3-tag, pooled elution fractions were incubated with his-Senp2 protease, followed by overnight dialysis at 4°C against 20 mM HEPES pH 7.5, 800 mM NaCl, 1 mM EDTA, 5% glycerol, 1 mM TCEP. After reverse affinity purification using 1 mL nickel Sepharose excel, the 6xHis-SUMO3-tag was found to be efficiently cleaved off from Tn5. The flow-through and wash fractions were pooled and concentrated to 1 ml. The protein was further purified by size exclusion chromatography on a SEC 650 10/300 column (Bio-Rad), equilibrated with 20 mM HEPES pH 7.5, 800 mM NaCl, 1 mM EDTA, 5% glycerol, 1 mM TCEP. Fractions containing dimeric Tn5 were pooled, concentrated and flash frozen in liquid nitrogen before storing at –80°C.

Tagmentation-based mapping of IPRs in the clones

Not all the IPRs were mapped in the clones by iPCR. They were additionally mapped using a Tn5 transposon based IPR mapping technique based on [Stern \(2017\)](#). In brief, PCR-products covering both the up- and downstream junctions between the genome and the inverted terminal inverted repeats (ITRs) were obtained by designing divergent internal primers. The library preparation is carried out as follows. 45 μ l of 100 μ M of TAC0101 & TAC0102 each were mixed with 10 μ l 10x TE and annealed using the following PCR reaction; 10 min at 95°C, 1 min at 90°C, followed by a slow ramp down (0.1°C/sec) until 4°C. The transposome is obtained by combining the adapters (1 μ l of 1:2 diluted adapters) and the Tn5 transposon (1.5 μ l of 2.7 mg/mL stock) in 18.7 μ l Tn5 dilution buffer (20 mM HEPES, 500 mM NaCl, 25% Glycerol) and incubating the mix for 1 hour at 37°C. The DNA tagmentation was performed by mixing 100 ng of gDNA with 1 μ l of transposome, 4 μ l 5x TAPS-PEG buffer (50mM TAPS-NAOH, 25mM MgCl₂, 8% vol/vol PEG8000) in a final volume of 20 μ l. This was incubated at 55°C for 10 minutes and quenched afterward with 4 μ l of 0.2% SDS. Library preparation was as follows. Both sides of the PiggyBac transposon were processed for the best mapping results by generating 3' ITR and 5' ITR libraries. First, we enriched our target region with a linear enrichment PCR amplification using TAC0006 (3' ITR) and TAC0099 (5' ITR). The PCR mix was 3 μ l of tagmented DNA, 1 μ l of 1 μ M primer, 2 μ l dNTPs (10mM), 4 μ l 5x Phusion® HF Buffer (Promega), 0.25

Phusion® HS Flex polymerase (2 U/μl - Promega), in a final volume of 20 μl and amplified as follows; 30 s at 98°C, 45 cycles of 10 s at 98°C, 20 s at 62°C and 30 s at 72°C. PCR1 of the library preparation was done with TAC0161 (3' ITR) and TAC0110 (5' ITR) in combination with N5xx (Nextera Index Kit – Illumina). The PCR mix was 5 μl of enrichment PCR, 1 μl of 10 μM primers, 2 μl dNTPs (10mM), 4 μl 5x Phusion® HF Buffer and 0.25 Phusion® HS Flex polymerase, in a final volume of 25 μl and amplified as follows; 30 s at 98°C, 3 cycles of 10 s at 98°C, 20 s at 62°C and 30 s at 72°C, 8 cycles of 10 s at 98°C, 50 s at 72°C. PCR2 of the library preparation was done with TAC0103 (both ITRs) and N7xx (Nextera Index Kit – Illumina). The PCR mix was 2 μl of PCR1, 1 μl of 10 μM primers, 2 μl dNTPs (10mM), 4 μl 5x Phusion® HF Buffer and 0.25 Phusion® polymerase (Promega), in a final volume of 22 μl and amplified as follows; 30 s at 98°C, 10 cycles of 10 s at 98°C, 20 s at 63°C and 30 s at 72°C. The library was then checked and purified with bead purification as for the indel libraries, quantitated with Qubit and sequenced on a Miseq (150 bp, paired-end).

Tagmentation-based indel and rearrangement detection

To assess potential rearrangements and very large deletions in our setup, we have adapted the protocol to map IPRs (above) to read out the IPR barcode and mutations. As input, we used three control replicates (clone 5) from the LBR & LMNA knock out experiments (Figure 6B) that were transfected either with the LBR2 sgRNA or an empty vector. The protocol follows the same steps as the IPR mapping except that we used an improved tagmentation buffer and the PCR primers are different. Instead of 5x TAPS-PEG buffer we used 10 ul of 2x TD buffer (20 mM Tris, 10mM MgCl₂, adjusted to pH 7.6 with 100% acetic acid before the addition of 20% (vol/vol) dimethylformamide). For the library PCR we first enriched the IPR with TAC0078. PCR1 was carried out with TAC0238 (primer F on Figure S2J) & Nextera N5xx. PCR2 was unchanged.

Transfection of sgRNA plasmids and ssODN

For transient transfection of the sgRNAs, 1 to 6 x 10⁶ cells (lower limit for clonal experiments, higher limit for pooled experiments) were resuspended in transfection buffer (100 mM KH₂PO₄, 15 mM NaHCO₃, 12 mM MgCl₂, 8 mM ATP, 2 mM glucose (pH 7.4)) (Hendel et al., 2014). After addition of 3.0-9.0 μg plasmid, the cells were electroporated in an Amaxa 2D Nucleofector using program T-016. DD-Cas9 was induced directly for ~16 hours after transfection with a final concentration of 500 nM Shield-1 (Aobious). For uncut controls we transfected either a GFP containing plasmid or pBlue-sgRNA vector without a sgRNA sequence. To probe SSTR, 3-9 μg sgRNA was co-transfected with 1.5-4.5 μg ssODN (5' TAGAATGCTAGCGTGACTGGAGTTCAGACGTTGTGCTCTTCCGATC-TAATTTCTACTTCATAATAAAGTGAAGTCCCAGGCCATCGACATCTCTTACCACTTCACCATCGGCAAATTTCTACTTGGCATT 3', Ultramer grade, IDT). The specific mutation that disrupts the PAM is underlined.

LMNA & LBR knock out generation

The guides for LBR and LMNA knock outs were selected using a combination of inDelphi (Shen et al., 2018) and CHOPCHOP (Labun et al., 2019; Montague et al., 2014) of optimal frameshift chances and efficiency. One million clone 5 cells were transfected with 3 μg of plasmid expressing the following sgRNAs per clone: LMNA_KO1 & LMNA_KO2 (each 1.5 μg) for LMNA KO 1 and 2; LMNA_KO4 for LMNA KO 3 and 4; LBR_KO1 for LBR KO 1-4 (Table S3), and cultured in complete RPMI medium with 500 nM Shield-1 (to activate Cas9) for 3 days. To obtain individual clones, cells were plated in two 96-well plates by limiting dilution (2 cells per ml; 100 μl per well). Each clone was then tested by TIDE (Brinkman et al., 2014) for frameshifts in all alleles (primers in Table S4). For each sgRNA we selected two clones with complete frameshifts for further experiments.

TIDE method

Prior to high throughput sequencing, all the samples were checked for general transfection and cutting efficiency by TIDE. For this we used the primers TAC0017 and TAC0018, that cover the endogenous LBR locus and the TIDE method was performed as described in Brinkman et al. (2014). Briefly, PCR reactions were carried out with ~100 ng genomic DNA in MyTaq Red mix (Bioline) and purified using the PCR Isolate II PCR and Gel Kit (Bioline) or by ExoSAP (for primers see Table S4). ExoSAP was done by adding 0.125 μl Shrimp Alkaline Phosphatase (1U/μl; New England Biolabs), 0.0125 μl Exonuclease I (20 U/μl; New England Biolabs) and 2.3625 μl H₂O per 10 μl PCR reaction. Samples were incubated 30 minutes at 37°C and inactivated for 10 minutes at 95°C. About 2 μl (50-100 ng) of purified PCR product was then subjected to Sanger Sequencing by Eurofins Genomics. The sequence traces were analyzed using the TIDE analysis tool (<https://tide.nki.nl>).

siRNA lipofection

All siRNAs were obtained from Dharmacon as ON-TARGETplus Smartpool and transfected on 1 million cells with the RNAiMAX Transfection kit (Invitrogen) at a final concentration of 25 nM, 24h prior to sgRNA electroporation according to the manufacturer's protocol for 6 well plates. Samples were partially collected 24h after electroporation for subsequent RNA isolation, reverse transcription and qPCR analysis. The rest was left to grow for 48 more hours CRISPR/Cas9 editing analysis.

RT-qPCR

Cells were collected in RLT buffer with 1% 2-mercaptoethanol and stored at –80C until processed. Total RNA was extracted using RNeasy mini kit (QIAGEN) with on-column DNase treatment (QIAGEN) according to the manufacturer's instructions. The RNA was eluted in 30 ul RNase-free H₂O and quantitated by Nanodrop. RNA was reverse-transcribed using Tetro Reverse Transcriptase

(Bioline) with random hexamers and Oligo(dT)20 primers (Invitrogen) according to the manufacturer's instructions. The RT reaction was diluted 10x and 4 μ L RT was used for the qPCR reaction. qPCR was performed using SensiFast no-ROX mix (Bioline) in a 10 μ L reaction. Melt curves after each PCR and all samples yielded a single peak. Gene-specific primers were obtained from PrimerBank (Wang et al., 2012), see Table S4. Data were normalized to the levels of TBP.

Western blots

Whole-cell extracts of $\sim 0.5 \times 10^6$ cells were prepared by washing cultures in PBS and lysing with 50 μ L lysis buffer (Tris pH 7.6, 10% SDS, Roche cComplete Protease Inhibitor Cocktail). Western blotting was performed according to standard procedures using the following antibodies and dilutions: H3K27me3 (1:1000 Cell Signaling C36B11, rabbit), H3K9me2 (1:1000 Millipore 07-441, rabbit), LMNA (1:800 Santa Cruz Biotechnology sc-376248, mouse), LBR (1:1000, Abcam, ab122919, rabbit).

Cell viability

Cells transfected with Rad51 siRNA were tested for cell viability using the Cell Titer Blue 48 hours after lipofection. After mixing, 100 μ L cells were plated in a 96-well opaque-walled tissue culture plates plus 20 μ L/well of CellTiter-Blue[®] Reagent (Promega), in 3 technical replicates per sample. The cells were briefly shaken and incubated for 3 hours in at cell culturing conditions. The fluorescence was measured by a Perkin Elmer EnVision plate reader. The three technical replicates were averaged, then the biological replicates were averaged, and the result was normalized of the non-targeting siRNA control. One sample t test was used for the statistics.

Inhibitor treatments

DNA-PKcs inhibitor NU7441 (Cayman; diluted 1:1000 from 1 mM stock in dimethylsulfoxide [DMSO]), M3814 (MedChemExpress; diluted 1:1000 from 1 mM stock in DMSO), GSK126 (Selleckchem; diluted 1:2000 from 1 mM stock in DMSO), BIX01294 (Sigma; diluted 1:1000 from 1 mM stock in H₂O), or respective solvent-only controls at equal volumes, was added to the cells at the same time when the cells were supplemented with Shield-1 to induce DD-Cas9 or 24 hours prior to nucleofection for GSK126 and BIX01294. DMSO was also present in the experiments in Figures 3, 4, 5, 6A, 7, S2A–S2F, S3A, S3B, S4, S5, and S7.

pA-DamID

pA-DamID maps were generated and processed as described (van Schaik et al., 2020). Briefly, 1 million cells were collected by centrifugation (3 minutes, 500 g) and washed in ice-cold PBS and subsequently in ice-cold digitonin wash buffer (DigWash) (20 mM HEPES-KOH pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.02% digitonin, cComplete Protease Inhibitor Cocktail). Cells were resuspended in 200 μ L DigWash with 1:100 mouse Lamin B2 antibody (Abcam, ab8983) and rotated for 2 hours at 4°C, followed by a wash step with 0.5 mL DigWash buffer. This was repeated with a 1:100 mouse anti-rabbit antibody (Abcam, ab6709) and 1 hour of rotation, and afterward with 1:100 pA-Dam (~ 60 NEB units). After two washes with DigWash, cells were resuspended in 100 μ L DigWash supplemented with 80 μ M SAM to activate Dam and incubated for 30 minutes at 37°C. Genomic DNA was extracted using the ISOLATE II Genomic DNA kit (Bioline cat. no. BIO-52067) and DNA was processed for high-throughput sequencing similar to conventional DamID (Leemans et al., 2019; Vogel et al., 2007), except that the DpnII digestion was omitted. To control for DNA accessibility and amplification bias, 1 million permeabilized cells (without any antibodies bound) were incubated with 4 units of Dam enzyme (NEB, M0222L) during the activation step. This sample functions as “dam-control” over which a log₂-ratio is determined. Log₂-ratios were converted to z-scores to account for small differences in dynamic range between experiments.

Generation of indel sequencing libraries

After 64 or 72 (for clone 5) hours incubation, the cells were collected, and genomic DNA was extracted using the ISOLATE II Genomic DNA kit (Bioline cat. no. BIO-52067). PCR was performed in two steps and pooled experiments were performed in triplicates for a higher coverage. IndelPCR1 was performed with 200 ng genomic DNA each using primers TAC0007 (indexed) and TAC0012 that amplify 1 bp upstream of the barcode 46 bp downstream of the cut-site (see Figure 1A; Table S4). indelPCR2 used 2 μ L of each indelPCR1 product with TAC0009 and either TAC0011 (non-indexed) or TAC0159 (indexed). Each sample was generated with a unique combination of one or two indexes. Both PCR reactions were carried out with 25 μ L MyTaq Red mix (Bioline cat. no. BIO-25044), 0.5 μ M of each primer and 50 μ L final volume. PCR conditions for both steps were 1 min at 95°C, followed by 15 s at 95°C, 15 s at 58°C and 1 min at 72°C (5x), followed by 15 s at 95°C, 15 s at 65°C and 1 min at 72°C (10x). The indelPCR2 was pooled per experiment after quantification on a 1% agarose gel and cleaned up using CleanPCR (CleanNA) beads at 0.8:1 beads:sample ratio. 5 μ L PCR product was run on a 1% agarose gel to check for remaining primer dimers and if required reloaded on a 1% agarose gel and cut out to remove remaining primer dimers and cleaned with PCR Isolate II PCR and Gel Kit (Bioline). The libraries were quantitated using the Qubit DNA dsHS Assay Kit. The purified libraries were sequenced on an Illumina HiSeq2500 or MiSeq (150bp, single-end) depending on the expected complexity of the library.

Time series

Sample collection in the time series experiments was done automatically using a Hamilton Microlab[®] STAR equipped with a Cytomat 2 C450 incubator. One million cells were transfected as described above with the sgLBR2. After 16 hours, 40,000 cells in 100 μ L medium were seeded per well in 96-well plates. The automated system then added 100 μ L RPMI medium (with 1 μ M Shield-1) to each

well, for a final 500 nM Shield-1 concentration. The first time point was directly collected and required a brief centrifugation step (10 s at 500 g) to precipitate the cells, before returning the cell culture plate to the robot. Then for each time point 170 μ L medium was removed from the well and discarded, the left-over was mixed and transferred to a new 96-well PCR plate at 8°C. Each newly collected well was then filled with 50 μ L of DirectPCR® Lysis (Viagen) buffer with 1 mg/ml Proteinase K (Bioline) to pre-lyse the cells. The cell culture plate was returned to the incubator and every 3 hours a new time point was collected as described above. One 96-well plate included 4 timeseries of each 24 time points. After 69 hours the collection of samples was finished, and the cell lysates were sealed and incubated for 3 hours at 55°C and heat-inactivated for 10 minutes at 95°C.

Library preparation for the timeseries was very similar to the pool experiments except for indelPCR1. 20 μ L of crude lysate was used in a total PCR volume of 80 μ L, with 40 μ L MyTaq HS Red mix (Bioline) and 0.5 μ M of each primer. PCR cycles were as described above.

Chromatin immunoprecipitation analysis of IPRs

Chromatin immunoprecipitation was performed as described (Schmidt et al., 2009). Main steps and modifications are described here. 50 μ L protein A Dynabeads were precleared with 0.5% BSA, 5 μ L of specific antibody (H3K4me1 [Abcam]; H3K27me3 [Active Motif]; H3K27ac [Active Motif], MRE11 [Novus]; DNA ligase IV [Genetex]; Rad51 [Santa Cruz Biotechnology]; DNA polymerase theta [Sigma]) and beads incubated at 4°C overnight. 10 million clone 5 cells (uncalibrated experiments) and 3 million cells (calibrated experiments) were fixed at a final concentration of 1% formaldehyde for 10 minutes. Fixation was quenched with 125nM glycine for 5min and a PBS wash. Equal amounts of clone 9 cells were fixed and quenched for calibrated experiments. Spike-in cells (clone 9) were mixed in a 1:1 ratio with clone 5 cells after fixation. After nuclear extraction, chromatin was sonicated (~10 cycles 30 s on/ 30 s off in BioRuptor Pico), Triton X-100 added to a final concentration of 1% and centrifuged to remove cell debris. Antibody coupled beads were washed with 0.5% BSA in PBS, chromatin was added (5% was kept as input) and rotated overnight at 4°C. Beads were washed 10 times with RIPA buffer and once with TBS. After last wash, 200 μ L of elution buffer was added and samples eluted and de-crosslinked at 65°C for 6 hours or overnight. 200 μ L of TE buffer and 0.9 μ L of 10mg/ml RNase A was added the samples and were incubated at 37°C for 1 hour and with 4 μ L of 20mg/ml Proteinase K at 55°C for 2 hours. DNA was extracted by phenol:chloroform extraction and resuspended in 50 μ L of 10mM Tris-HCl. IPR barcodes were collectively amplified using two step PCR. For indelPCR1, 100ng DNA was taken from input samples and same input volume added from pull-downs. PCR1 was performed in a final volume of 50 μ L with 25 μ L of MyTaq HS Red mix and 0.5 μ M of each primer (indelPCR: TAC00012 and TAC0007 or bcPCR: TAC0162 and TAC0007). 5 μ L of PCR1 was taken for PCR2 with 25 μ L of MyTaq Red Mix and 0.5 μ M of each primer (TAC0009 and TAC0159) for 12 PCR cycles (3 cycles with 58°C annealing followed by 9 cycles with 65°C annealing). PCR products were pooled, purified and quantitated as described above, and sequenced on an Illumina MiSeq.

For calibrated ChIP, clone 5 IPR barcode counts were normalized by clone 9 IPRs (spike-in clone). Each sample was normalized for library size (total clone 9 barcode counts) and input counts. Then, clone 9 read counts were used to normalize every sample to a reference sample. Normalized barcode counts were used for further plotting and analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

The data generated by the indel scoring was further processed and analyzed in *R* (R Core Team, 2020) with *Rstudio* (RStudio Team, 2020), the figures were generated using *ggplot2* (Wickham, 2016). The main packages and software are listed in the key resource table and an extensive bibliography is available in all the scripts (github & OSF for markdowns). Statistical details for individual experiments have been provided in the main text, figure legends, and Method Details, as well as in the available markdown documents on OSF.

Mapping of IPR integration sites by inverse PCR

Mapping of IPR integration sites was performed in two replicates by inverse PCR (iPCR) followed by 2 \times 150 bp paired end sequencing on an Illumina HiSeq2500 as previously described (Akhtar et al., 2014). Linking of IPR barcodes to the integration sites was adapted from Akhtar et al. (2013). Reads of both replicates were pooled. The first read in each read pair was used to extract the barcode. This was done using the 'GTCACAAGGGCCGCCACAAC' constant sequence followed by a regular expression 'TCGAG[ACGT]{16}TGATC'. From the sequence matching this regular expression, the 16 bp barcode was extracted. To identify barcodes arising from mutations during PCR and sequencing, *starcode v1.1* (Zorita et al., 2015) was used with the sphere clustering setting and a maximum Levenshtein distance of 2. The second read of each pair was used to locate the site of integration after removing the 'GTACGTCACAATATGATTATCTTTCTAGGGTTAA' sequence matching the transposon arm. The flanking sequence was aligned to GRCh38 using *bowtie2* (Langmead and Salzberg, 2012) using the *very-sensitive-local* option (20 seed extension attempts, up to 3 re-seed attempts for repetitive seeds, 0 mismatches per seed, with a seed-length of 20 and using a multi-seed function: $f(x) = 1 + 0.5\sqrt{x}$). Locations of integration sites were required to be supported by at least 5 reads with an average mapping quality larger than 10 at the primary location, having at least 95% of the reads located at this locus, with not more than 2.5% of the reads at a secondary location.

TagMeppR

To infer the integration sites, we created a software package (TagMeppR) specially designed to map, identify and visualize tagmentation mapping reads. This R-package is available at <https://github.com/robinweide/tagmeppr>. TagMeppR enables the creation of a hybrid reference genome made up of a full genome with additional two pseudo-chromosomes, consisting of the two PiggyBac ITRs. Next, the align-tool enables the mapping of read-pairs from both indexes separately with BWA MEM (arXiv:1303.3997v2 [q-bio.GN] - <https://github.com/lh3/bwa>). After mapping, suspected PCR-duplicate reads are removed with SAMtools *rmdup* (Li et al., 2009). For each putative IPR, TagMeppR will check if (1) reads from both indexes are enriched and (2) whether the read-densities are on opposite sides of the IPR. To assess the latter, we perform a binomial test on reads up- and downstream of the IPR. Next, we filter out IPRs in which both sets are biased on the same side of the IPR. We compute an IPR-specific p value with the conservative Edgington's sum-p method and compute the family-wise error rate. TagMeppR also enables the visualization of single insertions and genome-wide insertion maps.

The two IPRs that were mapped only with Tagmeppr were confirmed by PCR (with TAC0065 & TAC0128 for IPR5 [chr20:18569153]; TAC0065 & TAC0126 (Table S4) for IPR8 [chr7:13259711] followed by Sanger sequencing (see TIDE method) to identify the barcodes associated to them.

Indel scoring

Indel reads after induction and repair of DSBs consist of single-end reads of 150 bp that span both the DSB site and the barcode. Indel scoring was adapted from Brinkman et al. (2018). Barcodes were extracted from the reads with an in-house script using functions of *cutadapt 1.11* (Martin, 2011). The 16 bp barcode was located using the 20 bp constant 'GTCACAAGGGCCGGCCACAA' sequence preceding the barcode and 'TGATCGGT' expected immediately after the barcode. For the 20 bp constant sequence, 2 mismatches were allowed.

To determine the indel size in each read, we used the distance (number of nucleotides) between two fixed sequences at the start and at the end of the read. The indel size was calculated as the difference between the measured distance and the expected distance based on the wild-type sequence. We used the following anchor sequences: before the break site, 'TGATCGGT' and after the break site, 'TGGCCTT', 'GGAGTT', 'CACTTT', 'ATTATG', 'GAAGTA', 'ATTAGA' and 'GGAAGA'. The most proximal match found with the selection of these sequences found after the break site of the specific guide was used to calculate the indel size by subtracting the expected location from the observed location. Insertions and deletions have indel sizes > 0 and < 0, respectively. Wild-type sequence is defined as indel size 0. Point mutations were not analyzed. Per replicate experiment we observed a mean 17.3% (95% CI: 16.7%–17.8%) sequence reads in which we could not find a match with the constant parts; we discarded these reads in subsequent analyses. Potentially these represent large deletions, complex mutations, sequencing errors or a combination thereof.

Per barcode, the reads of all technical replicates were pooled if applicable. Mutated barcodes were included or discarded as described above for the mapping of IPR integrations. Because in the cell pools not all IPRs are equally represented (the cell pools consist of a mix of clones that each carry different IPRs, and some cell clones grow faster than others), we then discarded IPRs that were too underrepresented to provide reliable data. Specifically, we required that each IPR is represented by at least 50 cells among the ~100,000 cells that were used in each experiment. We assumed an average of 6 IPRs per cell. Accordingly, the number of total reads per IPR was divided by the library size and multiplied by 6 * 100,000 to obtain the estimated number of cells for each IPR. IPRs for which this score was > 50 were used for subsequent analyses. Then each replicate was normalized over library size and biological replicates were averaged. The frequency of each indel type as proportion of total reads was calculated on that average. Pathway frequency per IPR was calculated as a proportion of the specific mutation over all indels (excluding wild-type sequences).

For LBR2, 7bp deletions were classified as MMEJ, and 1bp insertions were classified as NHEJ. For the other guides, a slightly more complicated approach was used. For these guides all specific deletions with a microhomology at the site of deletion of 2 or more nucleotides were classified as MMEJ. Indels with an indel ratio of at least 0.01 in the DMSO setting and a significant decrease (adjusted p value < 0.05) in the M3814 inhibitor setting (one-sided Wilcoxon test) were classified as NHEJ.

Tagmentation to identify rearrangements

Similar methods were used to identify the barcode of the IPR for each read pair, similar to the iPCR mapping. After the barcode was identified, the sequence before the targeted break-site was removed using *fastx_trimmer* (http://hannonlab.cshl.edu/fastx_toolkit/). Only read pairs containing a barcode found in the pool were kept. The mate pair starting from the Tn5 part was scanned for the sequence of the Tn5 constant part and this sequence was removed whenever there was a match. The mate pair starting at the barcode was also scanned for the reverse complement of the Tn5 constant part, reads containing that fragment were discarded. The mate pairs were then tidied up using *BBTools* (Bushnell et al., 2017). The parts of both mates that were left were aligned to an *in silico* engineered genome in which transposon sequence was added at the previously identified sites of integration. The resulting bam file was converted to a bed file with the start and end of the complete fragment using *bedtools* (Quinlan and Hall, 2010) and *awk* (Aho et al., 1992). This bed file was used to identify overlap with each *in silico* engineered transposon site. The fragments with overlap of the transposon were used to quantify the small indels using the same method as for the general indel scoring, except with additional anchor sequences after the break site (TCTGA, CTAGC, GTTGA, TCTAT, AAGTT, AGAAC, TCGTA, AAGTC, TGAAT, AGTGA, ACGCC, AGCTC, TGCAC, GAAAG, TGCAT, ACGCA, GGGTT). Fragments without any overlap with any of the transposon sites were used to identify putative transposition and big deletion events (> 480bp) (Figure S2J, 3ii). For

the unique event counts of these fragments, PCR duplicates were removed by grouping all fragments with the exact same start (break site) and end (tagmentation adaptor) position, or 1 nucleotide difference on each side. The most abundant barcode found at these duplicates was assigned and the start of the mate pair beginning from the IPR was used as the putative side of translocation/big deletion event. To identify putative rearrangements/big deletions with the native LBR locus, we looked at the overlap of the putative rearrangement side and the gene annotation of the LBR gene. We were not able to identify rearrangements between IPRs for two reasons. First, only a fraction of the fragments was long enough to span outside of the IPR (517 bp from TAC0238 (Table S4) to the end of the transposon, see Figure S2J). Second, we detected a lot of PCR template switching between IPRs most likely due to the repetitive nature of the transposon ITR. This means that read pair 1 would pick up a specific barcode (starting from TAC0238) and read pair 2 from the Tn5 adaptor end would pick up genomic sites that is linked with another IPR. These events were detected in similar quantities in both cut and uncut setting, indicating that this is most likely a PCR artifact, and not a biological rearrangement after a DSB.

We estimated the chance of having rearrangements between IPRs to be < 1% by looking at the rearrangements with the cut-site in the native LBR locus. Simulations showed us that the chance of IPRs landing within 1Mb from each other are very small. In this simulation we assumed equal probability of landing in every 1Mb region of a fully triploid human genome. This resulted in an average 1.2 barcodes, meaning between 0 or one pair of barcodes landing within 1Mb of each other in our pools of 2000 cells. Considering this contribution to be minor we assumed rearrangements in *trans* between IPRs and the 4 LBR alleles could be used to calculate an estimate of rearrangements between IPRs. The ratio calculated for all rearrangements outside of the IPRs was multiplied by the ratio of LBR *trans* rearrangements. This was divided by the number of endogenous LBR alleles in K562 (4) and multiplied by the average of 6.8 IPRs per cell. This resulted in an estimate of 0.39% rearrangements between IPRs in *trans*. We therefore concluded that, together with very rare occurrence of IPRs being located in *cis* and within 1Mb of each other, the total percentage of rearrangements between IPRs is probably below 1%.

Preprocessing of previously published chromatin data

Published ChIP-seq data from various sources (Table S1) were re-processed for consistency. Raw sequencing data were obtained from the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>). Reads were aligned to the human genome GRCh38 using *bowtie2* with default options. Replicate datasets were processed separately, while the sequences from the input were combined. After alignment, reads were filtered on a minimum mapping quality of 30 (Tarasov et al., 2015). Duplicate reads were removed except for the reads coming from experiments using tagmentation, in which duplicates were kept. After this, genomic regions were masked based on blacklist regions identified by the ENCODE project (ENCFF419RSJ) (ENCODE Project Consortium, 2012) and putative artifact regions were identified based on the input reads using *chipseq-greylis* (<https://github.com/roryk/chipseq-greylis>) (Brown, 2020), a python implementation of GreyListChIPs. We considered ChIP-seq datasets to be of sufficient quality for our analyses if there was well annotated input and sample data available and consistent read lengths were used. Mean ChIP-seq signals for IPR integration sites were calculated by taking the sum of the reads in a region of 2 kb around the IPR using *deeptools2* (Ramirez et al., 2016), scaling input and sample counts by the smallest library size, adding a pseudo count of 1 and subsequently dividing sample over input normalized counts. After this, replicate experiments were averaged. For domain calling, ChIP-seq signals were calculated in similar fashion for bins of 5kb. *HMMt* (<https://github.com/gui11aume/HMMt>), an R package implementing a Hidden Markov model with t emission, was used to subsequently call domains.

DamID data of Lamin B1 are from Leemans et al. (2019). The DamID score was calculated by scaling counts to the smallest library size, adding a pseudocount of 1 and dividing over Dam-only. The normalized dam-only score was \log_2 -transformed before averaging between replicates to calculate the dam accessibility score. Replication timing data was obtained from the 4DN data portal in the form of read coverage for late and early fraction separately. Counts were processed in the same way as for the ChIP data. For TTseq coverage from forward and reverse tracks were summed and the lowest coverage score above zero was used as pseudo count before \log_2 -transforming and averaging between replicates. For DNase hypersensitivity data of both paired-end and single-end sequencing reactions were used from encode. Coverage tracks were used and for the single ended reaction a small pseudo count of half the minimum value above 0 was used before \log_2 transforming. Paired-end and single-end coverage was \log_2 transformed before averaging. Whole genome bisulfite sequencing tracks from encode were used and coverage was calculated and \log_2 transformed without the need for a pseudo count. Replicates were subsequently averaged. Data sources are available in Table S2.

Z-scores of above chromatin information for the clonal line was calculated by using the mean and standard deviation of the signals in the TRIP pool. For pA-DamID on the knock-out clones and clone 5, the scores were calculated in a window 10kb up and downstream of the IPR. Except for the different window size, pA-DamID scores were calculated similar to the overall DamID scores. Z-scores for pA-DamID were calculated using the mean and standard deviation of the pA-DamID score for 20kb binned tracks over the whole genome using the same formula as for the individual IPR's.

Correction of ChIP signals at IPRs for differences in cutting frequencies

Across IPRs, differences in ChIP signals of repair proteins may be confounded by differences in DSB frequencies. We assume that in a population of cells, an IPR that is cut twice as often as another IPR also yields a ChIP signal that is twice as strong. On top of this simple relationship there may be quantitative effects of chromatin context on the binding of these proteins. Because we are primarily interested in the latter chromatin effects, we computationally corrected the ChIP signals for differences in cutting frequency as

follows. **1.** As an approximation of the cutting rates, we used the total indel frequency of each IPR after 72 hours, i.e., after essentially all breaks were repaired. **2.** Because sgRNA-LBR exhibits a mild saturation of indel frequencies, we corrected for this. For this we correction we used the indel frequencies of sgRNAs LBR1, 12 and 15 (which do not show this saturation), using the loess fits of the pairwise scatterplots as shown in [Figure 3C](#); we averaged the results of these three independent corrections, yielding a value TIF_{cor} for each IPR. **3.** We then fitted for each protein a model $\log_2(ChIP) \sim 1 * \log_2(TIF_{cor}) + b$ using the `nls()` function of R. This fit forces a slope of 1, modeling the above assumption. **4.** The residuals of this fit were taken as the variation in ChIP signals due to chromatin context effects (now corrected for cutting efficiency) and are shown in [Figure 4G](#).

Time series analyses

Indels in clone 5 were identified and counted as described above. Indel frequencies and MMEJ:NHEJ balance was calculated before averaging across replicates. Only IPRs mapped to a single genomic location were used (19 total). Sigmoid curves were fitted to time series data using the following formula:

$$y = ae^{-be^{-ct}}$$

Where t is time and a , b and c are parameters that determine the shape and plateau of the curve. For the decay of wild-type sequence over time, the ratio was fitted as $1-y$. Fitting was done using the `nls` package in R. Starting values 20, 10 and 0.1 were used for fitting of the parameters a , b and c , respectively.

Molecular Cell, Volume 81

Supplemental information

Impact of chromatin context on Cas9-induced

DNA double-strand break repair pathway balance

Ruben Schep, Eva K. Brinkman, Christ Leemans, Xabier Vergara, Robin H. van der Weide, Ben Morris, Tom van Schaik, Stefano G. Manzo, Daniel Peric-Hupkes, Jeroen van den Berg, Roderick L. Beijersbergen, René H. Medema, and Bas van Steensel

SUPPLEMENTAL TABLES

Supplemental Table 1, Related to the STAR Methods

Epigenome ChIP datasets used in this study.

Label	description	chip_id	sra_chip	sra_input	reference
H2AFZ	Histone H2A.Z	GSM733786	SRR227661, SRR227662	SRR227650, SRR5331211, SRR5331212, SRR5331213	(Consortium, 2012)
EZH2	EZH2 (H3K27 methyltransferase)	GSM1003576	SRR568431, SRR568432	SRR227650, SRR5331211, SRR5331212, SRR5331213	(Consortium, 2012)
H3K79me2	Histone modification, mostly on active chromatin	GSM733653	SRR227378, SRR227379	SRR227650, SRR5331211, SRR5331212, SRR5331213	(Consortium, 2012)
H4K5acK8ac	Histone modification, mostly on active chromatin	GSE113635	SRR7070730, SRR7070731	SRR7070732	(Ott et al., 2018)
H3K9me2	Histone modification, specific type of heterochromatin	GSM1846169, GSM2152591	SRR2148301, SRR3503783	SRR2148307	(Salzberg et al., 2017)
CTCF	Insulator and looping factor	GSM1782717, GSM1782718	SRR2085871, SRR2085872	SRR2085882, SRR2085883, SRR2085884, SRR2085885, SRR2085886	(Schmidl et al., 2015)
H3K27ac	Histone modification, mostly on active chromatin	GSM1782721, GSM1782722	SRR2085875, SRR2085876	SRR2085882, SRR2085883, SRR2085884, SRR2085885, SRR2085886	(Schmidl et al., 2015)
H3K27me3	Histone modification, specific type of heterochromatin	GSM1782749, GSM1782750	SRR2085903, SRR2085904	SRR2085882, SRR2085883, SRR2085884, SRR2085885, SRR2085886	(Schmidl et al., 2015)
H3K36me3	Histone modification, mostly on active transcription units	GSM1782723, GSM1782724	SRR2085877, SRR2085878	SRR2085882, SRR2085883, SRR2085884, SRR2085885, SRR2085886	(Schmidl et al., 2015)
H3K4me1	Histone modification, mostly on active chromatin	GSM2773392, GSM2773394, GSM2773396	SRR6010166, SRR6010168, SRR6010170	SRR6010181	(Shah et al., 2018)
H3K4me2	Histone modification, mostly on active chromatin	GSM2773399, GSM2773400	SRR6010173, SRR6010174	SRR6010181	(Shah et al., 2018)

H3K4me3	Histone modification, mostly on active chromatin	GSM2773401, GSM2773403, GSM2773404, GSM2773406	SRR6010175, SRR6010177, SRR6010178, SRR6010180	SRR6010181	(Shah et al., 2018)
POL2AS2	RNA Polymerase II, phosphorylated at serine 2 of heptad repeat. Marks transcribed regions.	GSM935402	SRR502194, SRR502195	SRR502641	(Consortium, 2012)
SMC3	Subunit of cohesin	GSM935310	SRR502001, SRR502002	SRR502641	(Consortium, 2012)
POL2	RNA Polymerase II. Marks transcribed regions.	GSE91721	SRR5111542, SRR5111543	SRR5111209, SRR5111210	(Consortium, 2012)
HDAC1	Histone deacetylase	GSE105837	SRR6213961, SRR6213962	SRR5111209, SRR5111210	(Consortium, 2012)
HDAC2	Histone deacetylase	GSE91451	SRR5111049, SRR5111050	SRR5111209, SRR5111210	(Consortium, 2012)
HDAC3	Histone deacetylase	GSE127356	SRR8659957, SRR8659958	SRR5111896, SRR5111897	(Consortium, 2012)

Supplemental Table 2, Related to the STAR Methods

label	description	Source ID	PMID
DNase	DNase I accessibility	ENCF413AHU, ENCF936BDN	(Consortium, 2012)
Dam	Dam methylase accessibility	4DNESTAJJM3X	(Leemans et al., 2019)
LMNB1	DamID of Lamin B1; nuclear lamina interactions	4DNESTAJJM3X	(Leemans et al., 2019)
Late repli.	Late replicating DNA regions	4DNFIBIZK6EY, 4DNFIRKXOCUW, 4DNFI5TMO13R, 4DNFIUCL6QG2	(Dekker et al., 2017)
TTseq	TT-seq; transcribed regions	<i>Bigwig tracks provided by authors</i>	(Schwalb et al., 2016)
5mC	5-methyl-cytosine	ENCF872YSC, ENCF669KCI	(Consortium, 2012)

Supplemental Table 3, Related to the STAR Methods

Name	Sequence (5' -> 3')	Target location (GRCh38)
LBR1	GAAATTTGCCGATGGTGAAG	chr1 - 225,424,045-225,424,064 LBR exon 1
LBR2	GCCGATGGTGAAGTGGTAAG	chr1 - 225,424,038-225,424,057 LBR exon 1
LBR12	GTGAAGTGGTAAGAGGTCGA	chr1 - 225,424,031-225,424,050 LBR exon 1
LBR15	TCATAATAAAGTGAAGTCCC	chr1 - 225,424,031-225,424,050 LBR exon 1
LMNA_KO1	ACTGAGAGCAGTGCTCAGTG	chr1 - 156,130,700-156,130,719 LMNA exon 2
LMNA_KO2	TCTCAGTGAGAAGCGCACGC	chr1 + 156,130,713-156,130,732 LMNA exon 2
LMNA_KO4	GGCGAGCTGCATGATCTGCG	chr1 + 156,130,738-156,130,757 LMNA exon 2
LBR_KO1	AGGCCGACATTAAGGAAGCA	chr1 - 225,422,116-225,422,135 LBR exon 2

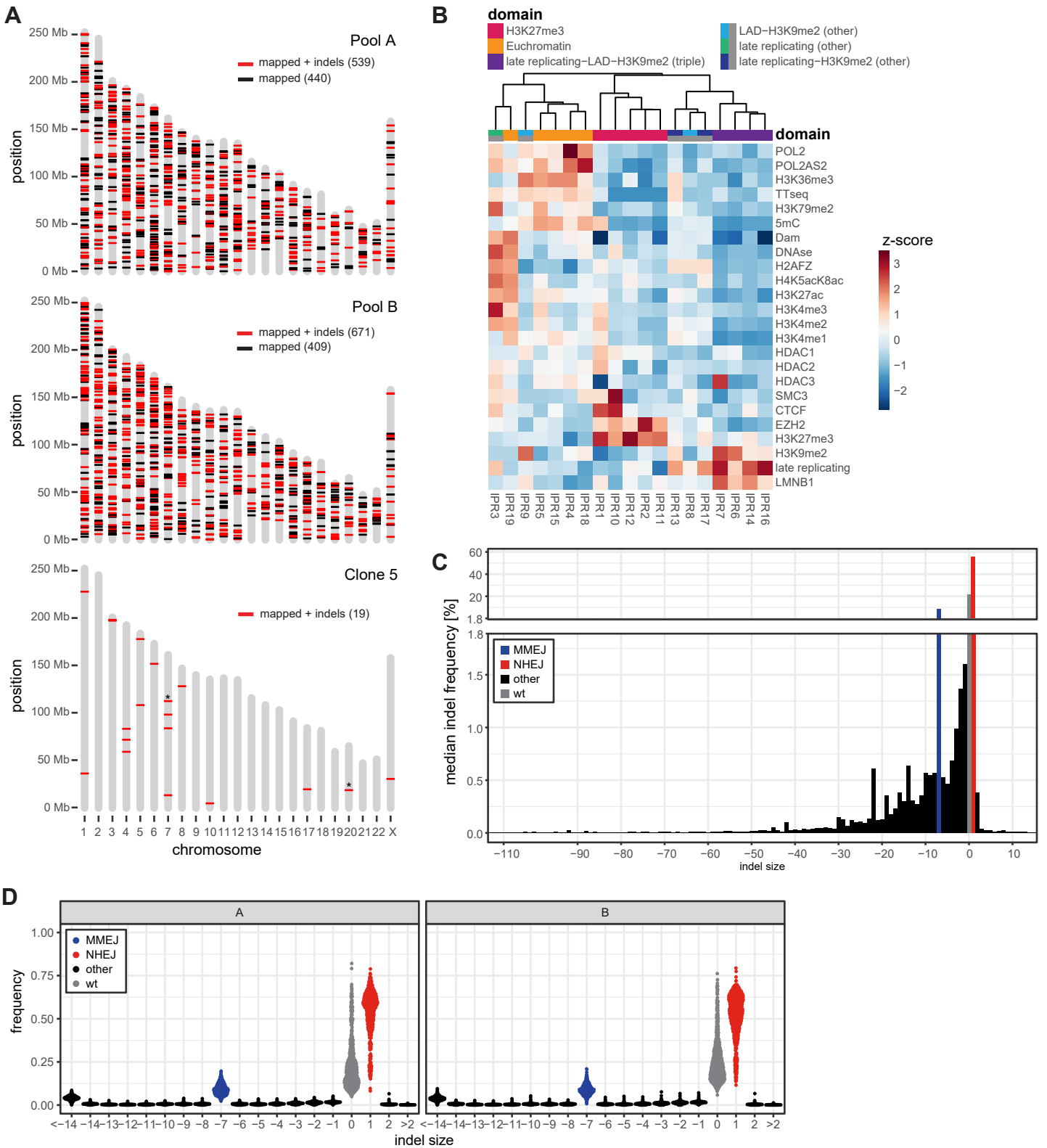
Supplemental Table 4, Related to Methods

Name	Number	Sequence (5' -> 3')
IPR-PB oligo-fw	ODS0001	ACAAC TAGAATGCTAGCGT GACTGGAGTTCA GACGTGTGCTCTTCCGATCTAATTTCTACTTC ATAATAAAGTGA ACTCCCAGGCCATCGACCT CTTACC
IPR-PB oligo-rv	ODS0002	TGATCGGTACCAACTCCAGCAGGACCATGTG ATCGAAAATGCCAAGTAGGAAATTTGCCGAT GGTGAAGTGGTAAGAGGTTCGATGGCCTGGG AG
PB_NheI_constr_fw	TAC0001	ACAAC TAGAATGCTAGCGTG
PB_KpnI_constr_rv	TAC0002	TGATCGGTACCAACTCCAG
barcode primer-fw	TAC0003	ACTGATCATGGGTACCGATCA(N)16TTGTGGC CGGCCCTTGTGACCTGCA
barcode primer-rv	TAC0004	AAAAGCGCGCATACTAGATTAACCCTAGAAA GATAATCATATTG
LBR12_oligo_fw	ODS0011	CACCGGTGAAGTGGTAAGAGGTCGA
LBR12_oligo_rv	ODS0012	AAACTCGACCTCTTACCACTTCACC
LBR15_oligo_fw	ODS0017	CACCGTCATAATAAAGTGA ACTCCC
LBR15_oligo_rv	ODS0018	AAACGGGAGTTCAC TTTATTATGAC
LBR21_oligo_fw	ODS0029	CACCGAGGCCGACATTAAGGAAGCA
LBR21_oligo_rv	ODS0030	AAACTGCTTCCTTAATGTTCGGCCTC
LMNAKO1_oligo_fw	ODS0033	CACCGACTGAGAGCAGTGCTCAGTG
LMNAKO1_oligo_rv	ODS0034	AAACCACTGAGCACTGCTCTCAGTC
LMNAKO2_oligo_f	ODS0035	CACCGTCTCAGTGAGAAGCGCACGC
LMNAKO2_oligo_rv	ODS0036	AAACGCGTTCGCTTCTCACTGAGAC
LMNAKO4_oligo_fw	ODS0039	CACCGGGCGAGCTGCATGATCTGCG
LMNAKO5_oligo_rv	ODS0040	AAACCGCAGATCATGCAGCTCGCCC
LBRKO_TIDE_fw	TAC0179	ACATAAAGCGGAAGACAAAAGGC
LBRKO_TIDE_rv	TAC0180	TGCATTTGTCTCATGAAAGATGGAT
LMNAKO_TIDE_fw	TAC0177	AGGATGCCCTCTCCTGGTAA
LMNAKO_TIDE_rv	TAC0178	CTGTGGTAGATCCCATTGGC
TIDE_endo_LBR_F	TAC0017	GTAGCCTTTCTGGCCCTAAAAT
TIDE_endo_LBR_R	TAC0018	AAATGGCTGTCTTTCCAGTAA
indelPCR1-fw-BC	TAC0007.1- 24	ACACTCTTTCCCTACACGACGCTCTTCCGATC T(N)10GTCACAAGGGCCGCCACA
indelPCR1-rv	TAC0012	GTGACTGGAGTTCAGACGTGTGCTCTTCCGA TCT
indelPCR2-fw	TAC0009	AATGATACGGCGACCACCGAGATCTACACTC TTTCCCTACACGACGCTCTTCCGATCT
indelPCR2-rv	TAC0011	CAAGCAGAAGACGGCATAACGAGATGTGACT GGAGTTCAGACGTGTGCTCTTCCGATCT
indelPCR2-rv-BC	TAC0159.1- 96	CAAGCAGAAGACGGCATAACGAGAT(N)6GTG ACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Tn5_adpt_A1	TAC0101	TCGTCGGCAGCGTCAGATGTGTATAAGAGAC AG
Tn5_adpt_A2_invdT	TAC0102	/5Phos/CTGTCTCTTATACACATC/3InvdT/
Tn5_enr_5ITR	TAC0099	CATTGACAAGCACGCCTCAC

Tn5_enr_3ITR	TAC0006	TTTTACGCATGATTATCTTTAACGTACGTC
Tn5_PCR1_F_5ITR	TAC0161	GTCTCGTGGGCTCGGAGATGTGTATAAGAGA CAGCGTCAATTTTACGCAGACTATC
Tn5_PCR1_F_3ITR	TAC0110	GTCTCGTGGGCTCGGAGATGTGTATAAGAGA CAGGTACGTCACAATATGATTATCTTTCTAG
Illumina_Nextera_N5xx	N5xx	ACGGCGACCACCGAGATCTACAC(N)8TCGTC GGCAGCGTC
Tn5_PCR2_R	TAC0103	AATGATACGGCGACCACCGA
Illumina_Nextera_N7xx	N7xx	CAAGCAGAAGACGGCATAACGAGAT(N)8GTCT CGTGGGCTCGG
Sanger_map_IPR_for	TAC0065	ATGCTAGCGTGAAGTGGAGTT
Chr20_IPR5_rev	TAC0126	AGCAGACAAGGCTCACAGCAGC
Chr7_IPR8_rev	TAC0128	ACTGCTGCCCGGCGAATTGT
Tn5_indel_enr	TAC0078	TCATTTTCGTATTTTATTTACGCCAGGG
Tn5_indel_PCR1_F	TAC0238	GTCTCGTGGGCTCGGAGATGTGTATAAGAGA CAGGTCACAAGGGCCGGCCACA
bcPCR_ChIP	TAC0162	GTGACTGGAGTTCAGACGTGTGCTCTTCCGA TCTGATCACATGGTCCTGCTGGAGTTG
qPCR_PrimerBank_PolQ_F	TAC0204	CTGCGTCGGAGTGGGAAAC
qPCR_PrimerBank_PolQ_R	TAC0205	CTGTAGGCTTGCATTCTCCTG
qPCR_PrimerBank_Lig4_F	TAC0206	AGCAAAAGTGGCTTATACGGATG
qPCR_PrimerBank_Lig4_R	TAC0207	TGAGTCCTACAGAAGGATCATGC
qPCR_PrimerBank_CtIP_F	TAC0208	CAGGAACGAATCTTAGATGCACA
qPCR_PrimerBank_CtIP_R	TAC0209	GCCTGCTCTTAACCGATCTTCT
qPCR_PrimerBank_Rad51_F	TAC0210	CAACCCATTTACCGTTAGAGC
qPCR_PrimerBank_Rad51_R	TAC0211	TTCTTTGGCGCATAGGCAACA
qPCR_PrimerBank_BRCA1_F	TAC0212	GAAACCGTGCCAAAAGACTTC
qPCR_PrimerBank_BRCA1_R	TAC0213	CCAAGTTAGAGAGTTGGACAC
qPCR_PrimerBank_BRCA2_F	TAC0214	CACCCACCCTTAGTTCTACTGT
qPCR_PrimerBank_BRCA2_R	TAC0215	CCAATGTGGTCTTTGCAGCTAT
qPCR_TBP_F	qPCR_TBP_F	CGGCTGTTTAACTTCGCTTC
qPCR_TBP_R	qPCR_TBP_R	CACACGCCAAGAAACAGTGA

SUPPLEMENTAL FIGURES

Supplemental Figure 1

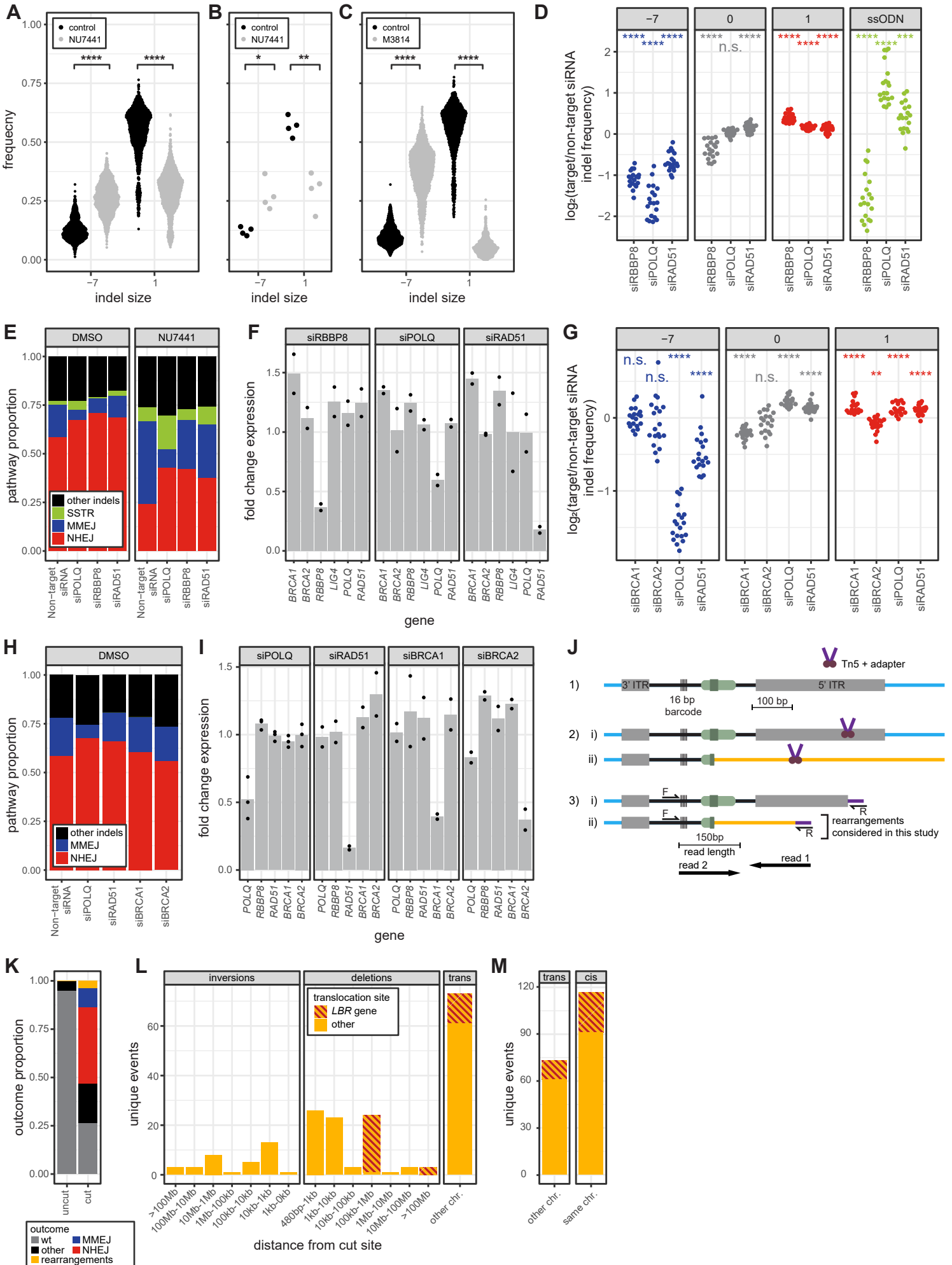


Supplemental Figure S1: Genomic location, chromatin context and indel frequencies of IPRs.

Related to Figure 2

(A) Genomic locations of the mapped IPRs in each of the two cell pools (Pools A and B) and in clone 5. Red: uniquely mapped IPRs with indel data that passed the quality criteria (also shown in [Figure 2A](#)); black: uniquely mapped IPRs that did not pass the indel quality criteria. The stars in clone 5 indicate the two IPRs mapped with tagmentation (see STAR Methods) (B) Heatmap of chromatin features at each IPR integration site in clone 5. Levels of chromatin features are represented as z-scores. IPRs are clustered based on similarities of the z-scores. (C) Median indel frequencies across all IPRs. Note that the y-axis is split in order to highlight the low frequency indels. Data are same as [Figure 2C](#) but the median frequencies are plotted over a wider range of indel sizes to illustrate that large indels are rare compared to -7 and +1. (D) Frequencies of the indels as in [Figure 2C](#) but plotted separately for cell pools A (539 IPRs) and B (690 IPRs). Data are average of 2-6 independent replicates.

Supplemental Figure 2

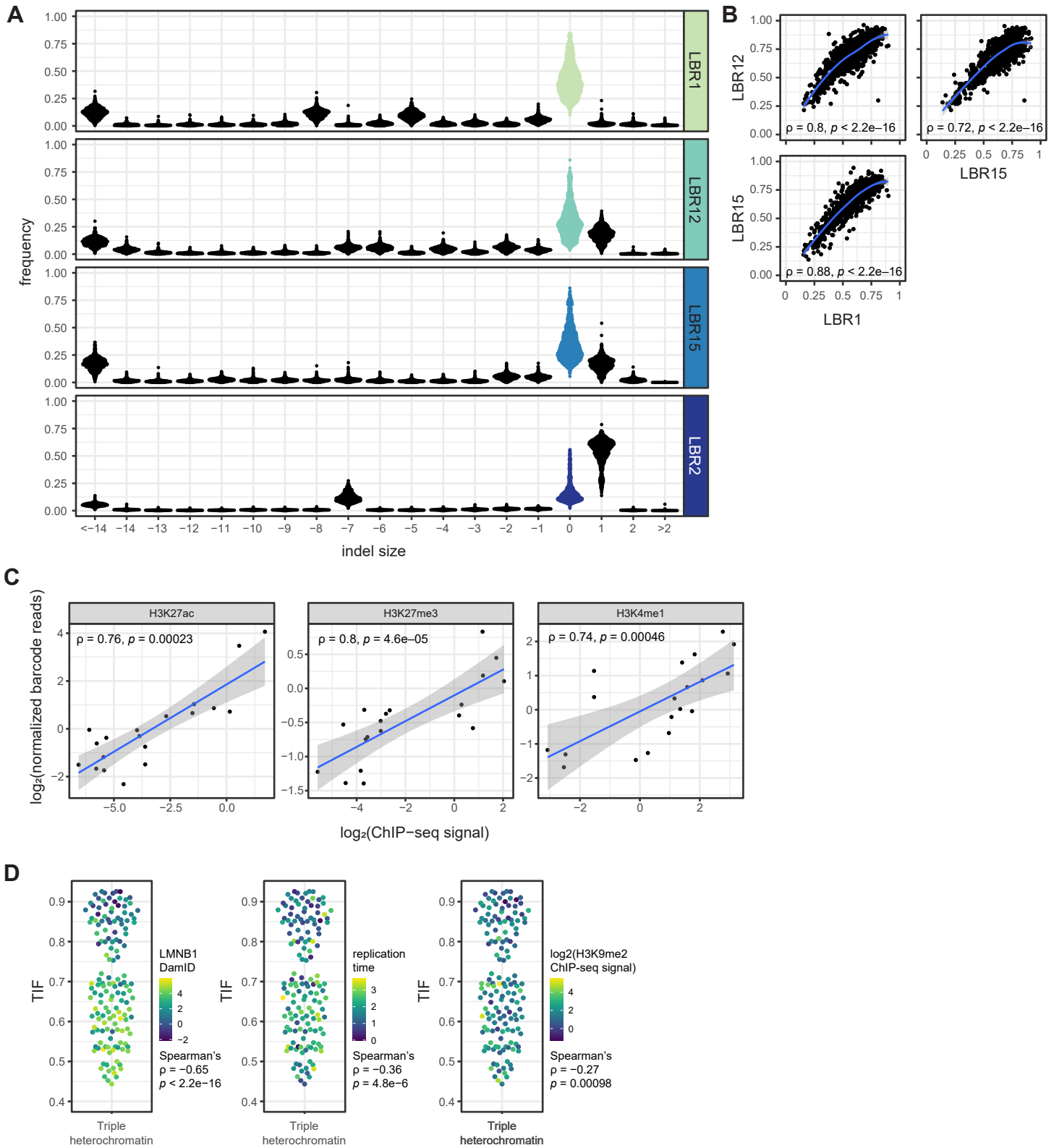


Supplemental Figure S2. Characterization of pathways that generate reporter indels.

Related to Figure 2

(A-C) Effects of chemical inhibition of DNA-PKcs on indel frequencies. P-values are according to a paired t-test. (A) Frequency of +1 and -7 indels for all IPRs (1118 IPRs) in the cell pools treated with the DNA-PKcs inhibitor NU7441 (gray, mean of $n = 2-4$) or with DMSO control (black, mean of $n = 2-4$). (B) Same data as in (A), but now shown as median indel frequencies of all IPRs, for each replicate experiment. (C) Frequency of +1 and -7 indels for all IPRs (975 IPRs) in the cell pools treated with the DNA-PKcs inhibitor M3814 (gray, mean of $n = 2$) or with DMSO control (black, $n = 2$). (D-E) Effect of knockdown of various DSB repair proteins on indel frequencies. (D) Log₂ fold-change in the frequencies of the wt, -7, +1 and ssODN-induced +2 indels at all 19 IPRs in clone 5, after siRNA-mediated knockdown of the indicated proteins compared to a control siRNA ($n = 2$). P-values are according to one sample t-test. Asterisks in panels A-D, G denote adjusted p-values: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. (E) Relative contribution of each pathway for all 19 barcodes in clone 5 combined, after indicated siRNA treatments. Red: +1 insertion (NHEJ); blue: -7 deletion (MMEJ); green: +2 insertion due to SSTR; black: other indels. Right-hand panel shows data from cells treated with NU7441 (1 μ M); left-hand panel shows data from control cells. (F) Mean mRNA expression fold change of several genes encoding DSB repair proteins after siRNA-mediated knockdown of indicated proteins, in same cells as in (E). Dots represent separate replicates. (G-H) Changes in indel frequencies and proportions after siRNA-mediated knockdown of indicated proteins. (I) Mean mRNA expression fold change in cell samples show in (G-H). Dots represent separate replicates. (J) Cartoon illustrating Tn5-mediated detection of rearrangements between barcoded IPRs and distal DNA sequences (orange) after Cas9-mediated cutting of the IPR. After purification of genomic DNA, random integration of Tn5 transposase loaded with an adaptor oligonucleotide allows for selective PCR amplification of the junctions using primers F and R. Large rearrangements were identified by paired-end sequencing of the PCR products, provided that the Tn5 integrates in a unique genomic sequence sufficiently close to the barcode (ii) but not inside the IPR (i). Furthermore, large rearrangements were only counted if their junction was located at the break site (read 2), and if read 1 matched the same genomic location as read 2. Smaller indels, which by definition do not involve a distal genomic sequence, were identified from read 2 alone. (K) Proportions of small indels in IPRs and of rearrangements between IPRs and unique genomic DNA sequences, Cas9 activation (cut) or in control cells (uncut). Rearrangements are only counted if the ligated sequence maps outside of the IPR. These values were not filtered on unique Tn5 integration sites. (L) Numbers and size distributions of detected rearrangements between IPRs and unique genomic sequences after Cas9 cutting, divided into inversions, deletions and translocations (trans). Rearrangements involving the *LBR* gene (which is also cut) are striped. Unique Tn5 integration sites are considered as unique events (M) Same data as in (L) with the insertions and deletions combined as *cis*.

Supplemental Figure 3



Supplemental Figure S3: IPRs generally adopt the chromatin state of their integration site.

Related to Figure 3

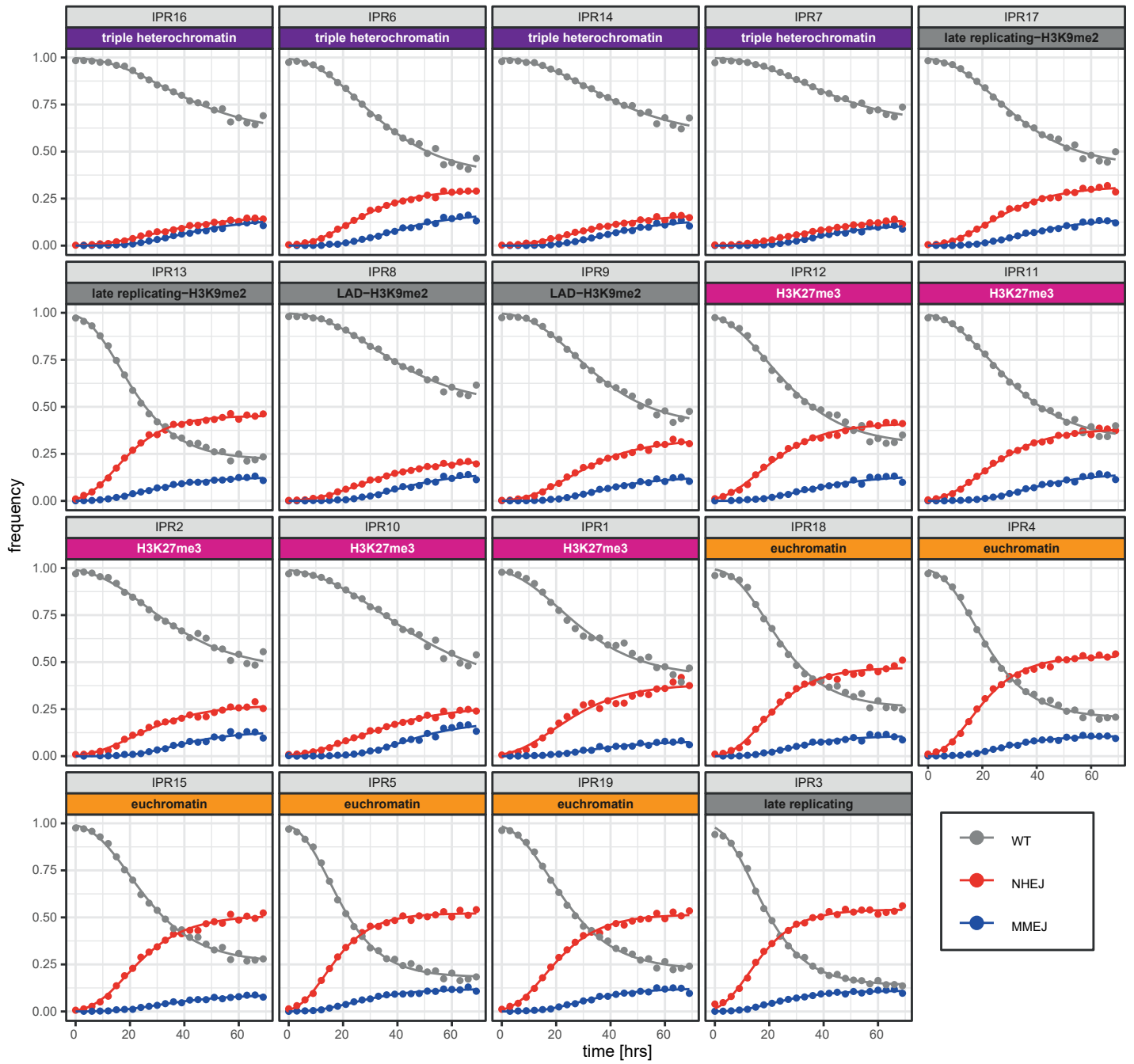
(A) Frequencies of the indels for all four tested sgRNAs targeting the IPRs. (B) Pairwise comparisons of total indel frequencies obtained with sgRNAs LBR1, LBR12 and LBR15. Blue lines are loess fits. ρ is Spearman's rank correlation coefficient. (C) IPRs generally adopt the chromatin state of their integration site. Vertical axes: \log_2 normalized barcode reads of 19 IPRs in clone 5 after ChIP of the indicated histone modifications, followed by PCR amplification of the barcodes and Illumina sequencing. Data are average of two independent replicates. Horizontal axes: \log_2 normalized signal of the indicated histone modifications in a window of 2kb centered on the IPR integration site ($n = 2$), according to public ChIP-seq data from K562 cells (see [Table S1](#)). Blue lines and gray shading show linear regression fits with 95% confidence intervals. ρ is Spearman's rank correlation coefficient.

Supplemental Figure S4: Correlations between pathway usage and chromatin features.

Related to Figure 4

(A) Distributions of MMEJ:NHEJ balances across IPRs, for each cell pool separately. (B) Scatterplots of the relative indel frequencies in the IPR cell pools versus local \log_2 H3K4me1 ChIP-seq signal (top row) or \log_2 Lamin B1 DamID signal (bottom row) at the IPR integration sites. ρ is Spearman's correlation coefficient; blue lines and grey shading show linear regression fits with 95% confidence intervals. (C) MMEJ:NHEJ balance for the four gRNAs used in this study in euchromatin and triple heterochromatin (as defined in [Figure 3E, 4D](#)), see Methods. P-values are according to Wilcoxon test. (D) Same as [Figure 4D](#), but now including heterochromatin types with 20 or fewer IPRs. Asterisks mark p-values according to the Wilcoxon test, compared to euchromatin IPRs (most left column): * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Supplemental Figure 5



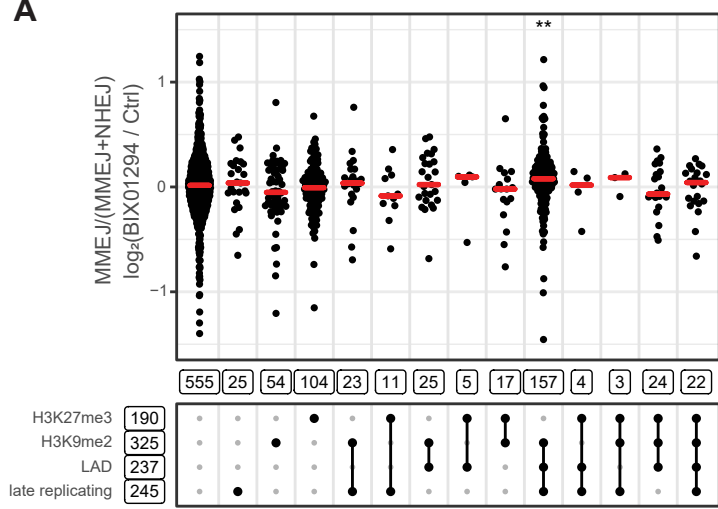
Supplemental Figure S5: Time series of -7 and +1 indel accumulation for all IPRs in clone 5.

Related to Figure 5

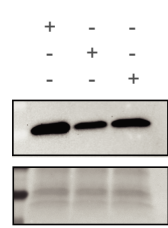
Time curves of the +1 insertion (red) and -7 deletion (blue) for all 19 individual IPRs. Dots are measured values; lines are fitted sigmoid curves.

Supplemental Figure 6

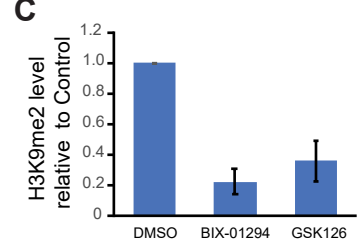
A



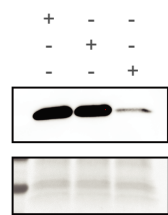
B



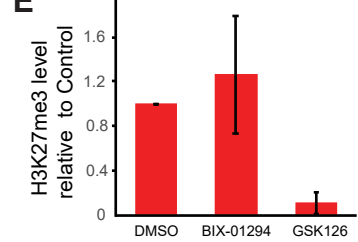
C



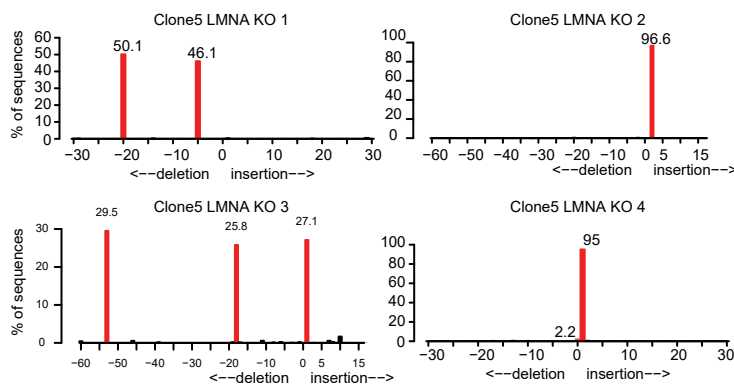
D



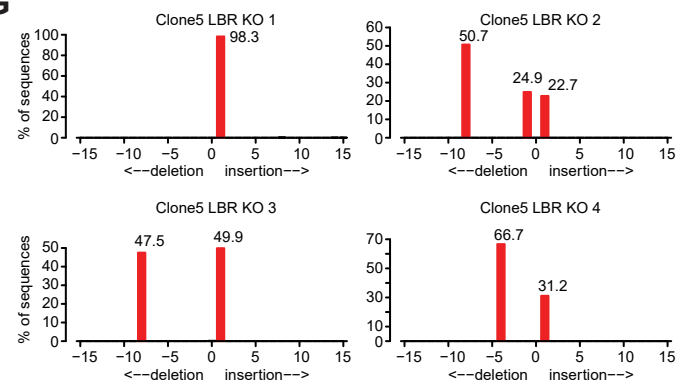
E



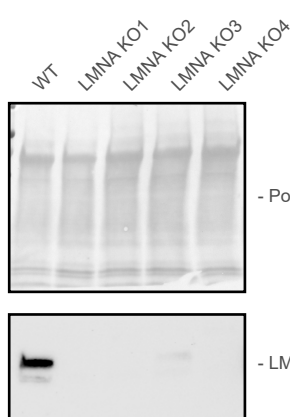
F



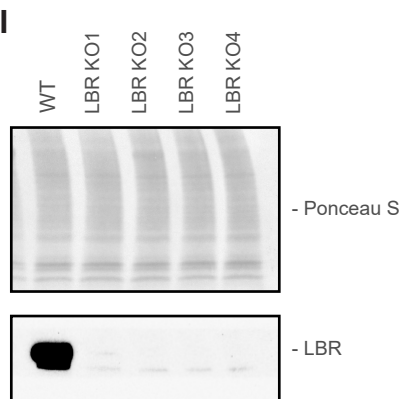
G



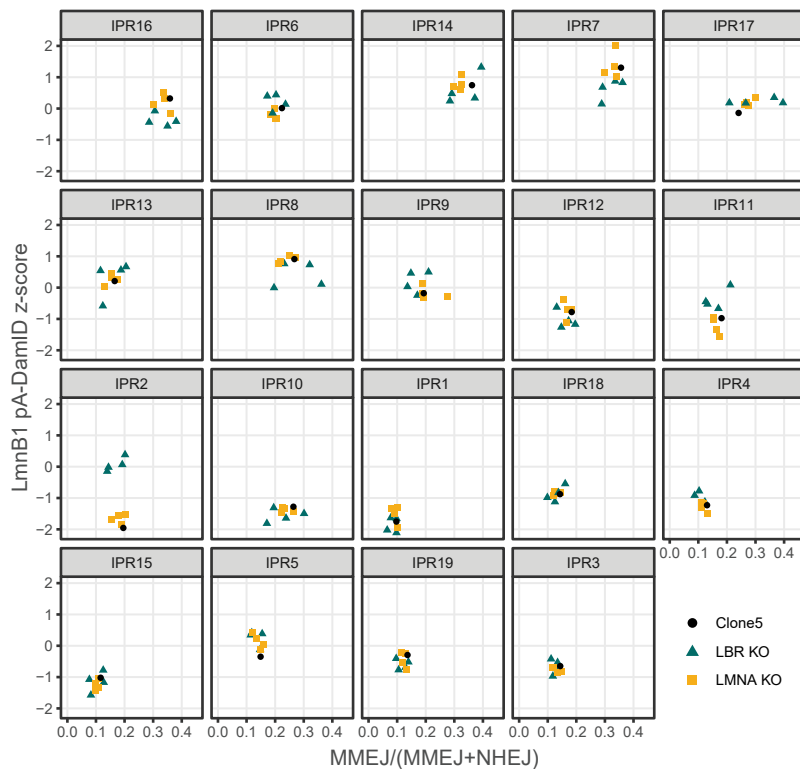
H



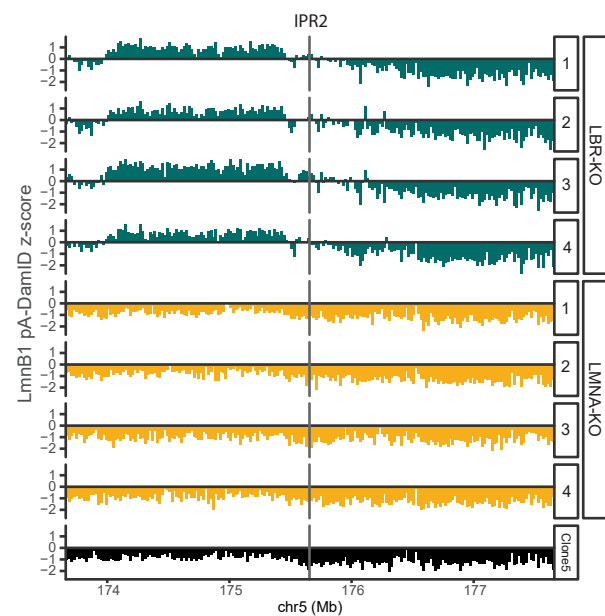
I



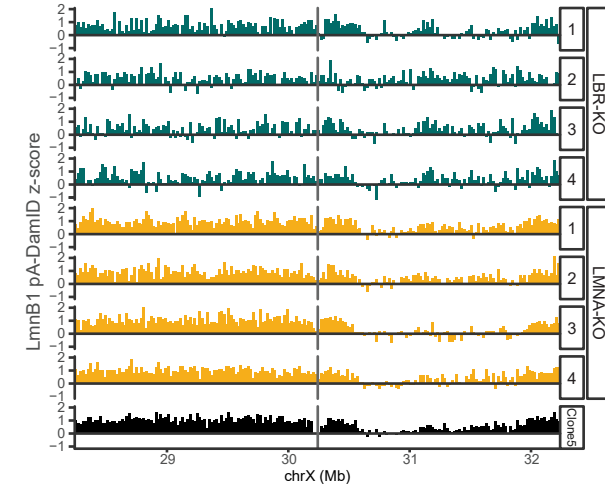
J



K



L

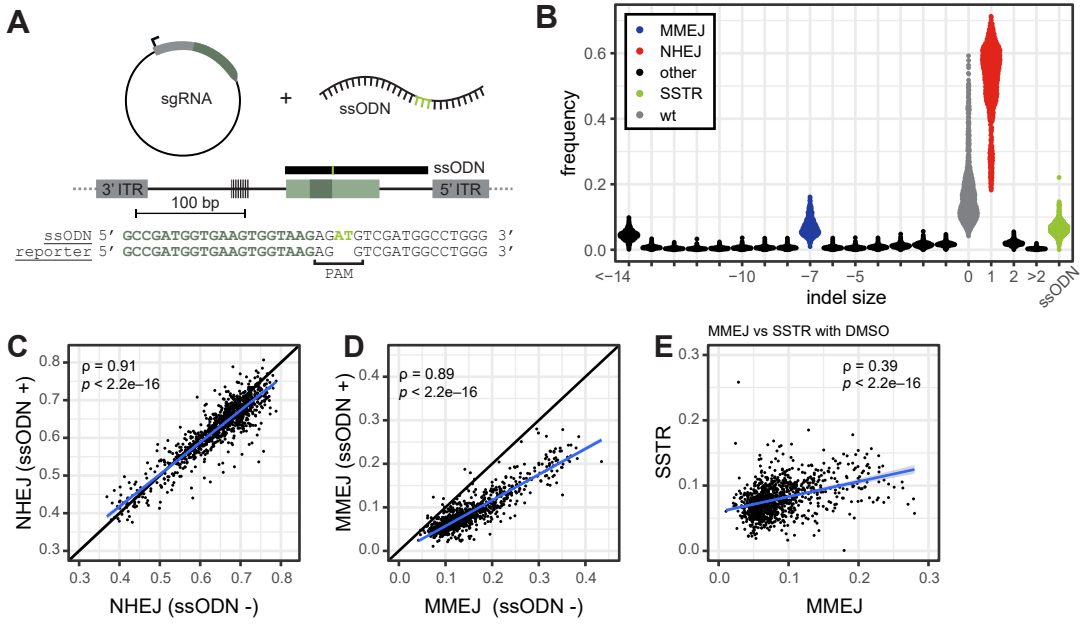


Supplemental Figure S6: Effects of heterochromatin perturbations.

Related to Figure 6

(A) Log₂ fold-change of MMEJ:NHEJ balance in BIX01294 treated cells (n = 2) compared to non-treated cells for 1029 IPRs divided by heterochromatin domains (n = 2-6; same data as in [Figure 2](#)). Wilcoxon test compared to non-heterochromatin IPRs (most left column), * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001. (B) Western blot of H3K9me2 in clone 5 cells after treatment with either 1 μM GSK126 or 500 nM BIX01294. (C) Quantification of Western blots (mean of two independent replicates, error bars show S.D.), normalized to H3K9me2 levels in control cells and red ponceau staining for protein content. (D-E) Same as B and C but for H3K27me3. (F-G) Indel patterns inside the *LMNA* (E) and *LBR* (F) genes in respective knockout sub-clones that were derived from clone 5, showing frameshifts (i.e., indel sizes that are not multiples of three) in all alleles. Note that chromosomes in K562 cells can be tri- or tetraploid. Indel spectra were obtained by TIDE (Brinkman et al., 2014). (H-I) Western blots of LMNA (H) and (I) LBR in WT (clone 5) and in the four knock-out clones. (J) Scatterplot of MMEJ:NHEJ balance compared to LMNB1 pA-DamID z-score for each IPR, averaged over 10kb up and downstream of the IPR. Black circle is clone5, green triangles are the LBR KO clones and yellow squares are LMNA KO clones. (K) Z-score of pA-DamID tracks for LMNB1 centered on IPR17 with 2Mb up and downstream. (n = 2) (L) Same as in K but for IPR2.

Supplemental Figure 7



Supplemental Figure S7: Chromatin context effects on the SSTR pathway.

Related to Figure 7

(A) Schematic of the strategy to probe SSTR simultaneously with NHEJ and MMEJ. Prior to Cas9 activation, the ssODN is co-transfected with a plasmid that encodes the LBR2 sgRNA. The ssODN (black bar) covers the reporter sequence but not the barcode, and contains a 2 bp insertion (green) at the PAM. (B) Indel frequencies generated by NHEJ, MMEJ and SSTR in 965 IPRs in the two cell pools, 64 hours after Cas9 activation (average of two replicate experiments). (C-D) Comparison of +1 (NHEJ; panel C) and -7 (MMEJ, panel D) indel frequencies in all IPRs in cell pools in the presence (+) or absence (-) of the ssODN. Black line: diagonal. (E) Comparison of MMEJ and SSTR frequencies across all IPRs in cell pools treated with the ssODN. (C-E) ρ is Pearson's correlation coefficient; blue lines and grey shading show linear regression fits with 95% confidence intervals.