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	ENCFF413AHU, ENCFF936BDN	(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, 4DNFIRKOXCUW, 4DNFI5TMO13R, 4DNFIUCL6QG2	(Dekker et al., 2017)
TTseq	TT-seq; transcribed regions	*Bigwig tracks provided by authors*	(Schwalb et al., 2016)
5mC	5-methyl-cytosine	ENCFF872YSC, ENCFF669KCI	(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	TCATAATAAAGTGAACTCCC	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

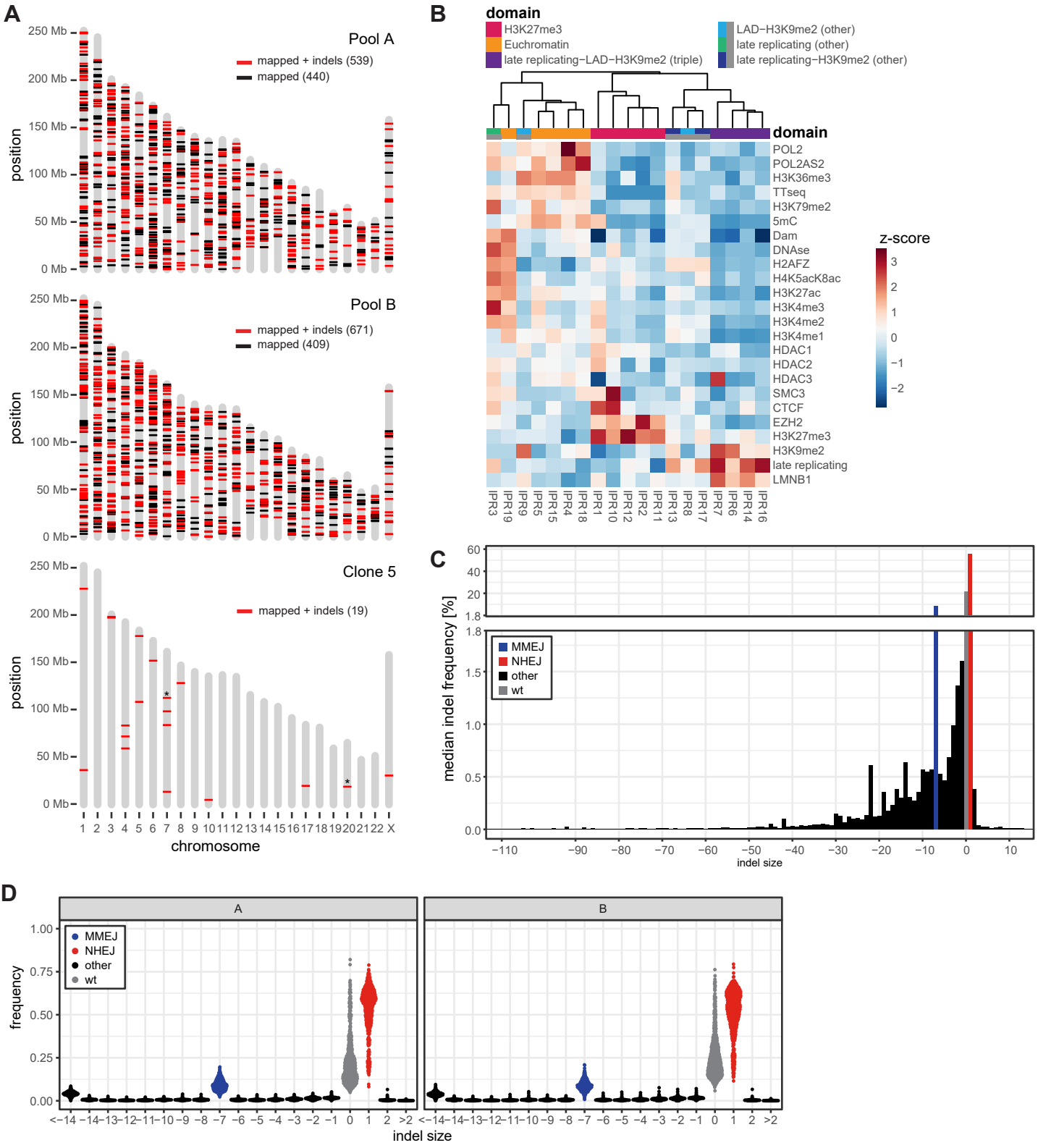
Supplemental Table 4, Related to Methods

Name	Number	Sequence (5' -> 3')
IPR-PB oligo-fw	ODS0001	ACAACTAGAATGCTAGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAATTTCTACTTCATAATAAAGTGAACTCCCAGGCCATCGACCTCTTACC
IPR-PB oligo-rv	ODS0002	TGATCGGTACCAACTCCAGCAGGACCATGTGATCGAAAATGCCAAGTAGGAAATTTGCCGATGGTGAAGTGGTAAGAGGTCGATGGCCTGGGAG
PB_NheI_constr_fw	TAC0001	ACAACTAGAATGCTAGCGTG
PB_KpnI_constr_rv	TAC0002	TGATCGGTACCAACTCCAG
barcode primer-fw	TAC0003	ACTGATCATGGGTACCGATCA**(N)16**TTGTGGCCGGCCCTTGTGACCTGCA
barcode primer-rv	TAC0004	AAAAGCGCGCATACTAGATTAACCCTAGAAAGATAATCATATTG
LBR12_oligo_fw	ODS0011	CACCGGTGAAGTGGTAAGAGGTCGA
LBR12_oligo_rv	ODS0012	AAACTCGACCTCTTACCACTTCACC
LBR15_oligo_fw	ODS0017	CACCGTCATAATAAAGTGAACTCCC
LBR15_oligo_rv	ODS0018	AAACGGGAGTTCACTTTATTATGAC
LBR21_oligo_fw	ODS0029	CACCGAGGCCGACATTAAGGAAGCA
LBR21_oligo_rv	ODS0030	AAACTGCTTCCTTAATGTCGGCCTC
LMNAKO1_oligo_fw	ODS0033	CACCGACTGAGAGCAGTGCTCAGTG
LMNAKO1_oligo_rv	ODS0034	AAACCACTGAGCACTGCTCTCAGTC
LMNAKO2_oligo_f	ODS0035	CACCGTCTCAGTGAGAAGCGCACGC
LMNAKO2_oligo_rv	ODS0036	AAACGCGTGCGCTTCTCACTGAGAC
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	AAATGGCTGTCTTTCCCAGTAA
indelPCR1-fw-BC	TAC0007.1-24	ACACTCTTTCCCTACACGACGCTCTTCCGATCT**(N)10**GTCACAAGGGCCGGCCACA
indelPCR1-rv	TAC0012	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
indelPCR2-fw	TAC0009	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
indelPCR2-rv	TAC0011	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
indelPCR2-rv-BC	TAC0159.1-96	CAAGCAGAAGACGGCATACGAGAT**(N)6**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Tn5_adpt_A1	TAC0101	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
Tn5_adpt_A2_invdt	TAC0102	/5Phos/CTGTCTCTTATACACATC/3InvdT/
Tn5_enr_5ITR	TAC0099	CATTGACAAGCACGCCTCAC

Tn5_enr _3ITR	TAC0006	TTTTACGCATGATTATCTTTAACGTACGTC
Tn5_PCR1_F_5ITR	TAC0161	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGTCAATTTTACGCAGACTATC
Tn5_PCR1_F_3ITR	TAC0110	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTACGTCACAATATGATTATCTTTCTAG
Illumina_Nextera_N5xx	N5xx	ACGGCGACCACCGAGATCTACAC**(N)8**TCGTCGGCAGCGTC
Tn5_PCR2_R	TAC0103	AATGATACGGCGACCACCGA
Illumina_Nextera_N7xx	N7xx	CAAGCAGAAGACGGCATACGAGAT**(N)8**GTCTCGTGGGCTCGG
Sanger_map_IPR_for	TAC0065	ATGCTAGCGTGACTGGAGTT
Chr20_IPR5_rev	TAC0126	AGCAGACAAGGCTCACAGCAGC
Chr7_IPR8_rev	TAC0128	ACTGCTGCCCGGCGAATTGT
Tn5_indel_enr	TAC0078	TCATTTCGTATTTTATTTACGCCAGGG
Tn5_indel_PCR1_F	TAC0238	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGTCACAAGGGCCGGCCACA
bcPCR_ChIP	TAC0162	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATCACATGGTCCTGCTGGAGTTG
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	CAACCCATTTCACGGTTAGAGC
qPCR_PrimerBank_Rad51_R	TAC0211	TTCTTTGGCGCATAGGCAACA
qPCR_PrimerBank_BRCA1_F	TAC0212	GAAACCGTGCCAAAAGACTTC
qPCR_PrimerBank_BRCA1_R	TAC0213	CCAAGGTTAGAGAGTTGGACAC
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

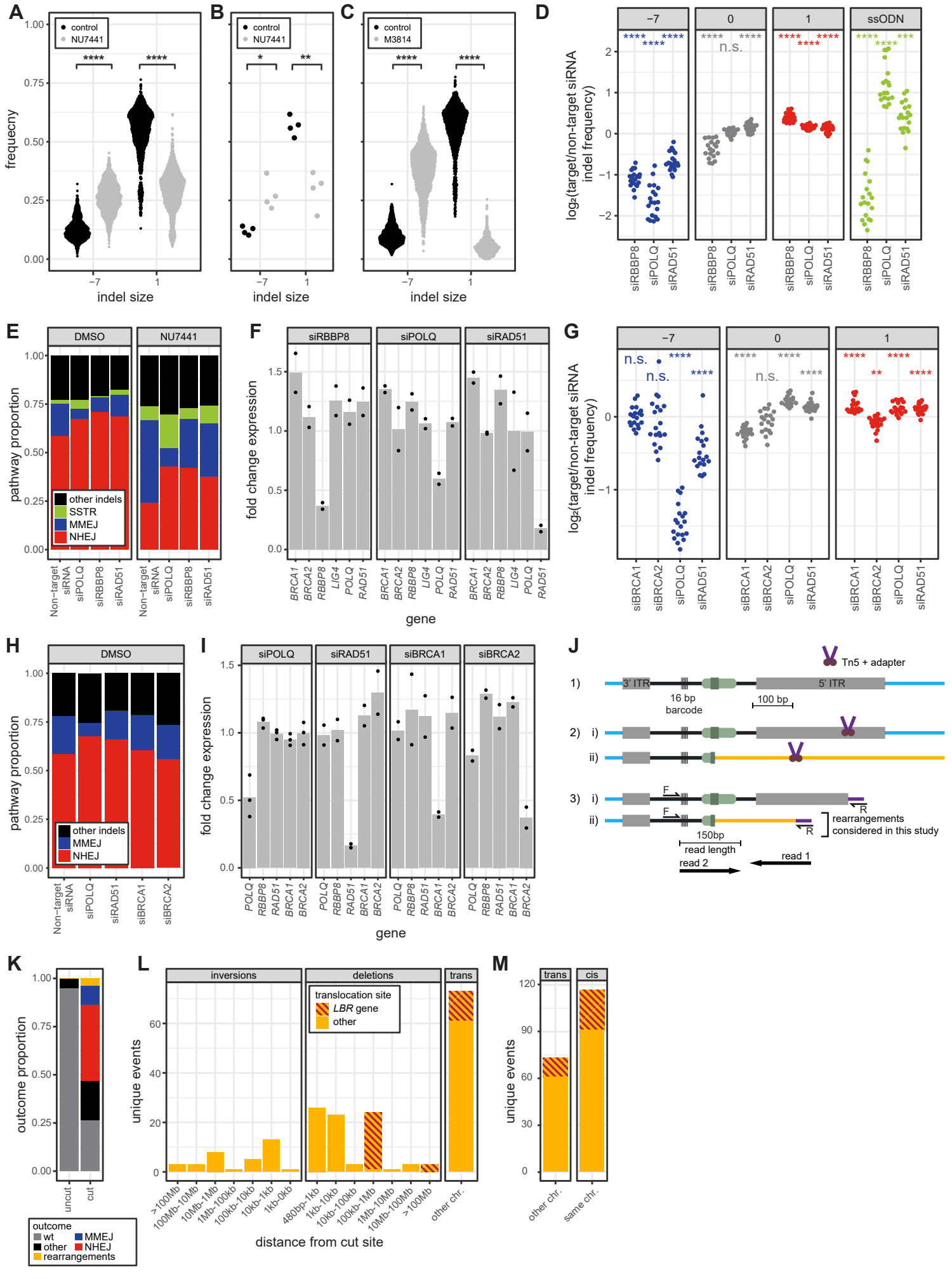


Supplemental Figure S1: Genomic location, chromatin context and indel frequencies of IPRs.

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

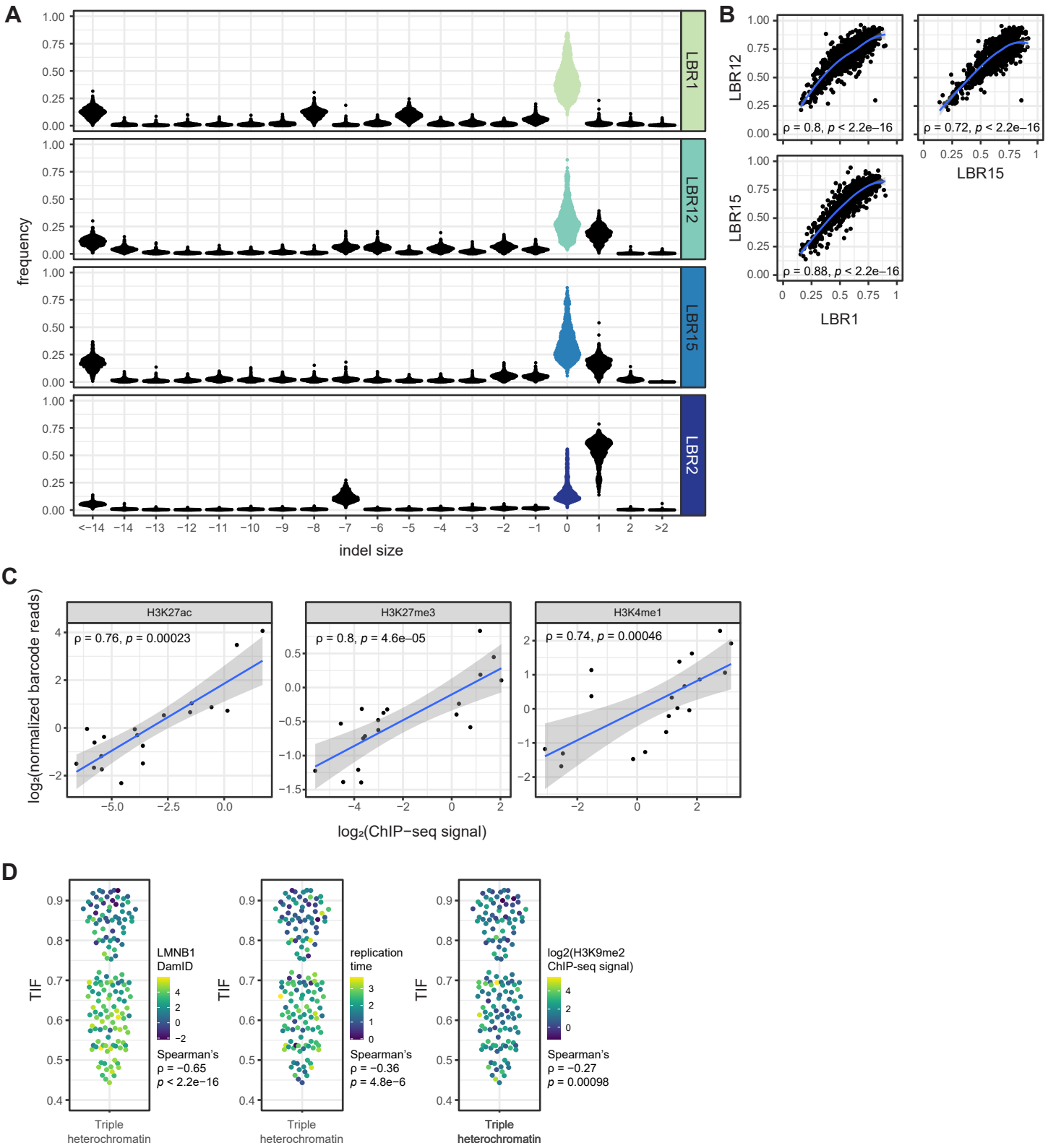


Supplemental Figure S2. Characterization of pathways that generate reporter indels.

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**) Log_2 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 μM); 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

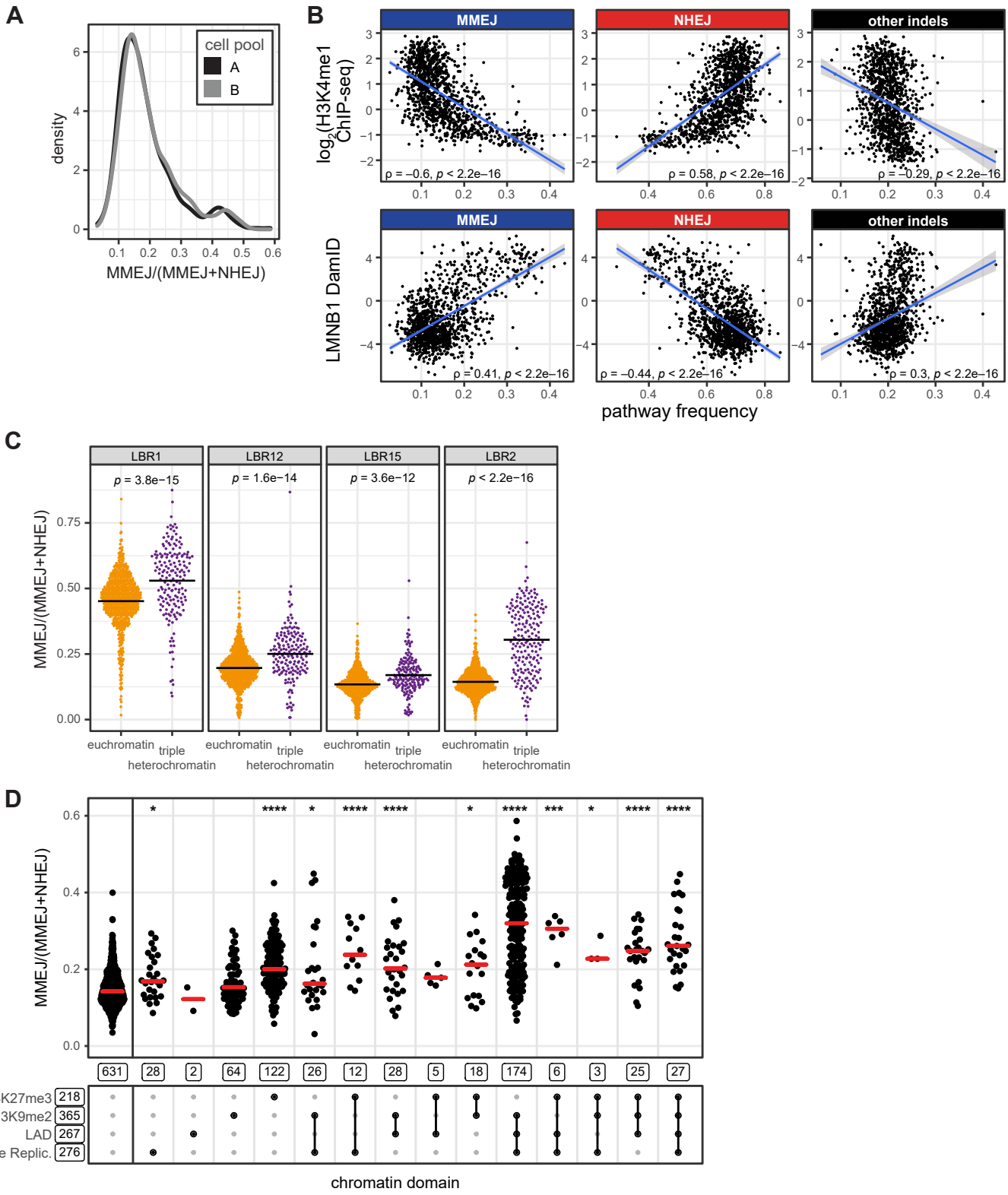


Supplemental Figure S3: IPRs generally adopt the chromatin state of their integration site.

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 4

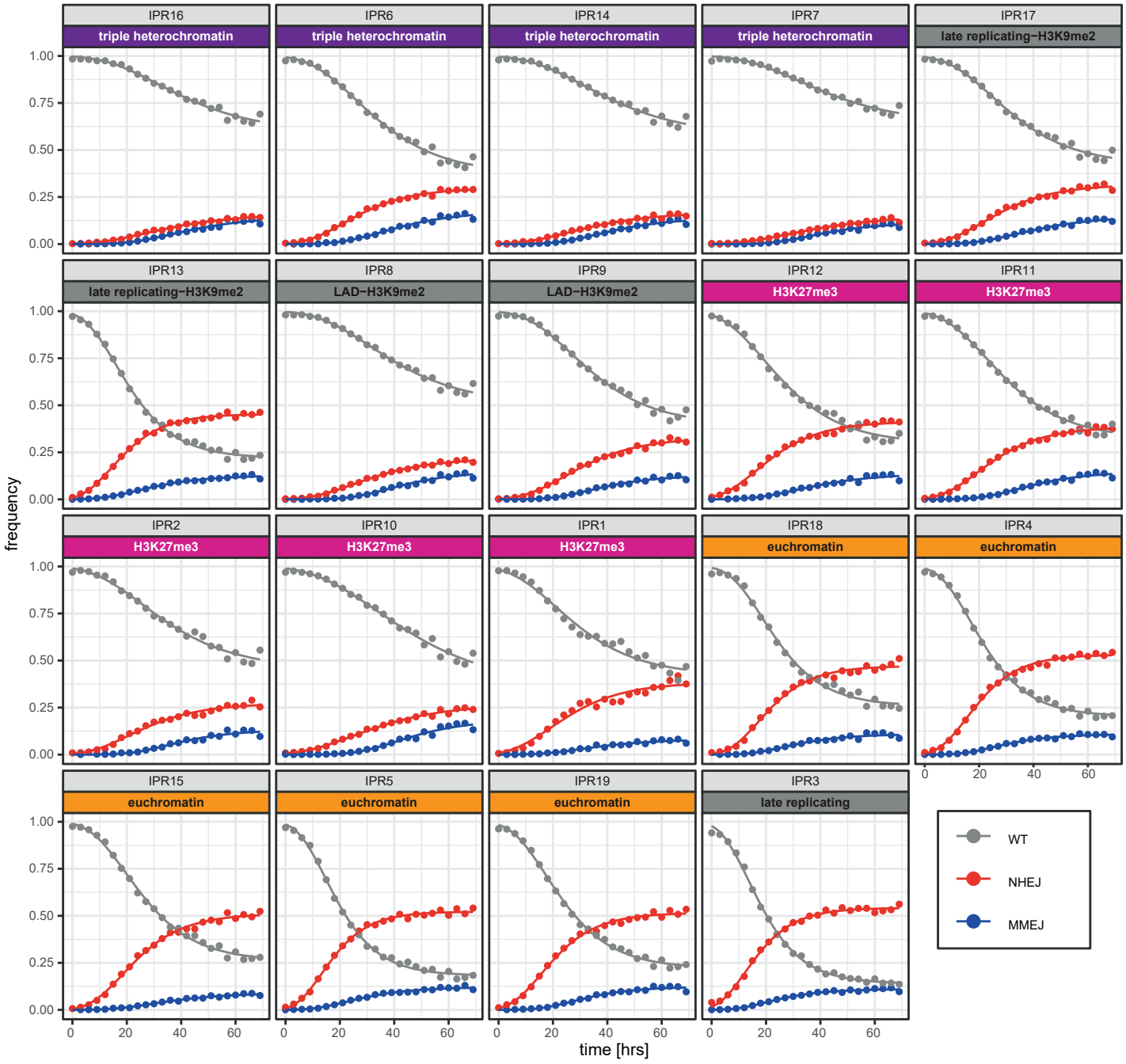


Supplemental Figure S4: Correlations between pathway usage and chromatin features.

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

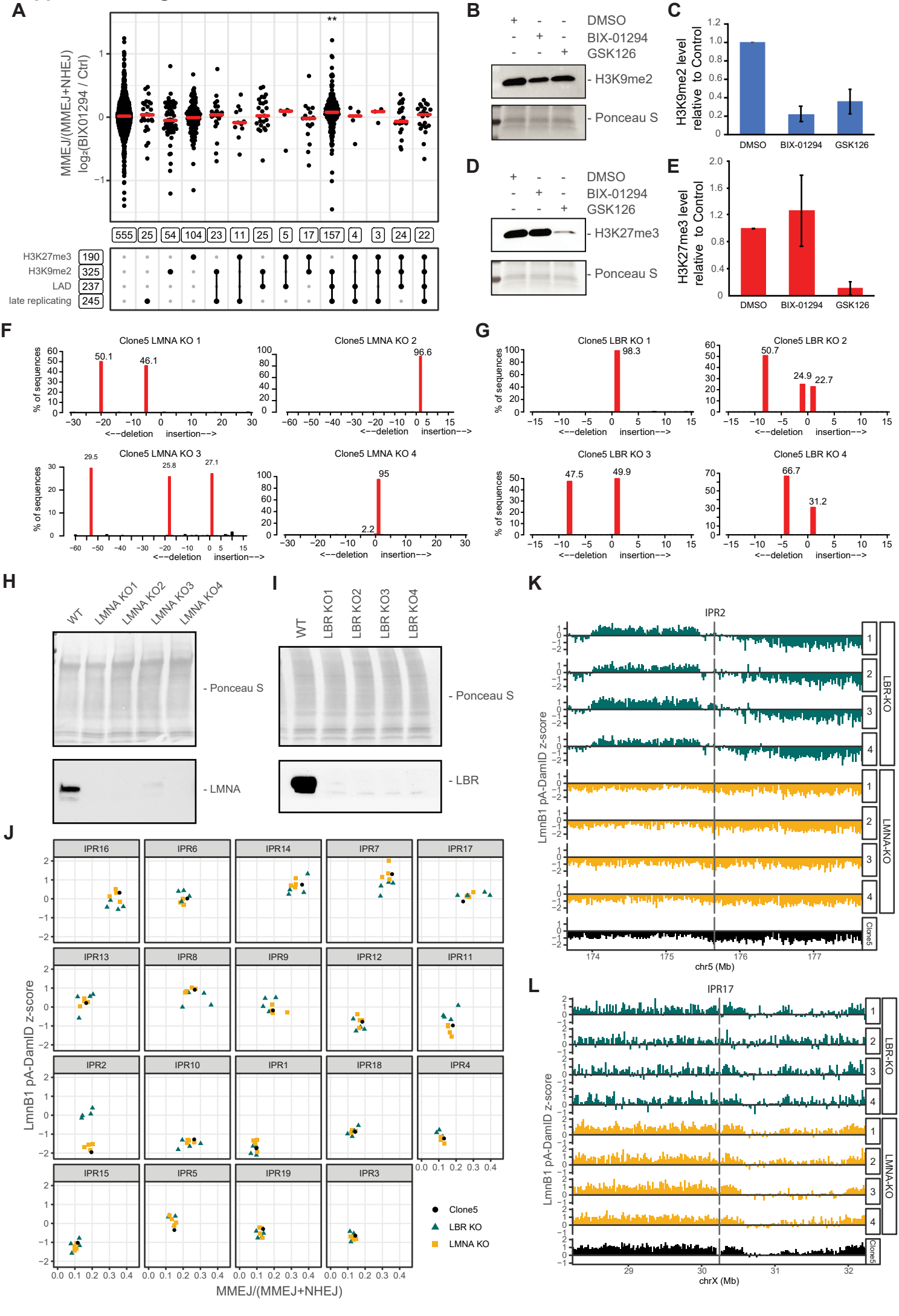


Supplemental Figure S5: Time series of -7 and +1 indel accumulation for all IPRs in clone 5.

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

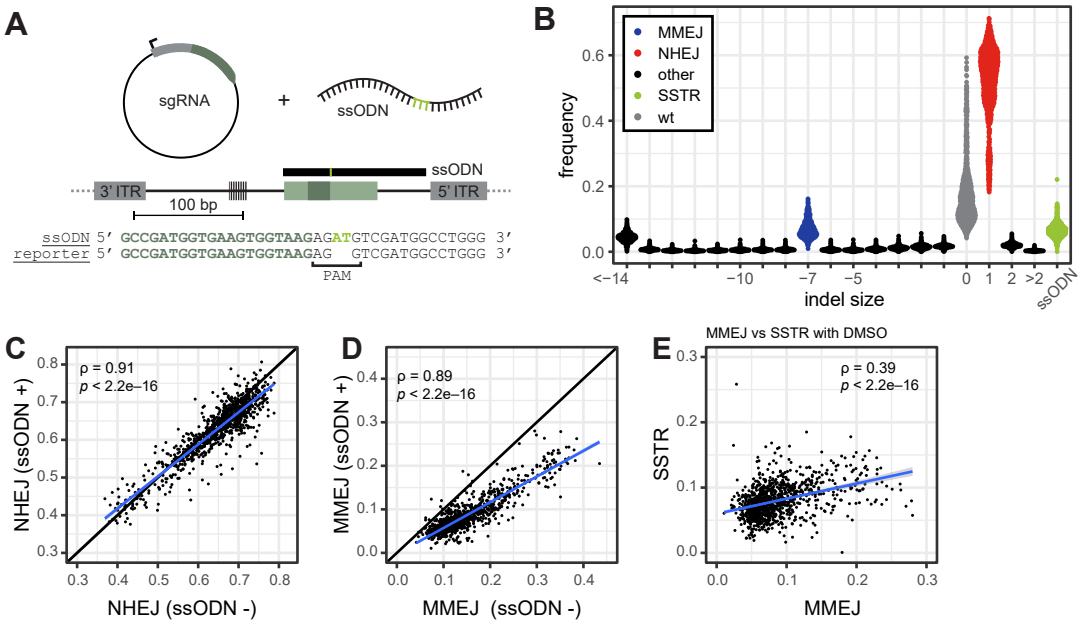


Supplemental Figure S6: Effects of heterochromatin perturbations.

Related to Figure 6

(**A**) Log_2 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.