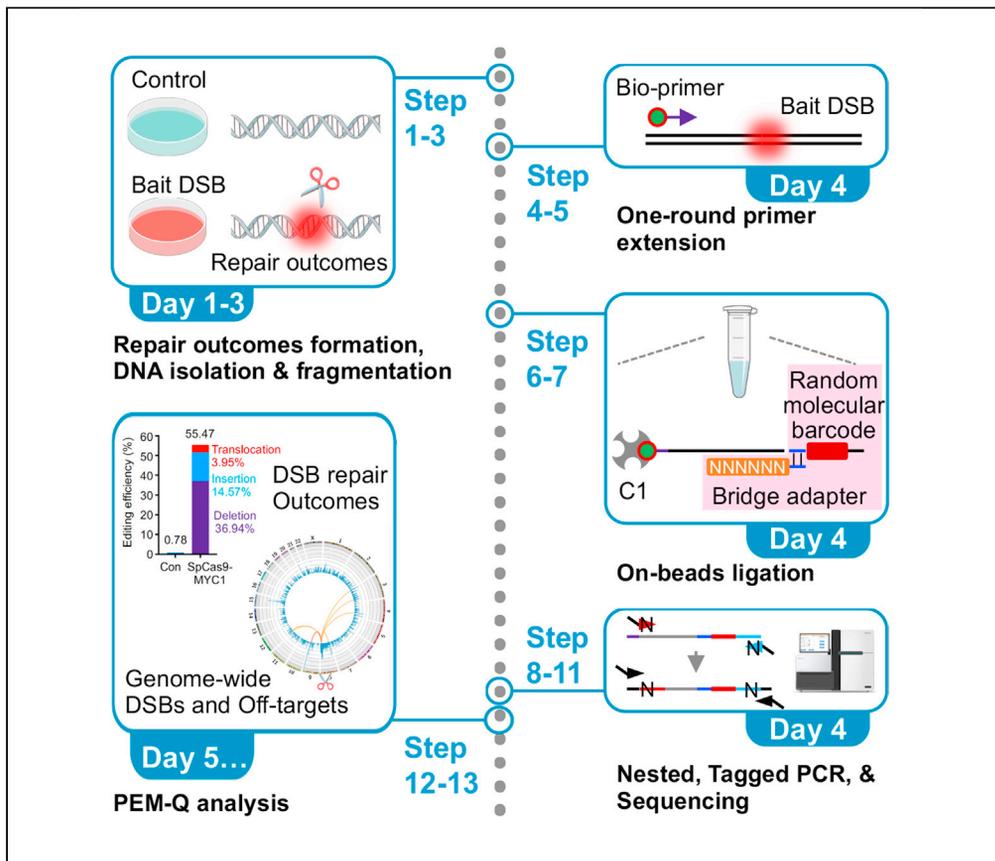


Protocol

PEM-seq comprehensively quantifies DNA repair outcomes during gene-editing and DSB repair



The repair products of double-stranded DNA breaks (DSBs) are crucial for investigating the mechanism underlying DNA damage repair as well as evaluating the safety and efficiency of gene-editing; however, a comprehensively quantitative assay remains to be established. Here, we describe the step-by-step instructions of the primer extension-mediated sequencing (PEM-seq), followed by the framework of data processing and statistical analysis. PEM-seq presents a full spectrum of repair outcomes for both genome-editing-induced and endogenous DSBs in mouse and human cells.

Yang Liu, Jianhang Yin, Tingting Gan, Mengzhu Liu, Changchang Xin, Weiwei Zhang, Jiazhi Hu

liu.y@pku.edu.cn (Y.L.)
hujz@pku.edu.cn (J.H.)

Highlights

PEM-seq comprehensively quantifies DSB repair outcomes

PEM-seq evaluates the efficiency and safety of genome-editing tools

PEM-seq studies the impact of DNA damage response pathways on DSB repair

PEM-seq identifies endogenous DNA damage sites and DNA fragment integrations

Liu et al., STAR Protocols 3, 101088
March 18, 2022 © 2021 The Author(s).
<https://doi.org/10.1016/j.xpro.2021.101088>



Protocol

PEM-seq comprehensively quantifies DNA repair outcomes during gene-editing and DSB repair

Yang Liu,^{1,2,3,*} Jianhang Yin,^{1,2} Tingting Gan,^{1,2} Mengzhu Liu,¹ Changchang Xin,¹ Weiwei Zhang,¹ and Jiazhi Hu^{1,4,*}

¹The MOE Key Laboratory of Cell Proliferation and Differentiation, School of Life Sciences, Center for Life Sciences, Genome Editing Research Center, Peking University, Beijing 100871, China

²These authors contributed equally

³Technical contact

⁴Lead contact

*Correspondence: liu.y@pku.edu.cn (Y.L.), hujz@pku.edu.cn (J.H.)
<https://doi.org/10.1016/j.xpro.2021.101088>

SUMMARY

The repair products of double-stranded DNA breaks (DSBs) are crucial for investigating the mechanism underlying DNA damage repair as well as evaluating the safety and efficiency of gene-editing; however, a comprehensively quantitative assay remains to be established. Here, we describe the step-by-step instructions of the primer extension-mediated sequencing (PEM-seq), followed by the framework of data processing and statistical analysis. PEM-seq presents a full spectrum of repair outcomes for both genome-editing-induced and endogenous DSBs in mouse and human cells.

For complete details on the use and execution of this profile, please refer to Gan et al. (2021), Yin et al. (2019), Liu et al. (2021a), and Zhang et al. (2021).

BEFORE YOU BEGIN

Double-stranded DNA breaks (DSBs) are intrinsic to DNA metabolism processes, including DNA replication (Liu et al., 2021b; Tubbs et al., 2018), transcription (Liu et al., 2021a; Meng et al., 2014), DNA damage repair (Tubbs and Nussenzweig, 2017), V(D)J recombination (Hu et al., 2015), and antibody class switch recombination (Dong et al., 2015). Besides, the emerging nuclease-mediated genome editing also induces DSBs, including FokI domain-containing nucleases, transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced palindromic repeats (CRISPR)-Cas (Li et al., 2020). DSBs are sealed by two main DSB repair pathways in mammalian cells, homologous recombination (HR) and non-homologous end joining (NHEJ) (Figure 1A). Although alternative end joining (a-EJ), including microhomology-mediated end joining (MMEJ), is discovered when NHEJ is deficient, it also participates in DSB repair in the presence of NHEJ and HR (Figure 1A). These DSB repair pathways are triggered under different scenarios and generate diverse repair outcomes that are mirrors of both DSBs and involved repair process(es) (Liu et al., 2021a). For instance, the repair of a single DSB would lead to perfect re-joinings, small indels, microhomology-mediated deletions, and large deletions (Figure 1A). While the repair of multiple DSBs induce not only the above mentioned products but also large deletions and intra- or inter-chromosomal translocations (Figure 1A). Regarding genome editing, large deletions and translocations usually are unwanted editing products. With these regards, a quantitative assay to profile a full spectrum of repair outcomes is urgently demanded to study DSB repair pathway(s) and evaluate the efficiency and safety of gene-editing tools (Saha et al., 2021).

To capture and quantify the DSB repair outcomes, we developed the primer extension-mediated sequencing (PEM-seq) by using primer-extension amplification and introducing random molecular



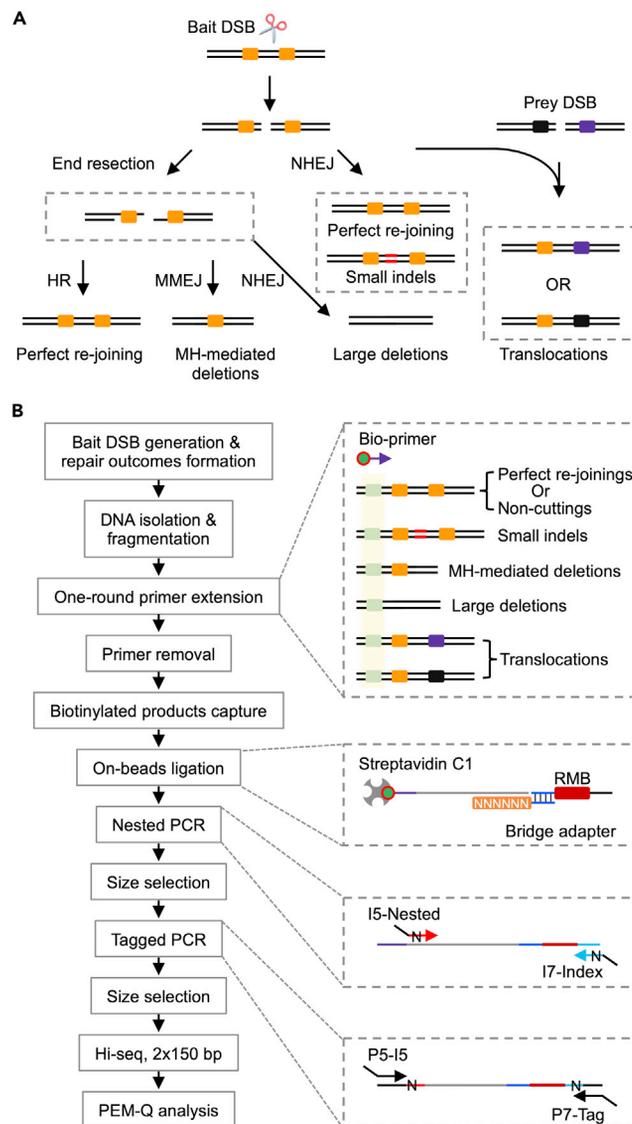


Figure 1. The design and experimental procedure of PEM-seq

(A) DSB repair pathways and repair outcomes. A DSB formed at the designed site, termed bait DSB, is mainly repaired by two types of repair processes, which are characterized by the end resection or not. With ends resection, bait DSB will be repaired by 1) the homologous recombination (HR) to form an error-free product that is identical to the reference sequence of bait DSB, termed perfect re-joining, 2) microhomology-mediated end joining (MMEJ) inducing microhomology (MH)-mediated deletions, or 3) non-homologous end joining (NHEJ) producing large deletions if 3'-overhang at the resected ends are removed and DNA ends are ligated. Without end resection, bait DSB will be re-joined by NHEJ to form perfect re-joining and small insertion or deletions (indels). When another DSB (prey DSB) is formed simultaneously with the bait DSB, they may join together and form intra- or inter-chromosomal translocations. The orange boxes are microhomology around the bait DSB.

(B) Procedures for the preparation of PEM-seq library and following analysis. With one round primer extension of a biotinylated primer and the following on-beads ligation with the barcoded bridge adapter, PEM-seq captures and quantifies multiple types of DSB repair outcomes at the bait DSB and genome-wide translocations. The major steps of PEM-seq are shown on the left; the right panel highlights indicated operations. Green boxes, with yellow-shadow background, show the bio-primer targeted regions. RMB, random molecular barcode.

barcode (RMB) (Liu et al., 2021a; Yin et al., 2019). As shown in Figure 1B, the procedure starts with the generation of DSB at the target site, followed by a limited duration to allow the formation of DSB repair products (Step 1). Extracted genomic DNA (Step 2) is sheared to 300–700 bp

fragments by sonication (Step 3). All products containing the complementary sequence of the biotinylated primer are then amplified by a one-round primer extension (Step 4). After removal of exceeded biotinylated primer (Step 5), biotin-labeled ssDNA is enriched by streptavidin C1 beads (Step 6) and ligated with a bridge adapter containing a 14-bp RMB (Step 7). Adapter-ligated ssDNA fragments are subjected to nested PCR (Step 8), size selection (Step 9), amplification with indexed Illumina primers (Step 10), size selection again (Step 11), and finally are sequenced by Hi-seq with 2 × 150 bp reads (Step 12). PEM-Q is applied to data processing and statistical analysis (Step 13) (Liu et al., 2021a).

PEM-seq can be used to determine the editing efficiency, off-target sites, and unwanted products of genome editing (Yin et al., 2019; Zhang et al., 2021). In addition, PEM-seq can also be applied to interrogate the DSB level, endogenous DSB hotspots, DSB repair pathway choice, and the underlying molecular mechanism (Liu et al., 2021a). Though couples of high-throughput sequencing approaches have been developed to evaluate the efficiency or off-target activity of gene-editing tools, including LAM-HTGTS, GUIDE-seq, DISCOVER-seq, CIRCLE-seq, SITE-seq, Digenome-seq, BLESS, etc., which have been well-reviewed elsewhere (Hu et al., 2016; Kim et al., 2019). However, PEM-seq is the only one that quantitatively profiles a full spectrum of repair outcomes. Here, we provide a step-by-step protocol for PEM-seq, based on our earlier publications (Gan et al., 2021; Liu et al., 2021a; Yin et al., 2019; Zhang et al., 2021). The protocol described below depicts the specific steps for using HEK-293T cells. However, we have also successfully applied this protocol in human primary T cells, cancer cell lines (HeLa, MRC-5, K562, etc.), and mouse *abelson* virus-transformed pro-B, CH12F3, and mouse embryonic stem cells (mESCs). Before initiating the experiment, the audience should select a desired DSB site, termed the bait DSB, and design primers for PEM-seq.

Bait DSB selection

PEM-seq depends on the use of recurrent DSB as bait DSB to achieve the best performance. If there is a recurrent DSB site in your experiments, such as V(D)J recombination loci in lymphocytes, antibody class-switch recombination loci in B cells, or genome editing target sites, please skip this section and start to design primers for PEM-seq analysis. If not, the audience should introduce a bait DSB into the cell. PEM-seq is compatible with multiple types of DSB, including blunt ends, sticky ends, ends with adducts, hairpin, and nick transformed DSBs (Figure 2A). Therefore, the bait DSB can also be generated by a broad scope of enzymes, such as *Asi*I, *I-Sce*I, RAG, AID, transposons (e.g., Cre), and genome editing tools, e.g., FokI domain-containing nucleases, TALENs, and CRISPR-Cas. Typically, we use CRISPR-*SpCas9* to introduce the bait DSB at the *c-Myc* locus in mouse and human cells, which are provided in Table 1.

Primer design

For each bait DSB, at least two primers are required, including one biotinylated primer for the one-round primer extension and a second primer for the nested PCR. The two primers should be anchored on the same side of the bait DSB, either upstream or downstream, with nested primer in proximity to the cleavage site (Figure 2B). Genomic DNA is sonicated to 300–700 bp to generate PEM-seq libraries, which are sequenced with a read length at 2 × 150 bp. Moreover, each R1 read shares an identical sequence, termed bait sequence, which is the same sequence from the nested primer to the bait cleavage site (Figure 2B). Hence, to achieve the best performance, the bait length between the nested primer and bait cleavage site should be 50–110 bp and the flanking length between the biotinylated and nested primer is best at 10–50 bp (Figure 2B). Of note, both the two primers and bait sequence should avoid any potential repetitive regions. Regarding multiple PEM-seq libraries with the same bait DSB, we recommend user-defined barcode sequences to be included at the 5' of the nested primer (Exemplified in Table 2). Finally, the universal primers for PEM-seq should also be synthesized before initiating the experiment, including bridge adapters, I7-index, P5-I5, and P7-tag (Listed in Table 2). Of note, primers with or without modification(s) should be purified by HPLC (High-Performance Liquid Chromatography) or PAGE (Polyacrylamide gel electrophoresis), respectively, when ordering.

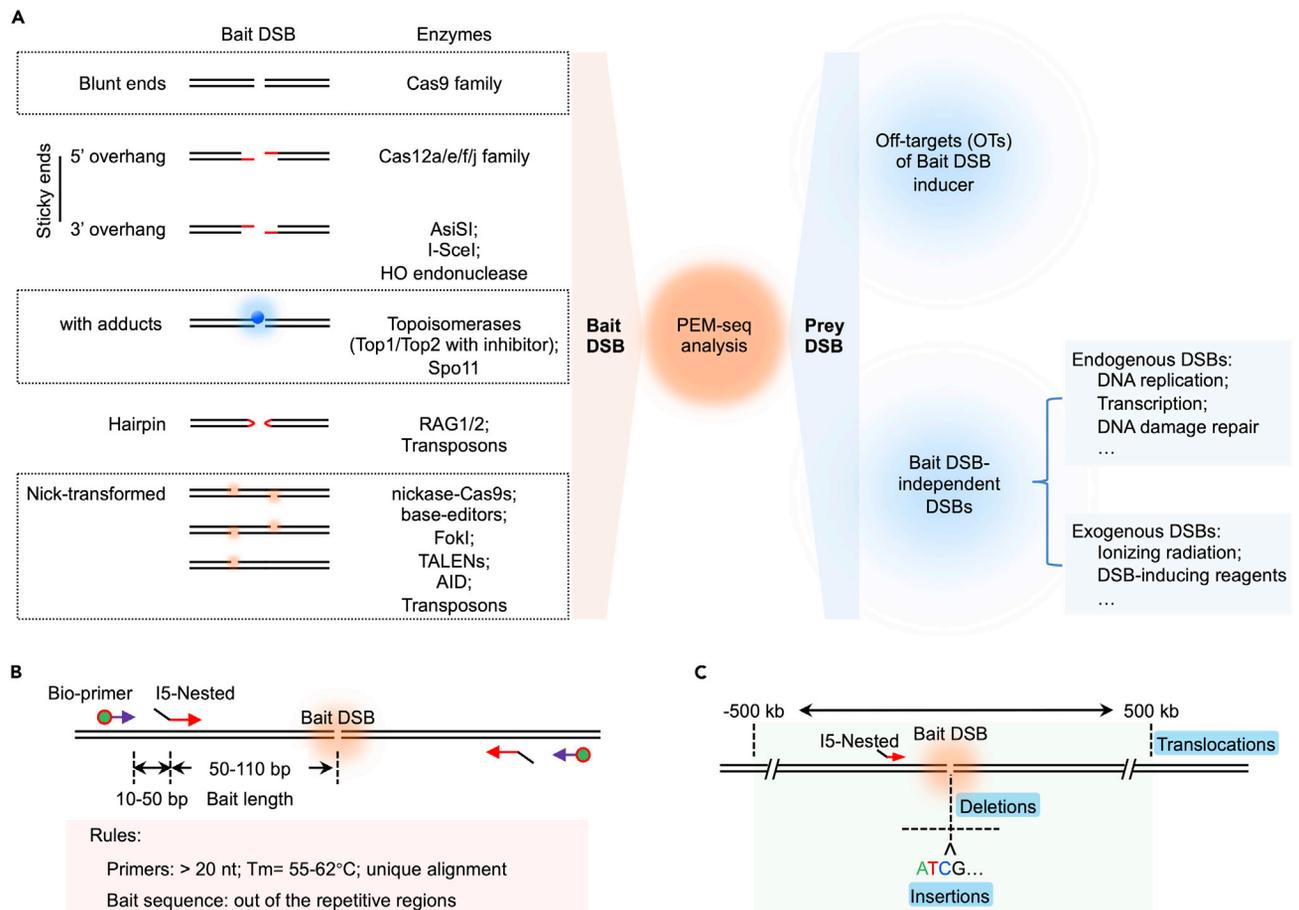


Figure 2. Principles for PEM-seq analysis

(A) PEM-seq is compatible with multiple types of bait and prey DSBs. Left: different types of bait DSBs, including blunt or sticky ends, ends with adducts, hairpin, nick-transformed DSB, etc., that can be analyzed by PEM-seq. Right: both off-target dependent and independent DSBs can be captured and analyzed by PEM-seq.

(B) Principles for the primer design. Generally, the distance from the start site of the nested primer to the cleavage site ranges from 50 to 110 bp, termed bait length, and the optimized distance between biotinylated primer (bio-primer) and nested primer is 10–50 bp.

(C) Definition of the repair products in PEM-Q. Events located in the ± 500 kb of bait DSB are grouped into insertions or deletions, and events out of the ± 500 kb of bait DSB are translocations.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals and recombinant proteins		
Nuclease-free water	Milli-Q, 0.22 mm filtered	N/A
DMEM	Corning	Cat#10-013-CV
opti-MEM	Gibco	Cat#31985070
Fetal Bovine Serum	Gibco	Cat#10091148
L-Glutamine	Corning	Cat#25-005-CI
Penicillin-Streptomycin Solution, 100×	Corning	Cat#30-002-CI
β -mercaptoethanol	Sigma-Aldrich	Cat#M3148
1× PBS, pH 7.4	Gibco	Cat#10010031
Polyethylenimine	Sigma-Aldrich	Cat#919012
Trypsin	Corning	Cat#25-052-CV
1 M Tris-HCl, pH 7.5	Invitrogen	Cat#15567027

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
5 M NaCl	Sigma-Aldrich	Cat#S6546
0.5 M EDTA, pH 8.0	Invitrogen	Cat#15575020
10% (wt/vol) SDS solution	Invitrogen	Cat#15553027
Proteinase K (20 mg/mL)	Invitrogen	Cat#AM2546
Isopropanol	Sigma-Aldrich	Cat#I9516
Ethanol	Sigma-Aldrich	Cat#1085430250
10× Isothermal Amplification Buffer II Pack	New England BioLabs	Cat#B0374S
Bst 3.0 DNA Polymerase	New England BioLabs	Cat#M0374L
dNTPs, 2.5 mM each	TransGen Biotech	Cat#AD101-01
Betaine solution, 5M	Sigma-Aldrich	Cat#B0300
Triton X-100	Sigma-Aldrich	Cat#T8787
Sodium hydroxide solution, 10 M	Sigma-Aldrich	Cat#72068
10× T4 DNA ligase buffer	Thermo Scientific	Cat#EL0011
T4 DNA ligase	Thermo Scientific	Cat#EL0011
PEG8000	Sigma-Aldrich	Cat#89510
10× EasyTaq buffer	TransGen Biotech	Cat#AP111-01
EasyTaq DNA Polymerase	TransGen Biotech	Cat#AP111-01
AMPure XP	Beckman Coulter	Cat#A63880
5× FastPfu buffer	TransGen Biotech	Cat#AP221-01
FastPfu DNA Polymerase	TransGen Biotech	Cat#AP221-01
Agarose	Thermo Scientific	Cat#75510019
Trans DNA Marker II	TransGen Biotech	Cat#BM411-01
6× DNA loading buffer	Beyotime Biotech	Cat#D0072
50× TAE buffer	Thermo Scientific	Cat#B49
GelRed Nucleic Acid Stain 10000× Water	Merck Millipore	Cat#SCT123
GeneJET Gel Extraction Kit	Thermo Scientific	Cat#K0692
Critical commercial assays		
Dynabeads MyOne streptavidin C1 beads	Invitrogen	Cat#65002
Experimental models: Organisms/strains		
Human cell line: HEK-293T	Lab stock	N/A
Oligonucleotides		
See Tables 1 and 2 for details	Sangon Biotech	N/A
Recombinant DNA		
pX330	(Cong et al., 2013)	Addgene; Cat#42230
pX330-MYC1	Lab stock	N/A; sgRNA targeting human c-MYC is cloned into pX330 by Bsal
pX330-GFP	Lab stock	N/A; SpCas9 is replaced by EGFP
Deposited data		
Example sequencing datasets & Original pictures for Figure 3	This paper; Mendeley Data	http://doi.org/10.17632/gjkh3wk4h4.1
Software and algorithms		
ImageJ 1.53c	(Schneider et al., 2012)	https://imagej.nih.gov/ij/download.html
PEM-Q	(Liu et al., 2021a)	https://github.com/liumz93/PEM-Q
Circos 0.69	(Krzywinski et al., 2009)	http://circos.ca/software/download/circos/
Prism 8	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
Others		
CO2 incubator	Nuaire	Cat#NU-5700
Thermomixer	Eppendorf	Cat#5382000023
Centrifuge	Eppendorf	Cat#5406000291
Spectrophotometer	DeNovix	Cat#DS-11 FX+
M220 Focused-ultrasonicator	Covaris	Cat#500295
MicroTUBE-130 AFA Fiber	Covaris	Cat#520045

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PCR machine	Eppendorf	Cat#6336000074
DynaMag-2	Invitrogen	Cat#12321D
DynaMa-PCR Magnet	Invitrogen	Cat#492025
Vortex Genie 2	VWR	Cat#G560E
VWR tube rotator UK plug	VWR	Cat#10136-084
Electrophoresis system	Thermo Scientific	Cat#FB-SBR-2025
UV Transilluminators	UVP	Cat#95-0461-02
ChemiDoc MP imaging system	Bio-Rad Laboratories	Cat#17001402
1.5 mL Microtubes	Axygen	Cat#MCT-150-C
0.2 mL PCR strip tubes	Axygen	Cat#PCR-0208-CP-C

MATERIALS AND EQUIPMENT

DMEM/10% FBS medium

Reagent	Final concentration	Amount
DMEM	n/a	440 mL
Fetal bovine serum	10% (vol/vol)	50 mL
L-Glutamine, 100× (vol/vol)	1× (vol/vol)	5 mL
Penicillin-Streptomycin Solution, 100× (vol/vol)	1× (vol/vol)	5 mL
β-mercaptoethanol (50 mM)	50 μM	0.5 mL
Total	n/a	500 mL

Filter-sterilize the medium with a 0.22 μm filter and store it at 4°C for up to one month.

⚠ **CRITICAL:** β-mercaptoethanol is toxic. Wear gloves, goggles, and face-shield to avoid skin exposure and inhalation.

PEI, 1 mg/mL

Reagent	Final concentration	Amount
Polyethylenimine	1 mg/mL	10 mg
ddH ₂ O	n/a	10 mL
Total	n/a	10 mL

Dissolve PEI in water that has been heated to 80°C; cool the solution to 25°C, and neutralize it to pH 7.0 by hydrochloric acid (HCl). Filter-sterilize the medium with a 0.22 μm filter, aliquot 1 mL into sterile tubes, and store it at –20°C for up to one year.

⚠ **CRITICAL:** Hydrochloric acid is toxic. Wear gloves, goggles, and face-shield to avoid skin exposure and inhalation.

Cell lysis buffer

Reagent	Final concentration	Amount
Tris-HCl, pH 7.5 (1 M)	10 mM	0.5 mL
NaCl (5 M)	200 mM	2 mL
EDTA, pH 8.0 (0.5 M)	2 mM	0.2 mL
10% SDS (wt/vol)	0.2% (wt/vol)	1 mL
ddH ₂ O	n/a	46.3 mL
Total	n/a	50 mL

Store it at room temperature (22°C–26°C, hereafter) for up to 6 months.

△ **CRITICAL:** SDS is toxic. Wear gloves and a mask to avoid skin exposure and inhalation.

TE buffer		
Reagent	Final concentration	Amount
Tris-HCl, pH 7.5 (1 M)	10 mM	0.5 mL
EDTA, pH 8.0 (0.5 M)	0.5 mM	0.05 mL
ddH ₂ O	n/a	49.45 mL
Total	n/a	50 mL

Store it at room temperature for up to 1 year.

1 × B&W buffer		
Reagent	Final concentration	Amount
Tris-HCl, pH 7.5 (1 M)	5 mM	0.25 mL
NaCl (5 M)	1 M	10 mL
EDTA, pH 8.0 (0.5 M)	1 mM	0.1 mL
ddH ₂ O	n/a	39.65 mL
Total	n/a	50 mL

Store it at room temperature for up to 1 year.

10 mM Tris-HCl, pH7.5		
Reagent	Final concentration	Amount
Tris-HCl, pH 7.5 (1 M)	10 mM	0.5 mL
ddH ₂ O	n/a	49.5 mL
Total	n/a	50 mL

Store it at room temperature for up to 1 year.

10% (vol/vol) Triton X-100		
Reagent	Final concentration	Amount
Triton X-100	10%	1 mL
ddH ₂ O	n/a	9 mL
Total	n/a	10 mL

Filter the solution through a 0.22 μm filter and store it at room temperature for up to 1 year.

Bridge adapter-upper/-lower		
Reagent	Final concentration	Amount
Bridge adapter-upper/-lower	400 μM	100 nmol
ddH ₂ O	n/a	250 μL
Total	n/a	250 μL

Dissolve the oligonucleotides in water, aliquot 50 μL into sterile tubes, and store it at –20°C for up to one year.

50% (wt/vol) PEG8000		
Reagent	Final concentration	Amount
PEG8000	50% (wt/vol)	5 g
ddH ₂ O	n/a	10 mL
Total	n/a	10 mL

Dissolve PEG8000 in ddH₂O at 50°C, and then adjust the volume to 10 mL. Prepare 1 mL aliquots and store them at –20°C for up to 1 year.

10 mM NaOH		
Reagent	Final concentration	Amount
NaOH (10 M)	10 mM	0.05 mL
ddH ₂ O	n/a	49.95 mL
Total	n/a	50 mL

Prepare 1 mL aliquots and store them at room temperature for up to 3 months.

⚠ **CRITICAL:** NaOH is toxic. Wear gloves to avoid skin exposure.

STEP-BY-STEP METHOD DETAILS

Generating bait DSB

⌚ **Timing:** 2 days

This step is aimed to introduce a bait DSB into cells and allow the formation of DSB repair products. The bait DSB can be induced by multiple approaches, including transfection, viral transduction, and nucleofection. Here, we describe the procedure of polyethylenimine (PEI)-mediated transfection with a plasmid containing *SpCas9-MYC1* in HEK-293T cells.

Note: If a bait DSB has been generated and then it underwent DSB repair in cells already, please skip this part and proceed with the genomic DNA isolation.

- Culture HEK-293T cells in DMEM/10% FBS medium to a 60%–80% coverage in the 6-cm dish. Two dishes of cells should be prepared. One is set as the control that is without the bait DSB, the other one is for bait DSB generation.
- PEI-mediated plasmid transfection.
 - Before transfection, bring all reagents to room temperature.
 - Remove the cell culture medium by aspiration, and add 4–5 mL pre-warmed (37°C) DMEM/10% FBS medium.
 - Prepare the transfection mixture.
 - Add 200 μ L opti-MEM and 5 μ g plasmid containing *SpCas9-MYC1* (inducing bait DSB by pX330-MYC1) or GFP (control, pX330-GFP) in a 1.5 mL tube. Mix well with gentle flaps.
 - Prepare two 1.5 mL tubes, each containing 200 μ L opti-MEM and 10–15 μ L PEI (1 mg/mL). Mix well by vortexing or pipetting.
 - Add one diluted PEI (from ii) to one diluted DNA (from i), and the total volume should be less than 450 μ L. Mix immediately with gentle flaps and incubate at room temperature for 15 min.
 - Dropwise add the DNA/PEI mixture to the cell culture by a pipette and put cells into a CO₂ incubator for 8–12 h.
 - Remove the medium by aspiration and add 5 mL pre-warmed (37°C) DMEM/10% FBS medium.
- Harvest transfected cells at 48–72 h post-transfection. Typically, over 0.1 million live cells are required for one PEM-seq library, and 1–3 million cells are recommended.
 - Aspirate the cell culture medium, wash cells with 3 mL 1 \times PBS, and aspirate the PBS.
 - Add 400 μ L 0.05% trypsin to cover all cells and incubate at 37°C for 2 min to allow cell detachment and separation.
 - Add 400 μ L DMEM/10% FBS medium to quench trypsin and collect cells into a 1.5 mL tube.
 - Spin at 300 g, 4°C, for 5 min and remove supernatant.
 - Wash once with 1 mL 1 \times PBS, spin, and discard the supernatant.

Isolating genomic DNA

⌚ **Timing:** 1 day

Table 1. Bait DSBs for PEM-seq analysis

Name of Bait DSB	Host	Tested Cell lines	Editing efficiency	Off-targets	Guide RNA sequence with PAM (5' > 3')
<i>Myc6</i>	Mouse	CH12F3, mouse Abelson virus-transformed pro-B, and mESC	40~60% in CH12F3	No	GCGGTGAGTCGTGATCT GAG CGG
<i>MYC1</i>	Human	HEK-293T, K562, HeLa, and human primary T cells	60~80% in K562	9 off-targets in HEK-293T	GGGATGTGGAGCTTGGC TAT GGG
<i>MYC-YH</i>	Human	HEK-293T and K562	30~40% in K562	No	GTGTCTCCTCATGGAGCA CC AGG

Note: The red bases show the protospacer adjacent motif (PAM) sequences of indicated sgRNA.

The genomic DNA for PEM-seq analysis is isolated and purified in this major step.

Note: Besides the procedures provided below, genomic DNA can be also isolated by any mainstream commercial kits, such as the PureLink Genomic DNA Purification Kit (Thermo), GenElute Mammalian Genomic DNA Miniprep Kits (Sigma-Aldrich), Monarch Genomic DNA Purification Kit (NEB), etc.

4. Prepare cell lysis master mixture for each sample by mixing 495 μ L of cell lysis buffer and 5 μ L of proteinase K (20 mg/mL).
5. Loosen the cell pellet with flaps, add 500 μ L of cell lysis master mixture, and incubate them in a thermomixer at 56°C, 500 rpm, for overnight (10–16 h).
6. Add 500 μ L of Isopropanol, mix immediately and thoroughly by inverting and shaking the microtube until the white pellet forms. [Troubleshooting 1](#)
7. Pick up the DNA pellet with a pipette tip, transfer it to a new 1.5 mL microtube containing 1 mL 70% (vol/vol) ethanol, and mix thoroughly by inverting and shaking.
8. Centrifuge it at 13,000 g for 5 min at 4°C.
9. Remove the supernatant completely, and air-dry the DNA pellet for 2–5 min.
10. Dissolve the DNA pellet in 150 μ L TE buffer and incubate at 56°C, 500 rpm, for at least 4 h.

⚠ CRITICAL: The DNA must be dissolved completely.

11. Determine the concentration of 1 μ L of the isolated genomic DNA with a spectrophotometer; the value of $A_{260/280}$ should be 1.8–2.0.

⏸ Pause point: Purified genomic DNA can be stored at –20°C for months or 4°C for a week.

Sonication

⌚ Timing: 2 h

In this section, the isolated genomic DNA should be sheared to small fragments with a peak length of 300–700 bp.

Table 2. DNA sequences for PEM-seq analysis

Usage	Name of adaptor or primer	Sequence (5' > 3')
Primer extension	Bio-primer	/5-Biotin/ EXTENSIONPRIMER
	Bio-primer-Myc6	/5-Biotin/CGGCACTAGGACTTGATGTTGGGCTAGCGC
	Bio-primer-MYC1 (+)	/5-Biotin/GGTGACTCACTTGGGAATCGGGAAGGTG
	Bio-primer-MYC1 (-)	/5-Biotin/CGGGAAGGAAGTTGGCATCTGTCCTG
	Bio-primer-MYC-YH	/5-Biotin/AGCTTGTACCTGCAGGATCTGAGCG
Bridge adapter ligation	Bridge adapter-upper	/5Phor/CCACGCGTGCTCTACA NNNTNNNNNTNNN AGATCGGAAGAGCACACGTCTGAACTCCAGT-NH2
	Bridge adapter-lower	TGTAGAGCACGCGTGGNNNNNN-NH2
Nested PCR	I5-Nested-1	ACTCTTCCCTACACGACGCTCTCCGATCT AGCG NESTEDPRIMER
	I5-Nested-2	ACTCTTCCCTACACGACGCTCTCCGATCT GCCT NESTEDPRIMER
	I5-Nested-3	ACTCTTCCCTACACGACGCTCTCCGATCT AGGAN NESTEDPRIMER
	I5-Nested-4	ACTCTTCCCTACACGACGCTCTCCGATCT TCAG NESTEDPRIMER
	I5-Nested-1-Myc6	ACTCTTCCCTACACGACGCTCTCCGATCT AGCG GGAAACCAGAGGGAATCCTC
	I5-Nested-1-MYC1 (+)	ACTCTTCCCTACACGACGCTCTCCGATCT AGCG CCTCAGAATAGGAGAGAGTG
	I5-Nested-1-MYC1 (-)	ACTCTTCCCTACACGACGCTCTCCGATCT AGCG AGAGCCATTCTCTGGCTCAG
	I5-Nested-1-MYC-YH	ACTCTTCCCTACACGACGCTCTCCGATCT AGCG AGTCCTGCGCTCGCAAGAC
	I7-Index-1	CAGAAGACGGCATAACGAGAT CGTGAT GTGACTGGAGTTCAGACGTGTGC
	I7-Index-2	CAGAAGACGGCATAACGAGAT ACATCG GTGACTGGAGTTCAGACGTGTGC
	I7-Index-3	CAGAAGACGGCATAACGAGAT GCCTAA GTGACTGGAGTTCAGACGTGTGC
	I7-Index-4	CAGAAGACGGCATAACGAGAT TGGTCA GTGACTGGAGTTCAGACGTGTGC
	I7-Index-5	CAGAAGACGGCATAACGAGAT CACTGT GTGACTGGAGTTCAGACGTGTGC
	I7-Index-6	CAGAAGACGGCATAACGAGAT ATTGGC GTGACTGGAGTTCAGACGTGTGC
	I7-Index-7	CAGAAGACGGCATAACGAGAT GATCTG GTGACTGGAGTTCAGACGTGTGC
	I7-Index-8	CAGAAGACGGCATAACGAGAT TCAAGT GTGACTGGAGTTCAGACGTGTGC
	I7-Index-9	CAGAAGACGGCATAACGAGAT CTGATC GTGACTGGAGTTCAGACGTGTGC
	I7-Index-10	CAGAAGACGGCATAACGAGAT AAGCTA GTGACTGGAGTTCAGACGTGTGC
	I7-Index-11	CAGAAGACGGCATAACGAGAT GTAGCC GTGACTGGAGTTCAGACGTGTGC
	I7-Index-12	CAGAAGACGGCATAACGAGAT TACAAG GTGACTGGAGTTCAGACGTGTGC
Tagged PCR	P5-I5	AATGATACGGCGACCACCGAGATCTACACACTCTTCCCTACACGACGC
	P7-Tag	CAAGCAGAAGACGGCATAACGAGAT

Note: 1. The orange bases with underline are the random molecular barcode (RMB) sequences on the bridge adapter. 2. The red and blue bases show the index on the I5-Nested and I7-Index primer, respectively. 3. The bold characters mark the primer sequences for primer extension or nested PCR. "+" and "-" is the strand of genome where the primer lies on.

12. Turn on the Covaris M220 focused-ultrasonicator and pre-cool it to 4°C.
13. Transfer 20–50 µg DNA to a microtube-130 and adjust the final volume to 130 µL with nuclease-free water.
14. Set the Covaris M220 by following the manufacturer's instructions (https://www.covaris.com/wp/wp-content/uploads/resources_pdf/pn_010252.pdf), and fragment the DNA to a target peak at 300–700 bp:

Parameter	Setting value
Temperature (°C)	4
Peak Incident Power (W)	50
Duty Factor (%)	20
Treatment Time (sec)	50–65
Cycles per Burst (cpb)	200

△ **CRITICAL:** The performance of sonication varies with different samples. Carry out a time course based on the above settings. In our lab, we usually set the treatment time at 60 or 50 s for human or mouse genomic DNA, respectively.

15. Run 1 µL of the sonicated DNA on a 1% (wt/vol) agarose gel in 1× TAE buffer; the size of DNA should be with a target peak at 300–700 bp. [Troubleshooting 2](#)

△ **CRITICAL:** The length distribution of DNA fragments is important, as the too short or too long DNA fragments will be lost in PEM-seq libraries. Of note, if the library DNA of PEM-seq is sequenced with a 2 × 250 bp read, the size of fragmented DNA can be larger, such as with a peak at 500–1000 bp. The longer sequencing reads are more precise to identify outcomes, however, a 2 × 150 bp read is good enough for most experiments and is much cheaper.

▣ **Pause point:** DNA fragments can be stored at –20°C for months or 4°C for a week.

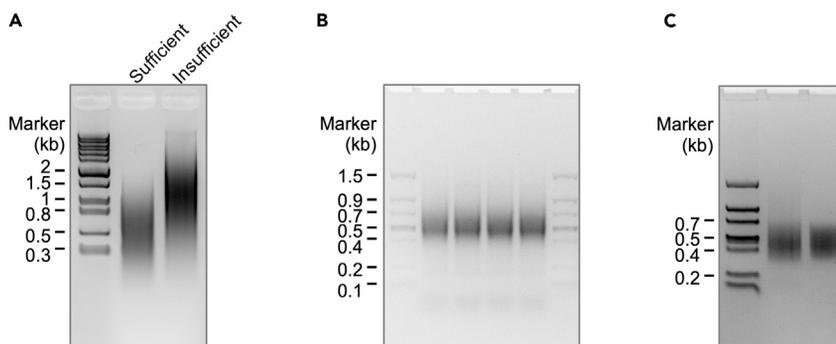


Figure 3. Representative images for preparing PEM-seq libraries

(A) Size distribution of fragmented DNA after sonication. The lane labeled sufficient shows the best length distribution of DNA fragments for PEM-seq, while the insufficient means DNA fragments are larger than the targeted length distribution.

(B) PCR products after the tagged PCR are separated by a 1.5% agarose gel. C. Size distribution of PEM-seq libraries after the last size selection. Each lane is a PEM-seq library DNA.

Primer extension

⌚ Timing: 1.5 h

DNA fragments containing the complementary sequence of the biotinylated primer are tagged with biotin at their 5' end by a one-round primer extension.

16. Prepare the primer anneal mixture on ice for each sample as below:

Reagent	Final concentration	Amount
10× Bst buffer	1×	16 μL
Bio-primer-MYC1 (+) (1 μM)	25 nM	4 μL
5 M Betaine	1 M	32 μL
sonicated DNA	6.25–250 ng/μL	1–40 μg
ddH ₂ O	n/a	To 160 μL
Total	n/a	160 μL

⚠ **CRITICAL:** The total amount of sonicated DNA for each sample should be 1–40 μg with an optimal amount of about 20 μg. If a larger amount of DNA is needed, scale up the mixture.

17. Aliquot each of the primer anneal mixtures into 4 PCR tubes.

18. Perform the primer anneal reaction with the following settings:

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	3 min	1
Denaturation	95°C	2 min	5 cycles
Annealing	58°C	3 min	
Final Annealing	58°C	3 min	1
Hold	10°C	Forever	

⚠ **CRITICAL:** The annealing temperature should be changed according to the melting temperature of biotinylated primer(s).

19. Set up the primer extension mixture on ice for each sample as below:

Reagent	Final concentration	Amount
10× Bst buffer	1×	4 μL
dNTPs (2.5 mM each)	50 μM	4 μL
Bst 3.0 DNA polymerase (8 U/μL)	0.1 U/μL	2.5 μL
ddH ₂ O	n/a	29.5 μL
Total	n/a	40 μL

⚠ **CRITICAL:** The amount of Bst 3.0 DNA polymerase is approximately 2 U/μg DNA. If the amount of DNA is more than 20 μg, scale up the polymerase.

20. Add 10 μL aliquots of the primer extension mixture to each PCR tubes in step 18, mix thoroughly by vortex.

21. Set the primer extension reaction as follows:

PCR cycling conditions			
Steps	Temperature	Time	Cycles
Primer extension	65°C	15 min	1
Inactivation	80°C	5 min	1
Hold	25°C	Forever	

Note: A 15-min of primer extension is sufficient to amplify DNA fragments with a peak length at 500–1000 bp.

Primer removal

⌚ Timing: 40 min

The exceeded biotinylated primers are removed in this major step, as the residual will be captured by streptavidin beads as well.

22. Place the AMPure XP beads at room temperature for at least 30 min before use.

23. Add 50 μ L of pre-warmed AMPure XP beads to each PCR tube with primer extension product by pipetting up and down 15–20 times and then incubate at room temperature for 5 min.

△ CRITICAL: The volume of AMPure XP beads, every new batch or after a long-time storage, should be tested before use by following the manufacturer's instructions (<https://www.beckman.com/reagents/genomic/cleanup-and-size-selection/pcr/a63880>). Principally, all of the free bio-primer should be removed and all DNA fragments larger than 200 bp should be kept. Typically, we use the beads ratio at 1.0 \times , which means that 50 μ L of the AMPure beads are mixed with 50 μ L of the primer extension product.

24. Place PCR tubes on the magnetic stand (DynaMa-PCR Magnet) at room temperature for 5 min.

25. Remove and discard the supernatant completely.

△ CRITICAL: Hold the PCR tubes on the magnetic stand and do not disturb the AMPure beads. Moreover, during the whole process of primer removal, the AMPure beads cannot be over air-dry.

26. Add 200 μ L of 70% ethanol to each tube and remove the supernatant completely.

△ CRITICAL: Leave the PCR tubes on the magnetic stand and do not disturb the AMPure beads.

27. Repeat step 26.

28. Add 50 μ L of 10 mM Tris-HCl, pH 7.5, and mix thoroughly by pipetting up and down 15–20 times.

29. Incubate at room temperature for 2 min.

30. Place PCR tubes on the magnetic stand at room temperature for 2 min.

31. Transfer 50 μ L of the clear supernatant from each PCR tube and pool the same sample together in a 1.5 mL microtube.

△ CRITICAL: Do not disturb or transfer any of the AMPure beads.

32. Check the concentration of the supernatant with a spectrophotometer; the total amount of the recovered DNA should be closed to the initial amount. [Troubleshooting 3](#)

▮▮ **Pause point:** DNA fragments can be stored at -20°C for months.

Streptavidin purification

⌚ **Timing:** 5 h

In this section, the biotinylated single-stranded DNA is captured by the streptavidin beads.

33. Incubate the purified DNA at 95°C for 5 min and immediately chilled on ice for 3 min.
34. Prepare the streptavidin binding mixture:

Reagent	Final concentration	Amount
Denatured DNA (from step 33)	n/a	200 μL
NaCl (5 M)	1 M	50 μL
EDTA (0.5 M, pH 8.0)	5 mM	2.5 μL
10% (vol/vol) Triton X-100	0.02%	0.5 μL
Total	n/a	253 μL

35. Prepare the streptavidin C1 beads.
a. Transfer 20 μL of streptavidin C1 beads for each sample to a new 1.5 mL microtube.

Note: If multiple samples are handled, scale up the amount of C1 beads.

- Add 400 μL of $1\times$ B&W buffer, mix well by pipetting.
- Place it on a DynaMag-2 holder at room temperature for 1 min.
- Remove and discard the supernatant completely.
- Repeat b-d for twice.
- Suspend the C1 beads with 20 μL of $1\times$ B&W buffer.

36. Add 20 μL of washed streptavidin C1 beads to the binding mixture from step 34 and incubate on a rotator at room temperature for 4 h.

⚠ **CRITICAL:** A 2 h incubation is sufficient to capture most of the biotinylated products, however, a 4 h incubation is recommended.

▮▮ **Pause point:** The incubation can be sustained overnight (12–16 h).

37. Spin the mixture at 200 g for 5 s.

⚠ **CRITICAL:** The centrifuge speed must be lower than 3000 rpm (or 1000 \times g), otherwise, the C1 beads will be broken, resulting in the loss of biotinylated products. Regarding this, a mini benchtop microcentrifuge is not recommended.

38. Place the mixture on a DynaMag-2 holder at room temperature for 1 min and remove the supernatant completely.
39. Suspend the C1 beads with 400 μL of $1\times$ B&W buffer, capture the beads on the magnet stand for 1 min and remove the supernatant completely.

40. Repeat step 39.
41. Resuspend the beads with 400 μL of 10 mM NaOH, immediately place the mixture on the magnet stand for 1 min, and discard the supernatant completely.

△ CRITICAL: The total incubation time with 10 mM NaOH must be less than 2 min.

42. Suspend the C1 beads with 400 μL of 10 mM Tris-HCl, pH 7.5, capture the beads on the magnet stand for 1 min and discard the supernatant completely.
43. Repeat step 42.
44. Suspend the C1 beads with 42.4 μL of 10 mM Tris-HCl, pH 7.5.

On-beads ligation

⌚ Timing: 5 h

The 3' end of single-stranded DNA, on the streptavidin beads, is ligated with the bridge adapter containing a 14-bp RMB.

45. Prepare the bridge-adapter mixture in a PCR tube:

Reagent	Final concentration	Amount
Bridge adapter-upper (400 μM)	200 μM	20 μL
Bridge adapter-lower (400 μM)	200 μM	20 μL
Total	n/a	40 μL

46. Set the bridge adapter assembly program:

PCR cycling conditions			
Steps	Temperature	Time	Cycle
Denaturation	95°C	3 min	1
Annealing	85°C, Ramp at 0.1°C/s	1 min	1
Annealing	80°C, Ramp at 0.1°C/s	1 min	1
Annealing	75°C, Ramp at 0.1°C/s	1 min	1
Annealing	70°C, Ramp at 0.1°C/s	1 min	1
Annealing	65°C, Ramp at 0.1°C/s	1 min	1
Annealing	60°C, Ramp at 0.1°C/s	1 min	1
Annealing	55°C, Ramp at 0.1°C/s	1 min	1
Annealing	50°C, Ramp at 0.1°C/s	1 min	1
Annealing	45°C, Ramp at 0.1°C/s	1 min	1
Annealing	40°C, Ramp at 0.1°C/s	1 min	1
Annealing	35°C, Ramp at 0.1°C/s	1 min	1
Annealing	30°C, Ramp at 0.1°C/s	1 min	1
Hold	10°C	forever	

△ CRITICAL: The ramp rate must be set at 0.1°C/s during the annealing processes.

47. Add 120 μL of nuclease-free water to bring the final concentration of annealed bridge adapter to 50 μM , prepare 80 μL aliquots and store them at -20°C for up to 1 year.
48. Set up an 80 μL ligation reaction as below:

Reagent	Final concentration	Amount
ssDNA on C1 beads (from step 44)	n/a	42.4 μ L
10 \times T4 DNA ligase buffer	1 \times	8 μ L
Bridge adapter (50 μ M from step 47)	1 μ M	1.6 μ L
T4 DNA ligase (400 U/ μ L)	20 U/ μ L	4 μ L
50% (wt/vol) PEG8000	15%	24 μ L
Total	n/a	80 μL

△ CRITICAL: Mix well all the reagents except PEG8000 by flaps, then add the PEG8000 with a cut-off P200 pipette tip and mix gently but thoroughly by pipetting up and down 10–15 times.

49. Incubate the ligation reaction at room temperature for 4 h or overnight (12–16 h) on a rotator with a rotation rate at 8 rpm.

△ CRITICAL: Do not spin the mixture. It is recommended to resuspend the mixture every 2 h during the incubation, however, the resuspension can be skipped during overnight ligation.

50. Add 320 μ L of 1 \times B&W buffer and mix thoroughly, capture the beads on the magnet stand for 1 min and remove the supernatant completely.

51. Resuspend the beads with 400 μ L of 1 \times B&W buffer, place the mixture on the magnet stand for 1 min, and discard the supernatant completely.

52. Repeat step 51.

53. Resuspend the beads with 400 μ L of 10 mM Tris-HCl, pH 7.5, place the mixture on the magnet stand for 1 min, and discard the supernatant completely.

54. Repeat step 53.

55. Resuspend the on-beads ligated DNA with 73 μ L of 10 mM Tris-HCl, pH 7.5.

Nested PCR

⌚ Timing: 1.5 h

The nested PCR is used to enrich DNA fragments around the bait DSB. Moreover, it also introduces Illumina indexes to distinguish multiple samples that will be sequenced in the same lane.

56. Set up a 100 μ L nested PCR mixture as below:

Reagent	Final concentration	Amount
On-beads ligation products (from step 55)	n/a	73 μ L
10 \times EasyTaq buffer	1 \times	10 μ L
dNTPs (2.5 mM each)	200 μ M	8 μ L
I5-Nested-1-MYC1 (+) (10 μ M)	400 nM	4 μ L
I7-index primer (10 μ M)	400 nM	4 μ L
EasyTaq DNA polymerase (5 U/ μ L)	0.05 U/ μ L	1 μ L
Total	n/a	100 μL

57. Mix thoroughly, aliquot the nested PCR mixture into 2 PCR tubes and perform amplification with the following program:

PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	5 min	1
Denaturation	95°C	1 min	15 cycles
Annealing	58°C	45 s	
Extension	72°C	1 min	
Final Extension	72°C	5 min	1
Hold	10°C	forever	

△ **CRITICAL:** DNA polymerase used in this step should be without the nuclease-dependent proofreading activity, as it will digest the beads-bound ssDNA. Moreover, do not spin the PCR mixture as it will greatly reduce the amplification efficiency.

Size selection

⌚ **Timing:** 40 min

DNA fragments larger than 300 bp are kept during the size selection to achieve the best performance of PEM-seq analysis.

58. Place the AMPure XP beads at room temperature for at least 30 min before use.
59. Add 40 μ L of pre-warmed AMPure XP beads to each PCR tube after the nested PCR by pipetting up and down 15–20 times and then incubate at room temperature for 5 min.

△ **CRITICAL:** The volume of AMPure XP beads should be tested before use. Principally, all DNA fragments larger than 300 bp should be kept. Typically, we use 40 μ L of the AMPure beads for 50 μ L of the PCR product.

60. Place PCR tubes on the magnetic stand (DynaMa-PCR Magnet) at room temperature for 5 min.
61. Remove and discard the supernatant completely.

△ **CRITICAL:** Hold the PCR tubes on the magnetic stand and do not disturb the AMPure beads. Moreover, the AMPure beads can be not air-dry.

62. Add 200 μ L of 70% ethanol to each tube and remove the supernatant completely.

△ **CRITICAL:** Leave the PCR tubes on the magnetic stand and do not disturb the AMPure beads.

63. Repeat step 62.
64. Add 35 μ L of 10 mM Tris-HCl, pH 7.5, and mix thoroughly by pipetting up and down 15–20 times.
65. Incubate at room temperature for 2 min.
66. Place PCR tubes on the magnetic stand at room temperature for 2 min.
67. Transfer 33 μ L of the clear supernatant from each PCR tube and pool the same sample together in a 1.5 mL microtube.

△ **CRITICAL:** Do not disturb or transfer the AMPure beads, as the AMPure beads greatly reduce the amplification efficiency of tagged PCR.

68. Check the concentration of 1 μL of the supernatant with a spectrophotometer; the concentration of the recovered DNA is less than 4 $\text{ng}/\mu\text{L}$, commonly ranging from 1 to 3 $\text{ng}/\mu\text{L}$. [Troubleshooting 4](#)

Tagged PCR

⌚ **Timing:** 1 h

In this section, PCR products are tagged by the Illumina adapter sequences with different indexes.

69. Set up a 100 μL tagged PCR mixture as below:

Reagent	Final concentration	Amount
PCR products (from step 67)	n/a	63 μL
5x FastPfu buffer	1x	20 μL
dNTPs (2.5 mM each)	200 μM	8 μL
P5-I5 (10 μM)	400 nM	4 μL
P7-tag (10 μM)	400 nM	4 μL
FastPfu DNA polymerase (2.5 U/ μL)	0.025 U/ μL	1 μL
Total	n/a	100 μL

70. Mix thoroughly, aliquot the tagged PCR mixture into 2 PCR tubes and perform PCR with the following program:

PCR cycling conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	95°C	2 min	1
Denaturation	95°C	30 s	15 cycles
Annealing	58°C	45 s	
Extension	72°C	1 min	
Final Extension	72°C	5 min	1
Hold	10°C	forever	

⚠ CRITICAL: To enhance the compatibility to different PCR machines and the efficiency of PCR, two 50 μL of aliquots are recommended. Moreover, the cycle number of PCR depends on the concentration of DNA in step 68. If the concentration is less than 2 $\text{ng}/\mu\text{L}$, perform 15 cycles; if the concentration is 2–4 $\text{ng}/\mu\text{L}$, perform 11–13 cycles; if the concentration is higher than 4 $\text{ng}/\mu\text{L}$, perform 10 cycles.

⏸ Pause point: The PCR products can be stored at -20°C for months.

Size selection

⌚ **Timing:** 3 h

PCR products ranging from 300 to 700 bp are kept for sequencing.

Note: Besides the procedures provided below, library DNA can be also size selected by AMPure XP beads (<https://www.beckman.com/reagents/genomic/cleanup-and-size-selection/pcr/a63880>) or SPRIselect beads (<https://www.mybeckman.cn/reagents/genomic/cleanup-and-size-selection/size-selection/b23317>) by following the manufacturer's instructions. Principally, DNA fragments ranging from 300 to 700 bp should be kept. Moreover, any mainstream gel extraction kit for DNA should work as well as the GeneJET Gel Extraction Kit.

71. Pool the same sample together, add 20 μL of 6 \times DNA loading buffer, mix well, and run the PCR products on a 1.5% (wt/vol) agarose gel in 1 \times TAE buffer.

△ CRITICAL: The time should be long enough to separate desired DNA products from by-products. Typically, the running time is over 1 h.

72. Excise gel with DNA fragments ranging from 300 bp to 700 bp, cut into small pieces, and transfer them to a 15 mL tube. [Troubleshooting 5](#)
73. Add 2 mL binding buffer from the GeneJET Gel Extraction Kit, mix thoroughly, and incubate at 56°C until all gel is dissolved.
74. Spin the mixture through a column, included in the GeneJET Gel Extraction Kit, at 10,000 g for 1 min, and then discard the flow-through.
75. Add 800 μL of wash buffer (from the gel extraction kit) to wash the column, spin it at 10,000 g for 1 min, and then discard the flow-through.
76. Spin the column at 10,000 g for 2 min, and transfer it to a new 1.5 mL microtube.
77. Add 50 μL of nuclease-free water to the column, incubate at room temperature for 2 min.
78. Spin the column at 10,000 g for 2 min and collect the eluted fraction.
79. Check the concentration of 1 μL of the library DNA with a spectrophotometer. The total amount of DNA should be 0.3–1 μg , with a concentration at 6–20 ng/ μL . [Troubleshooting 6](#)

⏸ Pause point: The library DNA can be stored at -80°C for months.

High-throughput sequencing

⌚ Timing: 3 days

In this section, pooled libraries with compatible indexes are sequenced on Illumina platforms.

80. Pool a proper number of PEM-seq libraries equally, and apply the mixed library DNA to be sequenced on Illumina sequencers.

△ CRITICAL: Make sure the pooled libraries have compatible indexes. Moreover, an optimal sequencing platform also has to be decided on, typically, we use Hi-seq 2500, with a 2 \times 150 bp. For each library, 3–5 million raw reads, with coverage at 1.5–2.5 \times as 1–2 million cells are recommended for one PEM-seq library, are sufficient for the following analysis.

81. Demultiplex each library based on its index(es), typically, this can be done with the bcl2fastq2 conversion software by following the manufacturer's instruction (https://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/bcl2fastq/bcl2fastq2-v2-20-software-guide-15051736-03.pdf).

Sequence reads processing

⌚ Timing: 3 h

Reads are processed by PEM-Q, which is available with a test sample datasheet and detailed instructions (<https://github.com/liumz93/PEM-Q>). Please install the pipeline following the instructions on the website.

82. Prepare information in the following order for each library:

Essential information for PEM-Q processing							
Genome	Name	Cut-site	Chromosome	P-start	P-end	P-strand	P-sequence

△ CRITICAL: "Genome" is the reference assembly, e.g. hg38, mm10 for alignment. "Name" stands for the basename of the input .fastq files, e.g. "/path/LY001_R1.fastq" should be "LY001". "Cut-site" is the breakpoint of the bait DSB on the strand of nested primer resident. "P-start", "P-end", "P-strand", and "P-sequence" are the starting point, endpoint, resident strand, and sequence of the nested primer. Of note, "P-sequence" must be in Upper characters, while lower characters are for the rest items except for the "Name". Information for *Myc6*, *MYC1 (+/-)*, and *MYC-YH* have been provided in the following table.

Example information for PEM-Q analysis								
Name of bait DSB	Genome	Name	Cut-site	Chromosome	P-start	P-end	P-strand	P-sequence (5' > 3')
<i>Myc6</i>	mm10	Name of your .fastq file	61986726	chr15	61986633	61986652	+	GGAAACCAGAGGGAATCCTC
<i>MYC1 (+)</i>	hg38	Name of your .fastq file	127743326	chr8	127743238	127743257	+	CCTCAGAATAGGAGAGAGTG
<i>MYC1 (-)</i>	hg38	Name of your .fastq file	127743326	chr8	127743381	127743400	-	AGAGCCATTCTCTGGCTCAG
<i>MYC-YH</i>	hg38	Name of your .fastq file	127738978	chr8	127738879	127738898	+	AGTCCTGCGCCTCGCAAGAC

Optional: The following information is only used for the vector integration analysis. Skip the following table, if no vector is used in your experiments.

Information for vector integration analysis only (Optional)							
Name	Name of the vector sequence file (.fa file)	Genome	Chromosome	P-strand	sgRNA-start	sgRNA-end	

Note: The "Name", "Genome", "Chromosome", and "P-strand" are the same as the essential information for PEM-Q processing. The "Name of vector sequence file" is the full title of the .fa file, e.g., "SpCas9_pX330.fa". "sgRNA-start" and "sgRNA-end" are the start and end point of the single guide (sg) RNA sequence on the vector. Examples are provided in the online instruction of PEM-Q.

83. Make a folder for each sample.

84. Move the sequence files to the indicated file folder and then execute step 85 under this folder.

85. Run the PEM-Q pipeline with the following command.

```
>PEM-Q.py ``Genome`` ``Name`` ``Cut-site`` ``Chromosome`` ``P-start`` ``P-end``
``P-strand`` ``P-sequence``
```

△ **CRITICAL:** The command must be executed in the folder containing the sequence files.

Optional: Execute the following command to identify vector integrations:

```
>vector_analyze.py ``Name`` ``Name of the vector sequence file`` ``Genome``  
``Chromosome`` ``P-strand`` ``sgRNA-start`` ``sgRNA-end``
```

86. After that, the output contains two types of files in the “results” folder, which will be used for the following analysis (Table below). The statistics file contains the number of each kind of editing event and the editing efficiency. Each .tab file is a kind of editing event, e.g., deletions, insertions, and intra- or inter-chromosomal translocations. Use the separated files for the analysis presented in the Quantification and statistical analysis section.

Editing events after PEM-Q analysis				
Types	Categories	Definition	Analysis	Files
Editing events	All editing events	Deletions, insertions, inversions, and translocations	Editing efficiency; Frequency of each type of editing events, & The top rank of editing events	*_Editing_events.tab *_Editing_events_dot_plot.pdf *_Editing_events.html *_statistics.txt
Deletions	Small deletions	Deletions, 1–100 bp around the bait DSB	Length distribution & Microhomology	*_Deletion.tab
	Large deletions	Deletions, 0.1–500 kb around the bait DSB		*_deletion_length.txt *_del_len_statistics.txt
Insertions	Small insertions	Insertions, < 20 bp in size, 0–500 kb around the bait DSB	Length distribution, Inserted sequences & Plasmid integration	*_Insertion.tab
	Large insertions	Insertions, ≥ 20 bp in size, 0–500 kb around the bait DSB		*_insertion_length.txt *_inser_len_statistics.txt
Translocations	Intra-chromosomal translocations	Junctions on the bait chromosome, but out of the +/- 500 kb around the bait DSB	Prey DSBs & Off-targets analysis	*_Translocation.tab
	Inter-chromosomal translocations	Junctions on other chromosomes		
Vector integrations	Vector integrations	Junctions on vector	Distribution of vector integrations	*_all_vector_2.2.tab

△ **CRITICAL:** Each line in the .tab file represents an editing event. The detailed description of output files is presented in the documentation of PEM-Q (<https://github.com/liumz93/PEM-Q>).

87. Convert the .tab files to .bdg files with commands listed below, and visualize the .bdg files in IGV.

```
>tab2bdg_PEMQ.py ``Full name of the .tab file`` ``Genome``
```

Optional: Execute the following command to visualize reads aligned to the vector sequence.

```
>vectorTab2bdg.py ``Full name of the .tab file containing vector`` ``the  
path/ Name of the vector sequence file (.fa)``
```

Table 3. Statistics analysis after PEM-Q processing

Library name	Control	SpCas9-MYC1 treated
Host & cell lines	Human, HEK-293T	Human, HEK-293T
Bait DSB	None	SpCas9-MYC1
Bio-primer & nested primer	Bio-/nested MYC1 (+)	Bio-/nested MYC1 (+)
Events	Hits or percentage	Hits or percentage
NoJunction (perfect re-joinings or non-cuttings)	587,577	118,768
Deletions	2,455	104,408
Small_deletions (<=100bp)	2,368	99,899
Large_deletions (>100bp)	87	4,509
Insertions	2,062	41,190
1_bp insertions	472	25,977
Small_insertions (<20bp)	985	34,048
Large_insertions (≥20bp)	1,077	7,142
Translocations	106	11,170
Vector integrations	106	3,119
Editing events	4,623	156,768
Total events	592,222	282,624
Editing efficiency (%)	0.78%	55.47%
Deletions (%)	0.41%	36.94%
Insertions (%)	0.35%	14.57%
Translocations (%)	0.02%	3.95%

EXPECTED OUTCOMES

Figure 3 shows representative examples of the size distribution of fragmented DNA for PEM-seq analysis, library DNA after tagged PCR, and library DNA for Hi-seq sequencing. Regarding the fragmented DNA by sonication, the peak length of DNA fragments should be at 300–700 bp (Figure 3A). A good DNA library should be 300–700 bp. Of note, the PCR products at 100–200 bp mainly only contain the Illumina primers (Figure 3B). Moreover, the total amount of DNA after size selection is from 300 ng to 1 μg. DNA subjected to sequencing should range from 300 to 700 bp as well, without any contaminations (Figure 3C).

QUANTIFICATION AND STATISTICAL ANALYSIS

PEM-seq quantifies all types of DSB repair outcomes, including perfect re-joinings or non-cuttings, insertions, deletions, and genome-wide translocations (Tables 3 and 4). Moreover, it also quantifies the editing efficiency and identifies the off-target activity of nuclease-dependent genome-editing tools. Following are the summary descriptions about the utilization:

1. The editing efficiency (E.E., Figure 4A) is output in the last row of the statistics file, calculated by the following formula:

$$\text{E.E.} = \text{Editing Events} / \text{Total Events}$$

2. Calculate the frequency of each DSB repair outcome, including perfect re-joining (NoJunction in the statistics file), deletions, insertions, intra- or inter-chromosomal translocations (Figures 4A and 4B). All required numbers can be found in the statistics file.
3. The length distribution of deletion is important to determine the end resection process during DSB repair. According to the distance, deletions are divided into small (shorter than 100 bp) and large deletions (ranging from 100 bp to 500 kb) (Figures 4C and 4D). If NHEJ defects, the frequency of large deletions will be increased.
4. Microhomology-mediated deletions (Figure 4E) are an important feature of the MMEJ pathway, as defective NHEJ promotes cells to utilize MMEJ, resulting in an increased frequency of

Table 4. Off-targets of SpCas9-MYC1 in HEK-293T cells

Off target	Chr	Start position	End position	Off-target sequences with PAM (5' > 3')	Junctions in control	Junctions in treated cells
OT1	chr9	127166926	127166949	AGGAAGTGGAGCTTGGCCTT GGG	0	62
OT2	chr8	19171846	19171869	AGGAAGTGGAGCTTGGCCTT GGG	0	31
OT3	chr8	19171768	19171791	GGGGTGTGGAGCTTGACTAT GAG	0	31
OT4	chr4	144444276	144444299	TGGGAGTGGAGCTTGGTTTT GGG	0	25
OT5	chr3	15065290	15065313	AGGATGAAGAGATTGGCTAT GGG	0	24
OT6	chr12	21299145	21299168	GGGAAGTGGAACTGGCTCT GGG	0	14
OT7	chr3	159731043	159731066	TGGATGTGCAGCCTGGCTAT TGG	0	10
OT8	chr8	19171905	19171928	GGAATTTGGCGCTTGATTAT AGA	0	5
OT9	chr12	3250772	3250795	GGTATGCAGAGCTTGGCTTT CGG	0	3

microhomology-mediated deletion and a longer length of utilized microhomology. Microhomology used by each editing event has been identified and listed in the .tab files.

- Small (<20 bp) and large (≥ 20 bp) insertions (Figure 4D) can be used to analyze the end of DSB and exogenous DNA integration, respectively. For instance, the staggering cleavage of SpCas9 at the target site induces a recurrent single-nucleotide insertion (Liu et al., 2021a). Moreover, the P-nucleotides resulting from RAG 1/2-mediated cleavage can also be detected in the small insertions fraction. Any exogenous DNA, such as the vector integrations during genome editing, can be identified from the large insertion fractions (Figure 4F). The insertion sequence has also been identified and listed in the .tab files.
- Translocation (Figure 4G) results from the joining between bait DSB and prey DSBs. The prey DSB may come from the off-target activity of nuclease-dependent genome-editing tools (restriction enzymes, FokI, TALEN, CRISPR/Cas), fragmented plasmids, endogenous DNA enzymes (such as RAG 1/2 and AID), and DNA metabolism (DNA replication, transcription, and DNA damage repair) (Figure 2A). Hence, the frequency of translocations indicates the DSB level and the translocation junctions indicate the breakpoints of prey DSBs. Combining the translocation frequency and junction positions help to identify recurrent DSBs in the genome. Specifically, for off-target identification, multiple translocation junctions should accumulate at the presumable cleavage sites of the off-target sequence that is similar to the target sequence. For example, the off-target activity of CRISPR-Cas9 induces DSBs at sequences similar to the sgRNA (Table 4). Of note, the vector integration can also be found in the translocation fractions. Our PEM-Q pipeline has already combined the insertion and translocation fractions during the vector integration analysis (Figure 4F).

LIMITATIONS

PEM-seq identifies the DSB repair outcomes and genome-wide DSBs by analyzing events around the bait DSB and chromosomal translocations between the bait DSBs and prey DSBs. Thus, samples without a bait DSB can't be used to prepare the PEM-seq library. Chromosomal translocation is a rare event formed between the bait and prey DSBs and requires a repair time of hours to days. Therefore, after the induction of DSB at the bait site, the treated cells must be cultured for at least 24 h, best at 48–72 h. In addition, a relatively large number of cells, typically > 0.1 million live cells are required for one library. PEM-seq tends to capture intra-chromosomal translocations with a higher frequency than inter-chromosomal translocations, similar to other translocation capture assays (Hu et al., 2016). In addition, both unsealed DSBs and repair outcomes losing bio- or nested primer binding site(s) are missed in the PEM-seq analysis. However, editing products without the primer binding site(s) are only less than 2.6% of total editing events (Liu et al., 2021a).

TROUBLESHOOTING

Problem 1

The genomic DNA forms a transparent and gel-like pellet in step 6.

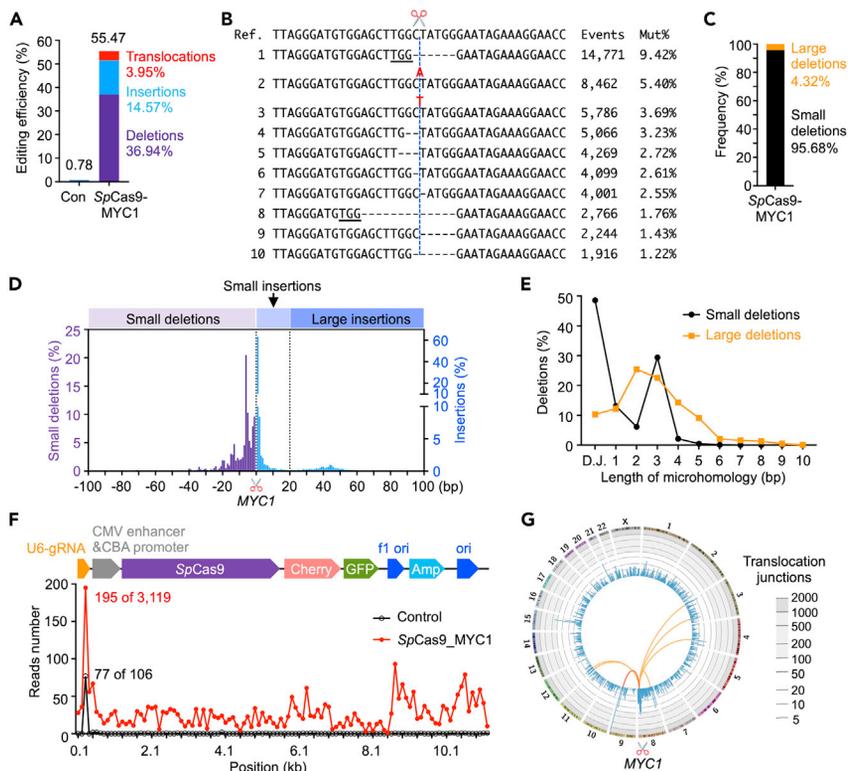


Figure 4. PEM-seq comprehensively quantifies editing outcomes of genome editing tools

(A) The editing efficiency of SpCas9 at MYC1 locus in HEK-293T cells. The frequency of deletion, insertion and translocation events are indicated. Con, control.

(B) The top ten editing events generated by SpCas9-MYC1 in HEK-293T cells. The red characters show the inserted bases and the dashed blue line marks the cleavage site. The underline shows microhomology. Mut% shows the frequency of each product in total editing events. Ref., sequence of the reference assembly.

(C) The frequency of small (≤ 100 bp) and large (> 100 bp) deletions induced by SpCas9-MYC1 in HEK-293T cells.

(D) The length distribution of small deletions (purple), small (< 20 bp) and large (≥ 20 bp) insertions (blue) generated by SpCas9-MYC1 in HEK-293T cells.

(E) Microhomology with indicated length in small or large deletions among the total deletions.

(F) The distribution of inserted sequence across the vector backbone (schematics on the top) at MYC1 in HEK-293T cells. The maximum reads and total event in control (black) and SpCas9-MYC1 (red) are showed.

(G) Circos plot showing the genome-wide translocations (blue bars) and the off-target sites (the color lines linked to the bait DSB, labeled with scissor) of SpCas9-MYC1 in HEK-293T cells.

Potential solution

The mixing procedure is not sufficient. Continuously invert and shake the microtube by hand until the white pellet is formed.

Problem 2

The peak length distribution of sonicated DNA is larger than 700 bp or smaller than 300 bp in step 15.

Potential solution

If the peak length of DNA fragments is larger than 700 bp, do the sonication again with an alternation of duration. If the DNA fragments are too small, take a new aliquot of 20–50 μg DNA and perform the sonication again with a shorter treatment time.

Problem 3

The total amount of recovered DNA is too low in step 32.

Potential solution

Two potential reasons will cause this, little DNA is bound on the beads, and (or) little beads-bound DNA is eluted. If little DNA is bound on the beads, more AMPure beads, thorough mixing, and a longer incubation time with beads will be helpful. A longer incubation time of beads with 10 mM Tris-HCl and the thorough mixing will increase the yield of purified DNA.

Problem 4

The concentration of purified DNA from nested PCR is too low or too high in step 68.

Potential solution

If the amount of DNA product is too little, it may result from multiple possibilities, including little DNA on the streptavidin C1 beads, insufficient on-beads ligation, and insufficient amount of AMPure beads. However, the most potential cause for a higher concentration is the insufficient removal of biotinylated primer, due to poor quality or inadequate storage of AMPure beads. The residual bio-primer can be captured by streptavidin C1 beads, ligated with the bridge adapter, and amplified during the nested PCR. Therefore, every new batch of commercial AMPure beads or any old beads after a long-time storage should be tested before use. Check and solve the problem based on these possibilities.

Problem 5

No or little DNA smear, a bright band between 100–200 bp, or very long DNA smear tail in step 72.

Potential solution

If there is no or little DNA smear, it can result from that: 1) AMPure beads are in the PCR reactions and (or) 2) the cycle number of PCR is few. The insufficient biotinylated primer removal will result in a bright band at 100–200 bp. Finally, a very long DNA smear tail indicates that the cycle number of tagged PCR is too many.

Problem 6

The concentration of purified DNA from tagged PCR is too low or too high in step 79.

Potential solution

Increase or decrease the cycle number of tagged PCR if the amount of purified DNA is too few or too many, respectively.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jiazhi Hu (hujz@pku.edu.cn).

Materials availability

This study did not generate new unique reagents. The PEM-Q pipeline is available at GitHub with instructions (<https://github.com/liumz93/PEM-Q>).

Data and code availability

The raw and analyzed sequence data from our original paper carrying out PEM-seq in HEK-293T cells (Zhang et al., 2021) can be found in National Omics Data Encyclopedia (NODE) under the accession number OEP001824. The run-id of SpCas9-MYC1 treated sample and control is OER195774 and OER195780, respectively. The original pictures for Figure 3 and testing high-throughput sequencing data are deposited to Mendeley Data (<http://doi.org/10.17632/gjkh3wk4h4.1>).

ACKNOWLEDGMENTS

We thank the members of the Hu laboratory for their helpful comments and critical reading. We also acknowledge the National Center for Protein Sciences at Peking University in Beijing, China for their

technical help. This work was supported by the National Key R&D Program of China (2017YFA0506700), the NSFC grant (31771485 and 32122018), and the SLS-Qidong Innovation Fund. J.H. is an investigator of PKU-TSU Center for Life Sciences and Y.L. is supported by the Boehringer Ingelheim-Peking University Postdoctoral Program.

AUTHOR CONTRIBUTIONS

J.H. conceptualized and designed the assay. Y.L., J.Y., T.G., and J.H. developed PEM-seq; M.L. wrote the PEM-Q pipeline. Y.L., J.Y., T.G., C.X., and W.Z. performed the experiments; Y.L., J.Y., T.G., M.L., C.X., W.Z., and J.H. analyzed the data. Y.L., J.Y., T.G., and J.H. wrote the paper. The authors read and approved the final manuscript.

DECLARATION OF INTERESTS

Patent applications have been filed relating to the PEM-seq assay. (Application number in China: CN201910199103.8, and the international application No. PCT/CN2020/098360)

REFERENCES

- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. <https://doi.org/10.1126/science.1231143>.
- Dong, J., Panchakshari, R.A., Zhang, T., Zhang, Y., Hu, J., Volpi, S.A., Meyers, R.M., Ho, Y.J., Du, Z., Robbiani, D.F., et al. (2015). Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching. *Nature* 525, 134–139. <https://doi.org/10.1038/nature14970>.
- Gan, T., Wang, Y., Liu, Y., Schatz, D.G., and Hu, J. (2021). RAG2 abolishes RAG1 aggregation to facilitate V(D)J recombination. *Cell Rep* 37, 109824. <https://doi.org/10.1016/j.celrep.2021.109824>.
- Hu, J., Meyers, R.M., Dong, J., Panchakshari, R.A., Alt, F.W., and Frock, R.L. (2016). Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.* 11, 853–871. <https://doi.org/10.1038/nprot.2016.043>.
- Hu, J., Zhang, Y., Zhao, L., Frock, R.L., Du, Z., Meyers, R.M., Meng, F.L., Schatz, D.G., and Alt, F.W. (2015). Chromosomal loop domains direct the recombination of antigen receptor genes. *Cell* 163, 947–959. <https://doi.org/10.1016/j.cell.2015.10.016>.
- Kim, D., Luk, K., Wolfe, S.A., and Kim, J.S. (2019). Evaluating and enhancing target specificity of gene-editing nucleases and deaminases. *Annu. Rev. Biochem.* 88, 191–220. <https://doi.org/10.1146/annurev-biochem-013118-111730>.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. <https://doi.org/10.1101/gr.092759.109>.
- Li, H., Yang, Y., Hong, W., Huang, M., Wu, M., and Zhao, X. (2020). Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduct Target Ther.* 5, 1. <https://doi.org/10.1038/s41392-019-0089-y>.
- Liu, M., Zhang, W., Xin, C., Yin, J., Shang, Y., Ai, C., Li, J., Meng, F.L., and Hu, J. (2021a). Global detection of DNA repair outcomes induced by CRISPR-Cas9. *Nucleic Acids Res.* 49, 8732–8742. <https://doi.org/10.1093/nar/gkab686>.
- Liu, Y., Ai, C., Gan, T., Wu, J., Jiang, Y., Liu, X., Lu, R., Gao, N., Li, Q., Ji, X., and Hu, J. (2021b). Transcription shapes DNA replication initiation to preserve genome integrity. *Genome Biol.* 22, 176. <https://doi.org/10.1186/s13059-021-02390-3>.
- Meng, F.L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberger, D., et al. (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159, 1538–1548. <https://doi.org/10.1016/j.cell.2014.11.014>.
- Saha, K., Sontheimer, E.J., Brooks, P.J., Dwinell, M.R., Gersbach, C.A., Liu, D.R., Murray, S.A., Tsai, S.Q., Wilson, R.C., Anderson, D.G., et al. (2021). The NIH somatic cell genome editing program. *Nature* 592, 195–204. <https://doi.org/10.1038/s41586-021-03191-1>.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. <https://doi.org/10.1038/nmeth.2089>.
- Tubbs, A., and Nussenzweig, A. (2017). Endogenous DNA damage as a source of genomic instability in cancer. *Cell* 168, 644–656. <https://doi.org/10.1016/j.cell.2017.01.002>.
- Tubbs, A., Sridharan, S., van Wietmarschen, N., Maman, Y., Callen, E., Stanlie, A., Wu, W., Wu, X., Day, A., Wong, N., et al. (2018). Dual roles of Poly(dA:dT) tracts in replication initiation and fork collapse. *Cell* 174, 1127–1142.e19. <https://doi.org/10.1016/j.cell.2018.07.011>.
- Yin, J., Liu, M., Liu, Y., Wu, J., Gan, T., Zhang, W., Li, Y., Zhou, Y., and Hu, J. (2019). Optimizing genome editing strategy by primer-extension-mediated sequencing. *Cell Discov* 5, 18. <https://doi.org/10.1038/s41421-019-0088-8>.
- Zhang, W., Yin, J., Zhang-Ding, Z., Xin, C., Liu, M., Wang, Y., Ai, C., and Hu, J. (2021). In-depth assessment of the PAM compatibility and editing activities of Cas9 variants. *Nucleic Acids Res.* 49, 8785–8795. <https://doi.org/10.1093/nar/gkab507>.