

Supplementary Materials for

Global detection of DNA repair outcomes induced by CRISPR-Cas9

Mengzhu Liu^{1,3}, Weiwei Zhang^{1,3}, Changchang Xin^{1,3}, Jianhang Yin¹, Yafang Shang², Chen Ai¹, Jiabin Li¹, Fei-long Meng², Jiazhi Hu^{1,*}

This PDF file includes:

Supplementary Figure S1 Analyzing repair outcomes by PEM-Q

Supplementary Figure S2 Repair outcomes at *c-Myc* and *Bcr* loci

Supplementary Figure S3 Microhomologies are commonly used in large deletions

Supplementary Figure S4 Insertions at the target sites

Supplementary Figure S5 Translocations enriched at recurrent DSBs in the genome

Supplementary Figure S6 Editing events detected by primers with different distances from the cut-site

Supplementary Table S1 Primer sequences for PEM-seq and 3C-HTGTS

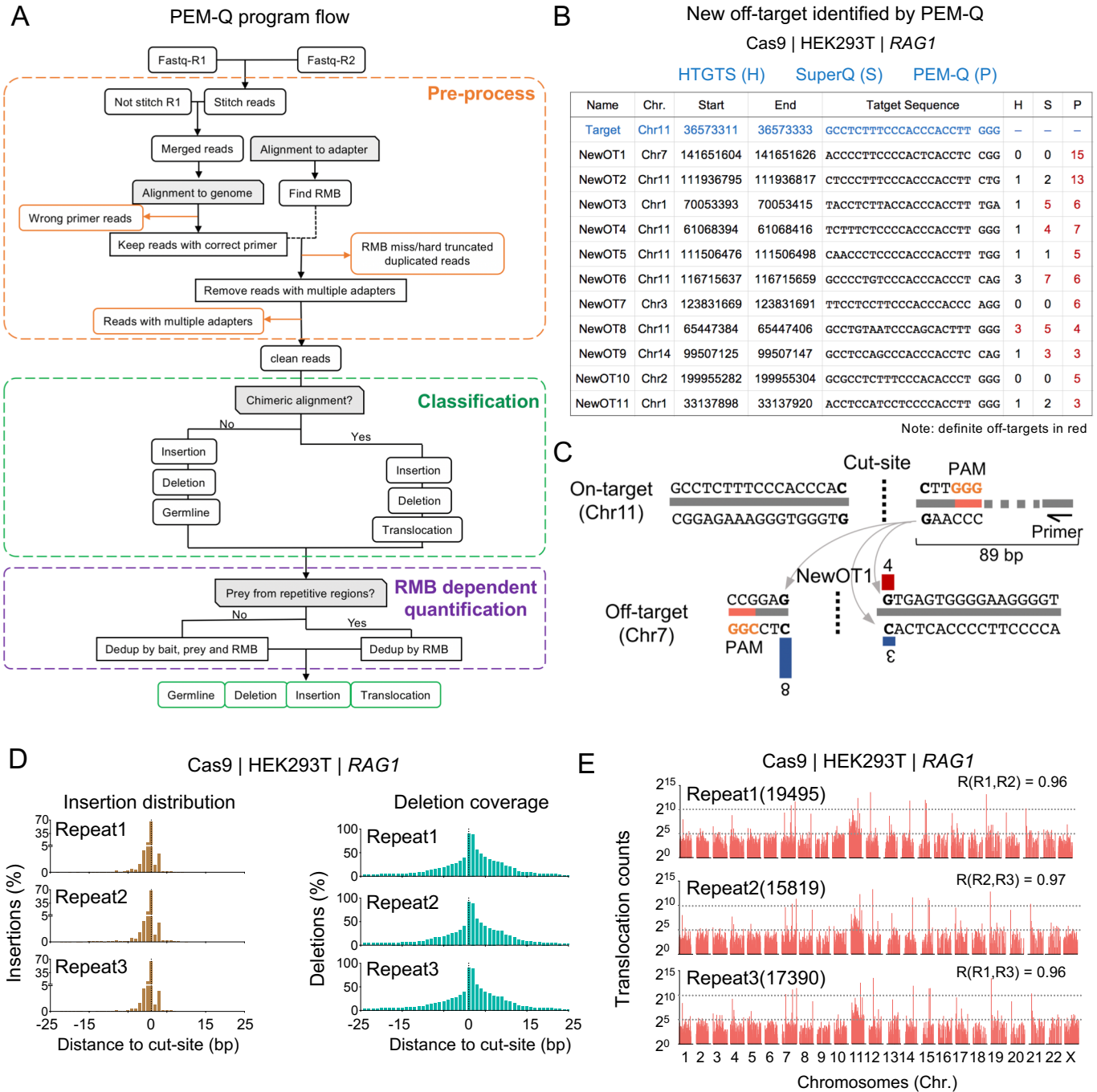
Supplementary Table S2 sgRNA sequences for PEM-seq

Supplementary Table S3 Detected 45 plasmid integrations sequences at the *c-Myc* locus in CH12F3 cells.

Supplementary Table S4 Comparison of PEM-Q with other *in silico* methods analyzing CRISPR-Cas9 editing repair outcomes

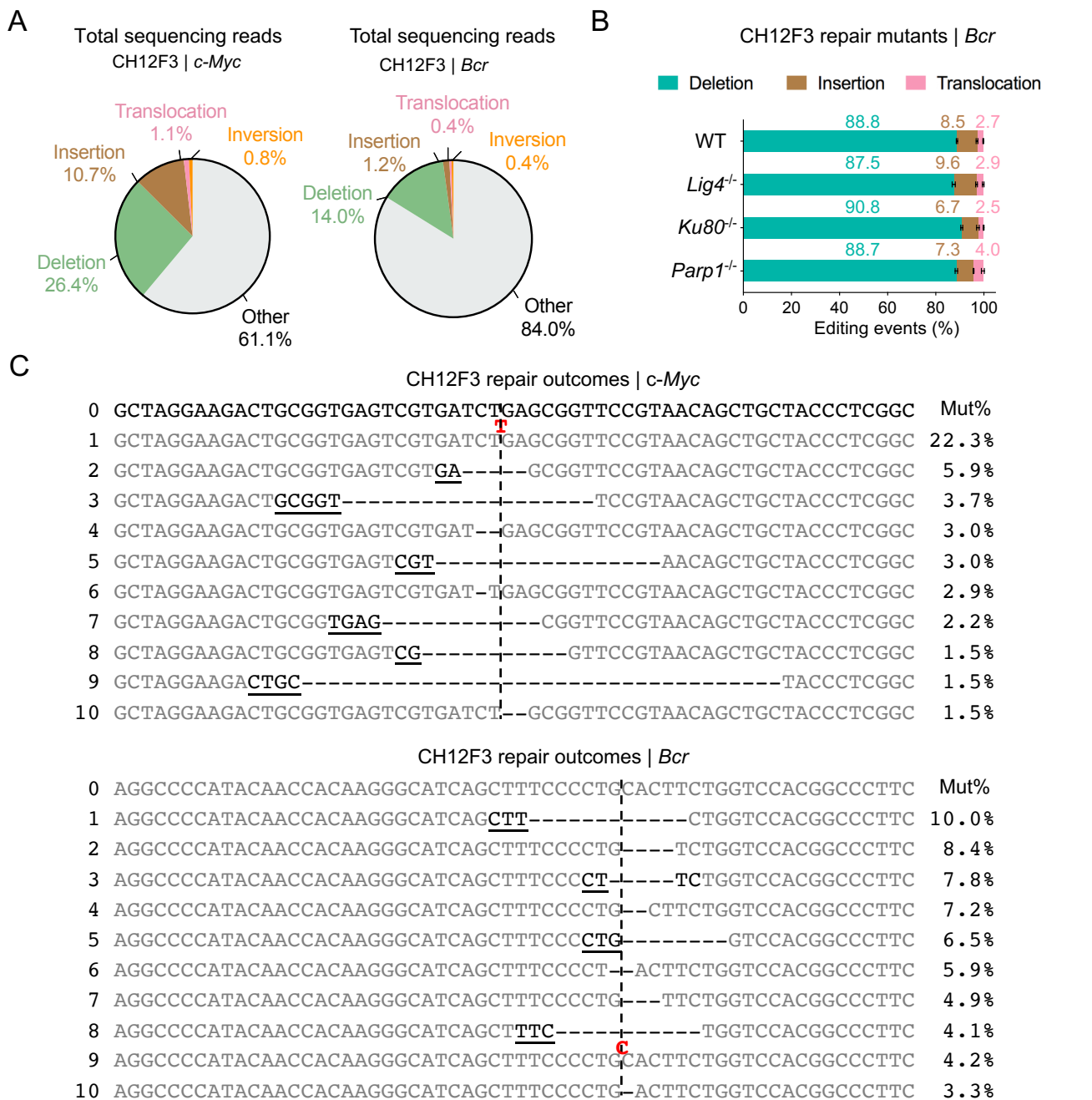
Supplementary Table S5 Editing events and editing efficiency detected by PEM-Q

Supplementary Figure S1



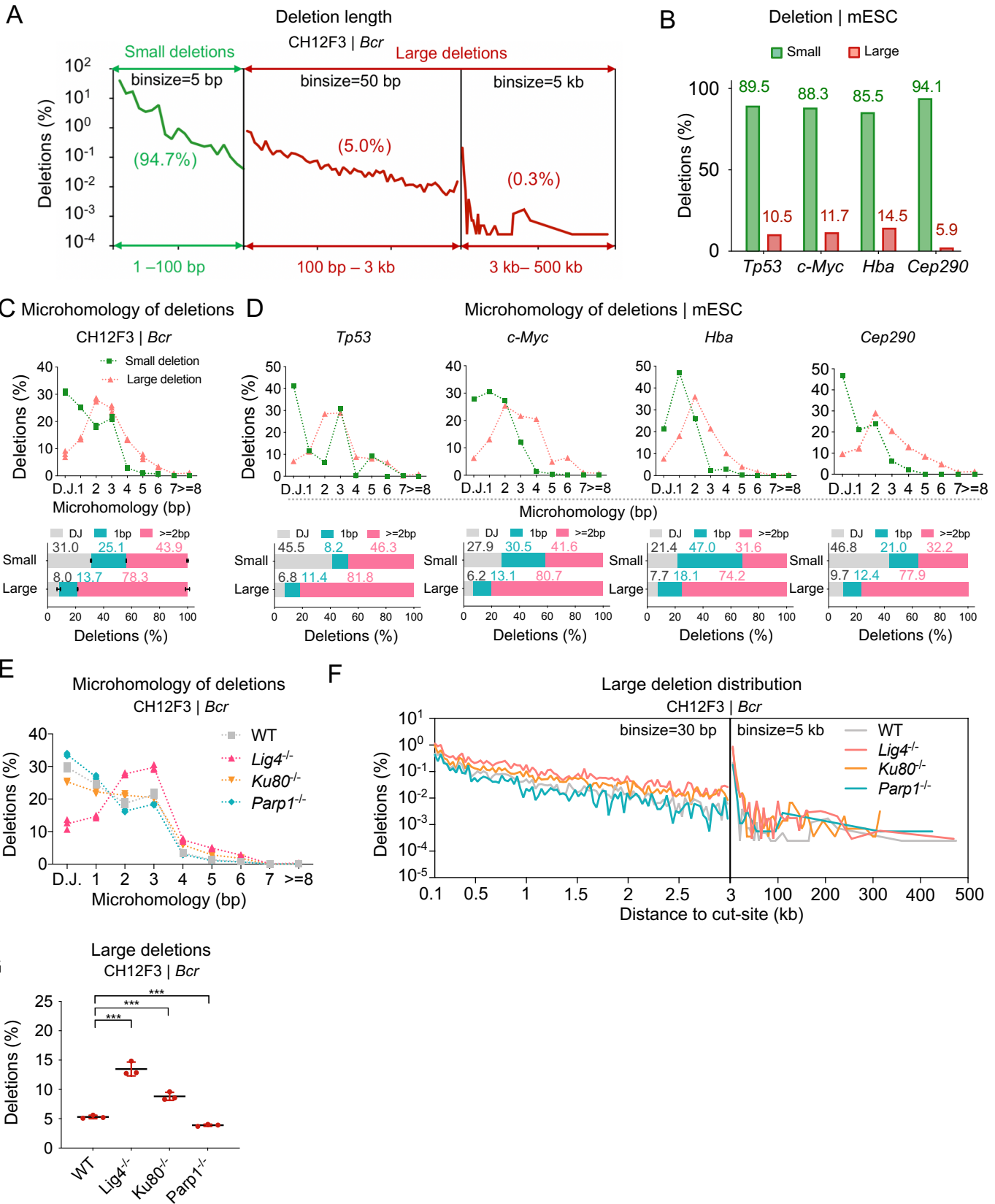
Supplementary Figure S1. Analyzing repair outcomes by PEM-Q. (A) PEM-Q pipeline flow diagram. (B) 11 New off-targets (NewOT) identified by PEM-Q at the *RAG1* locus in HEK293T cells. Translocation junctions on NewOTs detected by HTGTS (H), SuperQ (S) and PEM-Q (P) are listed in the table. The target locus is highlighted in blue. Off-targets consistent with the frequently-used criteria (see methods for details) are highlighted in red. (C) Schematic diagram showing the distribution of translocation junctions at NewOT1. PAM sequences are highlighted in orange and bases at cut-site are in bold. Primer is 89 bp downstream from cut-site as shown by a black arrow. Positive-strand translocation junctions are in red, while negative-strand in blue. (D) Distribution of insertions among total insertions and coverage of deletions among total deletions for three repeat PEM-seq libraries at the *RAG1* locus in HEK293T cells. (E) Genome-wide translocations of three repeat PEM-seq libraries at *RAG1* locus in HEK293T cells. Translocation junctions are plotted with 2-Mb intervals on a log scale. Total numbers (n) of translocations are shown. Pearson correlation (R) between each other repeat was 0.96-0.97.

Supplementary Figure S2



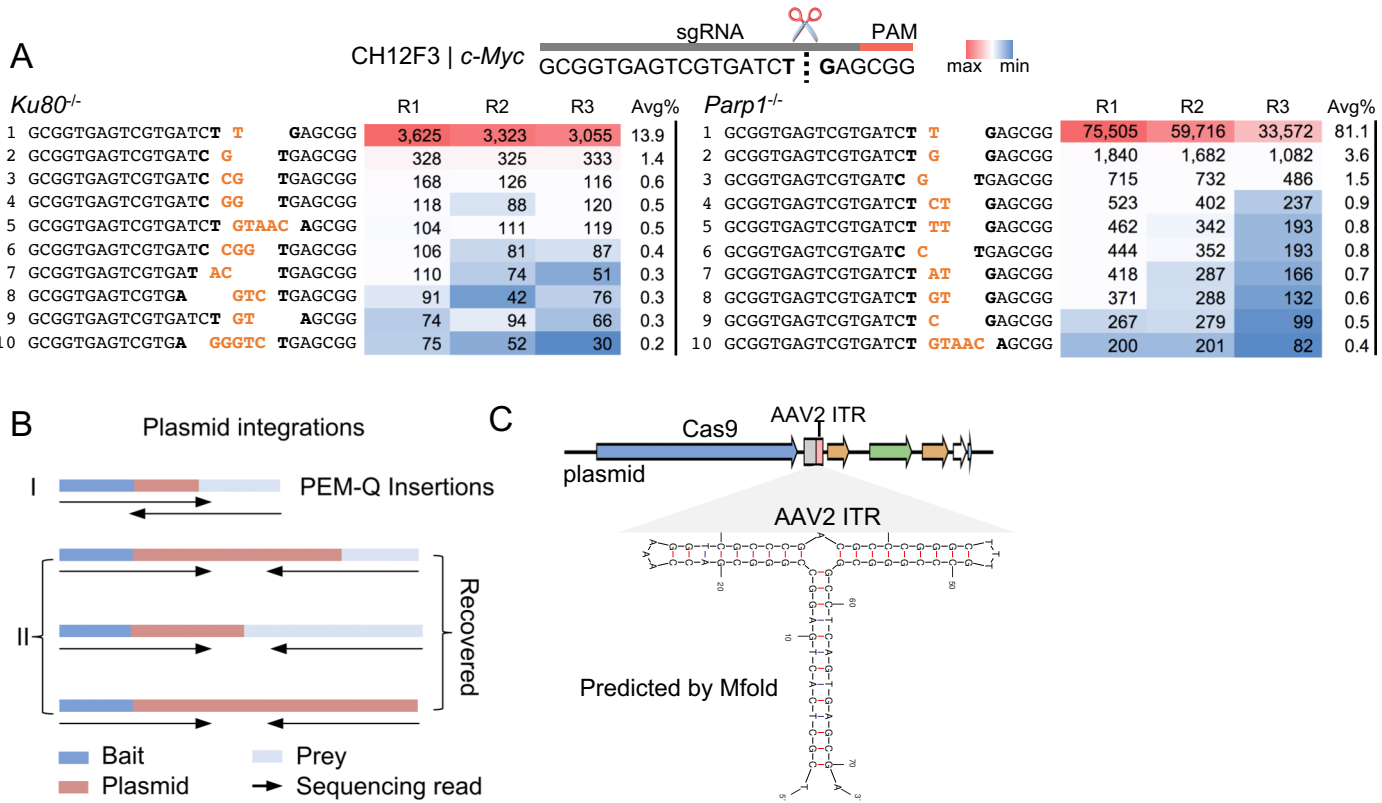
Supplementary Figure S2. Repair outcomes at *c-Myc* and *Bcr* loci. (A) Pie chart of total sequencing reads at the *c-Myc* (left), *Bcr* (right) loci in CH12F3 cells. Total sequencing reads include deletion, insertion, translocation, inversion, and other reads (including germline sequences). The average percentages of three repeats are shown. (B) Bar charts showing percentages of deletion (cyan), insertion (brown) and translocation (pink), the *Bcr* locus in CH12F3 cells with indicated backgrounds. Error bars, mean \pm SD. (C) Top 10 most frequent mutation sequences at the *c-Myc* (top) and *Bcr* (bottom) loci in CH12F3 cells. #0 in the top indicate the target sequences. The horizontal dashed line represent deleted bases and the vertical dashed line marks the cut-site. Bases with underline represents the microhomology, the base of insertion is shown in red. The average mutation frequency (Mut%) of three repeats among all editing events are listed on the right side.

Supplementary Figure S3



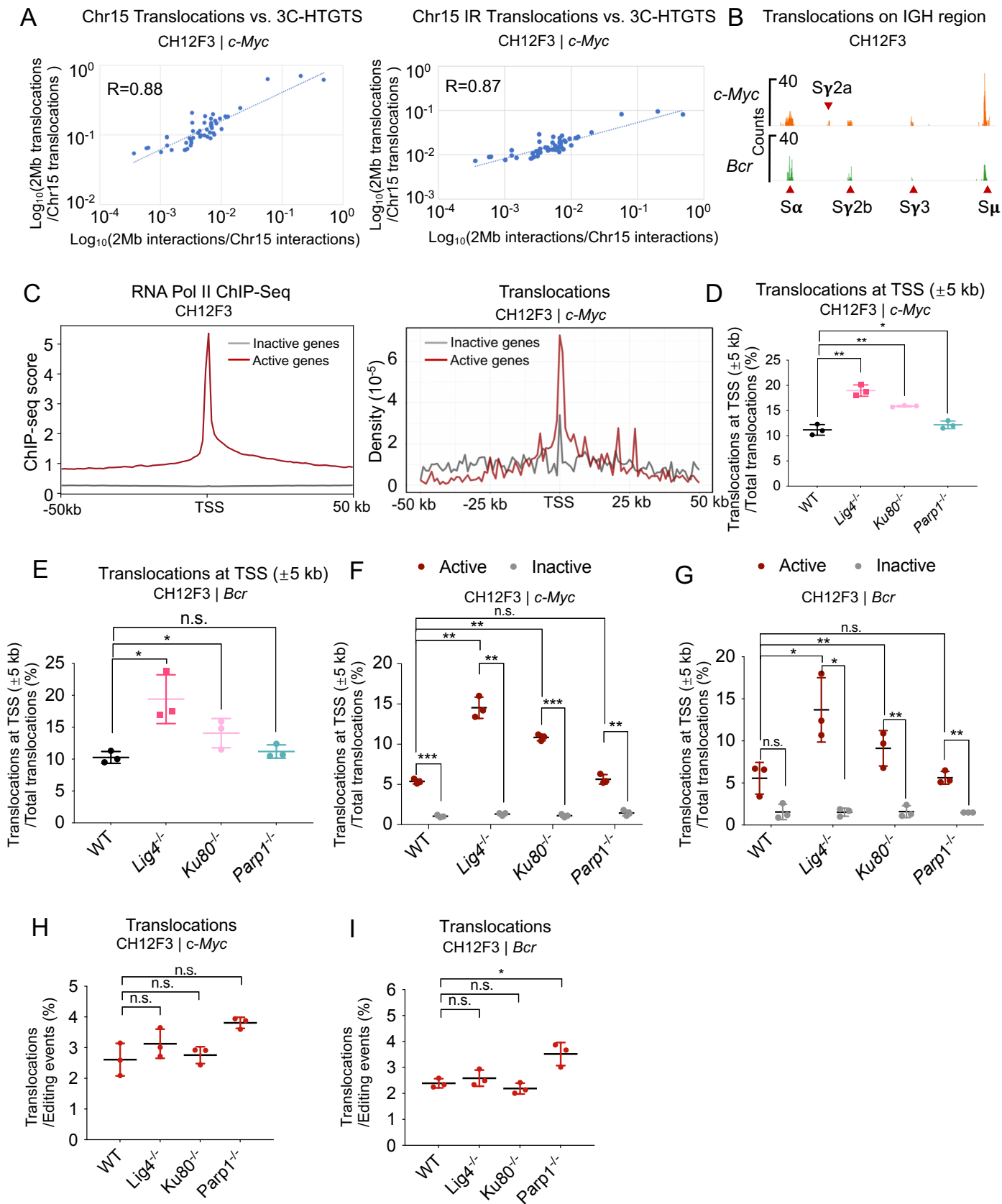
Supplementary Figure S3. Microhomologies are commonly used in large deletions. (A) The distribution pattern of deletions at the *Bcr* locus in CH12F3 cells. Total junctions of three repeats are plotted on a log scale. Percentages of deletions within each region are shown in the brackets. Please note that 5 bp, 50 bp, and 5 kb bin-sizes are used for the three regions, respectively. (B) Percentages of small and large deletions among the total deletions at indicated loci in mESCs. (C and D) Line plot (top) and bar chart (bottom) of microhomologies with indicated length in small or large deletions at the *Bcr* locus in CH12F3 cells (C) at indicated loci in mESCs (D). D.J., direct joining. Error bars, mean \pm SD. (E) Line plot of microhomologies with indicated length in total deletions at the *Bcr* locus in CH12F3 cells with indicated backgrounds. (F) The distribution patterns of large deletions at the *Bcr* locus in CH12F3 cells with indicated backgrounds. Please note that 30 bp and 5 kb bin-sizes are used for two regions, respectively. (G) Percentages of large deletions among total deletions at the *Bcr* locus in CH12F3 cells with indicated backgrounds. One-tailed t-test, ***, $p < 0.0005$. Error bars, mean \pm SD.

Supplementary Figure S4



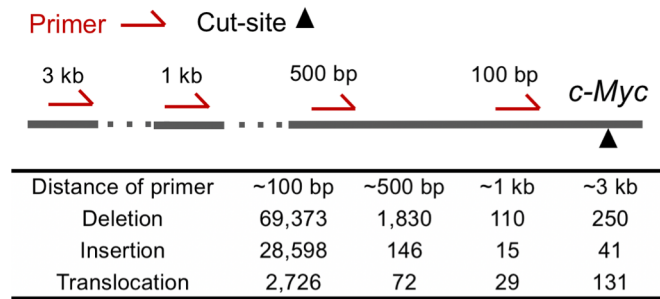
Supplementary Figure S4. Insertions at the target sites. (A) Top 10 most frequent insertions at the *c-Myc* locus in *Ku80*^{-/-} and *Parp1*^{-/-} CH12F3 cells. The target site is shown on the top. Bases at cleavage site are in bold and inserted bases are in orange. The numbers from three repeats (R1, R2, R3) and average percentages (Avg%) of each type insertions among the total insertion are listed in the table, filled with gradient color from the maximum (red) to the minimum (blue) frequencies. (B) Schematics showing the two types of plasmid integrations. The sequencing reads cover the entire inserted fragments in Structure I, while only partially cover the inserted fragment in Structure II. (C) Predicted secondary structure of AAV2 ITR from plasmid predicted by Mfold (<http://www.unafold.org/mfold/applications/dna-folding-form.php>).

Supplementary Figure S5



Supplementary Figure S5. Translocations enriched at recurrent DSBs in the genome. (A) Correlation of junctions on Chr15 between 3C-HTGTS and PEM-seq without (left) or with (right) IR. Each dot represents translocation frequency versus 3C-HTGTS interaction frequency in each 2 Mb bin along Chr15. Pearson correlation (R) is shown. (B) The distribution of translocation junctions in the S region. The positions of junction-enriched S regions are indicated by red triangles. (C) The distribution of RNA Pol II ChIP-seq signals (left) and translocation junctions (right) around TSSs (± 50 kb, 1kb bins). Genes were sorted by means of RNA Pol II signal at TSS (± 5 kb), top 25% genes were grouped into active genes, while bottom 25% genes were grouped into inactive genes. Translocations around active genes' TSSs and inactive genes' TSSs are plotted in the corresponding color. (2 kb bins). For note, translocations distributed on the *IgH* region were removed. (D and E) The percentages of translocation junctions around TSSs (± 5 kb) at the *c-Myc* (D) and *Bcr* (E) loci in CH12F3 cells. (F and G) The percentages of translocation junctions around active and inactive genes' TSSs (± 5 kb) at the *c-Myc* (F) and *Bcr* (G) loci in CH12F3 cells. (H and I) Percentages of translocations among editing events at the *c-Myc* (H) and *Bcr* (I) loci in CH12F3 cells. One-tailed t-test, *, $p < 0.05$, **, $p < 0.005$, ***, $p < 0.0005$; n.s., not significant. Error bars, mean \pm SD.

Supplementary Figure S6



Supplementary Figure S6. Editing events detected by primers with different distances from the cut-site. Numbers of deletions, insertions and translocations captured by primers with distances of 100 bp, 500 bp, 1 kb, and 3 kb are listed in the table. For note, the numbers were normalized by total deduped sequencing reads. The black triangle indicates the cut-site. The red arrows represents primers.

Supplementary Table S1 Primer sequences for PEM-seq and 3C-HTGTS

Cell line	Locus	Bio primer	Red primer
mESCs	<i>Tp53</i>	GAGGCTATCCGGAGCTAAGAG TCGCTC	GTCGGGCAAGTCTCGCTGAG
	<i>c-Myc</i>	CTGACAGCCTGGGACCGACAC GGAG	GCGGCGATCGCAACCCGTCC
	<i>Hba</i>	CCCTAGGAAGGGCTTGGGGGT CC	AGAGGCATCAGGGTGTCCAC
	<i>Cep290</i>	GTATTCTAGATTAGCCAATAAT CAGATGG	GACAGTAGCATACCTGGTAATG
CH12F3	<i>c-Myc</i> (~100 bp)	CGGCACTAGGACTTGATGTTG GGCTAGCGC	GGAAACCAGAGGGAATCCTC
	<i>c-Myc</i> (~500 bp)	AGGTTACTATGGGCTGACGCT GACCCGGCC	GATATGTGTCCTTTGAGGGGTC AAAC
	<i>c-Myc</i> (~1 kb)	GCTCTGGAGTGAGAGGGGCTT TGCTCCG	GCGACTGACCCAACATCAGCGG CCGC
	<i>c-Myc</i> (~3 kb)	GCCTTGGGGCGAGGAGTCCG GAATAAG	GTGTAGGATAAGCAAATCCCGA GGG
	<i>Bcr</i>	CACTGACCACAGCTGTTCTTC CCAGGGAG	CAGGTTGCCCTTCTGAGGCTAC

Supplementary Table S2 sgRNA sequences for PEM-seq

Cell line	Locus	sgRNA Sequence
mESCs	<i>Tp53</i>	CGGAAGAGGAAAGCGGACTC CGG
	<i>c-Myc</i>	AGCCTGACCCCCGCGGCACT AGG
	<i>Hba</i>	TCCAGAGAGGCATGCACCGC GGG
	<i>Cep290</i>	ATTAGTGTCAAGTACCCCAT AGG
CH12F3	<i>c-Myc</i>	GCGGTGAGTCGTGATCTGAG CGG
	<i>Bcr</i>	GCCGTGGACCAGAAGTGCAG GGG

Supplementary Table S3 Detected 45 plasmid integrations sequences at the *c-Myc* locus in CH12F3 cells.

No.	Bait	Prey	Aligned Vector Sequence
1	61986728 (Chr15)	61986731 (Chr15)	TGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTTAGGT AACAACTTACGTTACGTAACAGCTAGTTACGTAACGTA
2	61986729 (Chr15)	61986731 (Chr15)	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTTACTTA ACAGCTTACGTTACGTAACAGCTAGTTACGTAACGTA
3	61986729 (Chr15)	61986731 (Chr15)	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTTACGTA TCAGCTTACGTTACGTAACAGCTAGTTACGTAACGTA
4	61986717 (Chr15)	61986732 (Chr15)	TCTCCCCATCTCCCCCCTCCCCACCCCAATTTGTATTATTTA TTTTTAATTATTTGTGCAGCGATGGGGGCGGGGGGGGGG
5	61986705 (Chr15)	113260965 (Chr12)	TGGCGAGGCGGCGGCGGCGGCGGCCCTATAAA
6	61986724 (Chr15)	61986730 (Chr15)	TGATCACCGACGAGTACAAGGTGCCAGCAAGAAATCAAGGTG CTGGGCAACACCGACCGGCACAGC
7	61986724 (Chr15)	61986730 (Chr15)	TGATCACCGACGAGTACAAGGTGCCAGCAAGAAATCAAGGTG CTGGGCAACACCGACCGGCACAGC
8	61986689 (Chr15)	61986733 (Chr15)	AATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACA GAGAGATGATCGAGGAACGGCTGTAATGCTTGTGTTGCTATTAA AGCATT
9	61986714 (Chr15)	61986733 (Chr15)	AATTATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGG ACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACA GAGAGATGATCGAGGAACGGCTGAAA
10	61986725 (Chr15)	61986728 (Chr15)	GGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGA ACCAGACCACCCAGAAGGGACAGA
11	61986727 (Chr15)	61986728 (Chr15)	GCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACC AGACCACCCAGAAGGGACAGA
12	61986723 (Chr15)	61986737 (Chr15)	ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAG GGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGG GCATCAAAG
13	61986723 (Chr15)	61986737 (Chr15)	ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAG GGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGG GGATCAAAG
14	61986723 (Chr15)	61986737 (Chr15)	ACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAGAAG GGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGG GCATCAAAG
15	61986724 (Chr15)	61986727 (Chr15)	AAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCC CCAAGTGAATA
16	61986728 (Chr15)	61986731 (Chr15)	CAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGAGAGGTTTTCA CCGTCATCACCGATGAGT
17	61986728 (Chr15)	61986726 (Chr15)	CAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGAGAGGTTTTCA CCGTCATCACCGA
18	61986728 (Chr15)	61986726 (Chr15)	CAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTGAGAGGTTTTCA CCGTCATCACCGA
19	61986726 (Chr15)	3142041 (Chr11)	TTGGTTGAGTACTACCAGTACAGAAAAGCATCTTACGGATGGC ATGACAGTAAGAGAATTATGCAGTGCTG
20	61986726 (Chr15)	61986726 (Chr15)	CGCTCGGCCCTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA GCCGG
21	61986726 (Chr15)	61986726 (Chr15)	CGCTCGGCCCTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA GCCGG
22	61986726 (Chr15)	61986726 (Chr15)	CGCTCGGCCCTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA GCCGG
23	61986726 (Chr15)	61986726 (Chr15)	CGCTCGGCCCTCCGGCTGGCTGGTTTATTGCTGATAAATCTGCA GCAGG
24	61986726 (Chr15)	61986729 (Chr15)	AGTAACAGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGG CGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGAT

25	61986721 (Chr15)	61986730 (Chr15)	GAAAAAAAAAGTAAAGGTGAGTTACCATCTTCGGCAACATCGTGG ACGAGGTGGCCTACAAAGGTGAGTT
26	61986721 (Chr15)	61986730 (Chr15)	TAAAAAAAAAGTAAAGGTGAGTTACCATCTTCGGCAACATCGTGG ACGAGGTGGCCTACAAAGGTGAGTT
27	61986721 (Chr15)	61986730 (Chr15)	TAAAAAACAGTAAAGGTGAGTTACCATCTTCGGCAACATCGTGG ACGAGGTGGCCTACAAAGGTGAGTT
28	61986717 (Chr15)	61986730 (Chr15)	TAAAAAAAAAGTAAAGGTGAGTTACCATCTTCGGCAACATCGTGG ACGAGGTGGCCTACAAAGGTGAGTTACGC
29	61986721 (Chr15)	61986730 (Chr15)	TAAAAAAAAAGTAAAGGTGAGTTACCATCTTCGGCAACATCGTGG ACGAGGTGGCCTACAAAGGTGAGTT
30	61986726 (Chr15)	61986726 (Chr15)	TCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGG
31	61986726 (Chr15)	61986726 (Chr15)	TCCAGCTGGTGCAGACCTACAACCAGCTGTTTCGAGG
32	61986717 (Chr15)	61986729 (Chr15)	GCGTGGGAGGCCATCCTCCGAGGGGATGTACCCCGAGGACGGCG CCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACG
33	61986722 (Chr15)	61986729 (Chr15)	GGAGGCCATCCTCCGAGGGGATGTACCCCGAGGACGGCGCCCTG AAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACG
34	61986722 (Chr15)	61986736 (Chr15)	GGAGGCCATCCTCCGAGGGGATGTACCCCGAGGACGGCGCCCTG AAGGGCGAGATCAAGCAGAGGCTGCCGCTGAAGCACCTGTTATT
35	61986662 (Chr15)	61986772 (Chr15)	ATTCGGTCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTA ACGCGAATTTT
36	61986720 (Chr15)	61986728 (Chr15)	CCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTT TTGCTCACATATAAGCGGTTACATATT
37	61986720 (Chr15)	61986728 (Chr15)	CCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTT TTGCTCACATATAAGCGGTTACATATT
38	61986724 (Chr15)	61986942 (Chr15)	GTTTTTCGAGTTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTT TTAGT
39	61986725 (Chr15)	61986726 (Chr15)	TGCCGACAAGAAGTACAGCATCGGCCTGGACA
40	61986691 (Chr15)	61986728 (Chr15)	GACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGTACT
41	61986721 (Chr15)	61986728 (Chr15)	GGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTG CCCTCCAAATATGTGAACTTCCTGTACTGTTTCATCCAGCTTGTT
42	61986729 (Chr15)	61986730 (Chr15)	GGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCG CTAGCGC
43	61986682 (Chr15)	62003290 (Chr15)	AAGTAGGAATCTGTTGCCGGATCAAGAGCTACCAACTCTTTTTCC GAAGGTAAGTGGCTTCAGCAGAG
44	61986717 (Chr15)	61986729 (Chr15)	AGAACTCTGTAGCACCGCCTACATACCTCG
45	61986717 (Chr15)	61986729 (Chr15)	AGAACTCTGTAGCACCGCCTACATACCTCG

Supplementary Table S4 Comparison of PEM-Q with other in silico methods analyzing CRISPR-Cas editing repair outcomes

In silico program	Sequencing method	Eliminate PCR bias	Quantification	Off target analysis	Indel analysis	Large deletion detection	Translocation detection	Microhomology detection	Vector integration analysis	Editing efficiency analysis	Ref.
TIDE	Sanger sequencing	–	Indel alignment	–	Indel lengths and counts	Limited to templates' length	–	–	–	✓	47
CRISPResso	Targeted sequencing	–		–			–	–	–	✓	31
ampliCan	Targeted sequencing	–		–			–	–	–	✓	49
CRIS.py	Targeted sequencing	–		–			–	–	–	✓	48
HTGTS	HTGTS	–	Translocation junction	✓	–	Any length	✓	✓	–	–	42
	LAM-HTGTS	✓		✓			✓	–	–	33	
	PEM-seq	✓		✓			✓	–	–	28	
CAST-seq	CAST-seq	✓	Linker with translocation junction	✓	Use CRISPResso		✓	✓	–	Use CRISPResso	27
SuperQ	PEM-seq	✓	Barcodes with indel alignment or translocation junction	✓	Indel counts		✓	✓	–	✓	28
PEM-Q	PEM-seq	✓		✓	Indel lengths, sequences and counts		✓	Lengths and sequences	Lengths and sequences	✓	This study

Supplementary Table S5 Editing events and editing efficiency by PEM-seq

Sample	Cell line	Deletion	Insertion	Translocation	Editing Efficiency (%)
Myc-WT-1	CH12F3	96104	6151	3504	23.67
Myc-WT-2	CH12F3	78895	4969	3248	25.00
Myc-WT-3	CH12F3	102150	7408	3666	25.10
Myc-Lig4 -/- -1	CH12F3	80179	8588	4299	25.30
Myc-Lig4 -/- -2	CH12F3	70788	6802	3241	27.20
Myc-Lig4 -/- -3	CH12F3	88315	9512	4992	25.10
Myc-Ku80 -/- -1	CH12F3	68171	4341	3067	19.89
Myc-Ku80 -/- -2	CH12F3	71441	5035	2927	17.94
Myc-Ku80 -/- -3	CH12F3	58673	3827	2298	19.45
Myc-Parp1 -/- -1	CH12F3	47086	3029	1939	13.16
Myc-Parp1 -/- -2	CH12F3	45101	2795	2102	12.04
Myc-Parp1 -/- -3	CH12F3	61609	3913	2854	12.76
Bcr-WT-1	CH12F3	84458	7722	2848	12.89
Bcr-WT-2	CH12F3	125864	12549	3717	17.90
Bcr-WT-3	CH12F3	130172	12007	4024	16.74
Bcr-Lig4/- -1	CH12F3	58747	6564	2305	12.77
Bcr-Lig4/- -2	CH12F3	164842	16899	5141	28.56
Bcr-Lig4/- -3	CH12F3	52361	5756	1769	15.82
Bcr-Ku80/- -1	CH12F3	73445	6000	1863	11.39
Bcr-Ku80/- -2	CH12F3	116415	8271	3082	17.51
Bcr-Ku80/- -3	CH12F3	44326	3105	1280	9.60
Bcr-Parp1/- -1	CH12F3	32888	2692	1696	7.63
Bcr-Parp1/- -2	CH12F3	63068	5117	2520	10.80
Bcr-Parp1/- -3	CH12F3	58344	4532	3122	10.36
MYC1-1	HEK293T	91661	22162	4255	42.70
MYC1-2	HEK293T	100493	24400	4759	43.02
MYC1-3	HEK293T	145408	33929	7258	34.05
MYC2-1	HEK293T	58215	26060	2809	21.55
MYC2-2	HEK293T	94922	44884	4167	18.06
MYC2-3	HEK293T	45435	20280	2046	23.86
Dnmt1-1	HEK293T	98723	22235	4417	44.03
Dnmt1-2	HEK293T	148520	33713	8582	39.55
Dnmt1-3	HEK293T	162452	36548	7112	44.81
Dnmt2-1	HEK293T	127000	53751	6282	46.83
Dnmt2-2	HEK293T	141535	57826	6576	40.89
Dnmt2-3	HEK293T	145725	57110	5378	39.82
P53	mESC	49110	5794	1213	6.21
c-Myc	mESC	49232	9739	1896	7.26
Hba	mESC	22285	5607	948	17.32
Cep290	mESC	19882	1704	862	4.95
SpCas9_RAG1A-1	HEK293T	218058	41004	19494	33.54
SpCas9_RAG1A-2	HEK293T	181393	32731	15818	29.48
SpCas9_RAG1A-3	HEK293T	177486	30246	17389	31.93
eCas9_RAG1A-1	HEK293T	123012	15987	5451	36.09
eCas9_RAG1A-2	HEK293T	113242	14552	5258	29.20
eCas9_RAG1A-3	HEK293T	69808	8651	3021	30.26
FeCas9_RAG1A-1	HEK293T	40072	6148	2845	39.90
FeCas9_RAG1A-2	HEK293T	61231	11964	4586	50.02
FeCas9_RAG1A-3	HEK293T	31395	4995	2618	36.98
HF1_RAG1A-1	HEK293T	95531	13704	4647	32.06
HF1_RAG1A-2	HEK293T	142130	21126	6710	28.50
HF1_RAG1A-3	HEK293T	111758	16438	4792	35.24