

RESEARCH ARTICLE

Open Access



DNA polymerases in precise and predictable CRISPR/Cas9-mediated chromosomal rearrangements

Mohammadreza M. Mehryar^{1,2†}, Xin Shi^{1,2†}, Jingwei Li^{1,2} and Qiang Wu^{1,2*} 

Abstract

Background Recent studies have shown that, owing to its cohesive cleavage, Cas9-mediated CRISPR gene editing outcomes at junctions of chromosomal rearrangements or DNA-fragment editing are precise and predictable; however, the underlying mechanisms are poorly understood due to lack of suitable assay system and analysis tool.

Results Here we developed a customized computer program to take account of staggered or cohesive Cas9 cleavage and to rapidly process large volumes of junctional sequencing reads from chromosomal rearrangements or DNA-fragment editing, including DNA-fragment inversions, duplications, and deletions. We also established a sensitive assay system using *HPRT1* and *DCK* as reporters for cell growth during DNA-fragment editing by Cas9 with dual sgRNAs and found prominent large resections or long deletions at junctions of chromosomal rearrangements. In addition, we found that knockdown of *PoIQ* (encoding Pol θ polymerase), which has a prominent role in theta-mediated end joining (TMEJ) or microhomology-mediated end joining (MMEJ), results in increased large resections but decreased small deletions. We also found that the mechanisms for generating small deletions of 1 bp and >1 bp during DNA-fragment editing are different with regard to their opposite dependencies on Pol θ and Pol λ (encoded by the *PoLL* gene). Specifically, Pol θ suppresses 1 bp deletions but promotes >1 bp deletions, whereas Pol λ promotes 1 bp deletions but suppresses >1 bp deletions. Finally, we found that Pol λ is the main DNA polymerase responsible for fill-in of the 5' overhangs of staggered Cas9 cleavage ends.

Conclusions These findings contribute to our understanding of the molecular mechanisms of CRISPR/Cas9-mediated DNA-fragment editing and have important implications for controllable, precise, and predictable gene editing.

Keywords Gene editing, CRISPR/Cas9, DNA-fragment editing, DNA polymerase, *PoIQ*, *PoLL*, *PoIM*, Large resection, Staggered cleavage, Chromosomal rearrangement

[†]Mohammadreza M. Mehryar and Xin Shi contributed equally to this work.

*Correspondence:

Qiang Wu

qiangwu@sjtu.edu.cn

¹ Center for Comparative Biomedicine, Ministry of Education Key Laboratory of Systems Biomedicine, State Key Laboratory of Systems Medicine for Cancer, School of Life Sciences and Biotechnology, Institute of Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China

² WLA Laboratories, Shanghai 201203, China



Background

CRISPR gene editing outcomes are generated from cellular ligations of double-strand break (DSB) ends after Cas9 cleavages. This occurs either via homologous recombination (HR) during the S and G2 cell cycle phases or via non-homologous end joining (NHEJ) throughout all four phases of the cell cycle. The former results in precise modifications while the latter is associated with indels that are difficult to predict [1–4]. Recent studies revealed that NHEJ can be further divided into cNHEJ (canonical NHEJ) and alt-NHEJ (alternative NHEJ), a.k.a microhomology-mediated end joining MMEJ or TMEJ. While cNHEJ may be accurate and requires Ku70/80 and Pol λ , MMEJ is error-prone and requires Pol θ . However, a major outstanding issue, particular with regard to the NHEJ process, is our incomplete understanding of the underlying mechanisms, including the identity of the cellular DNA polymerases that are involved in repairing the DSB ends following Cas9 cleavages [5–9].

An excellent model system to obtain a better understanding of CRISPR gene editing is the use of dual sgRNAs to investigate mechanisms of Cas9-mediated chromosomal rearrangements and 3D genome engineering [4, 10–13]. In particular, Cas9 programmed with dual sgRNAs can result in chromosomal rearrangements including DNA-fragment deletions, inversions, and duplications (Additional file 1: Fig. S1) [11, 14–19]. The mechanism underlying these different types of chromosomal rearrangement is not known but may be related to the NHEJ pathway by direct ligations of two of the four DSB ends resulted from the double cutting [11, 12, 20]. Details of this mechanism can thus also inform processes associated with normal chromosomal rearrangements, which are known to promote genome instability in cancers or generate immune diversity during development [21, 22].

Cas9-mediated chromosomal rearrangements programmed with dual sgRNAs are a model system to investigate mechanisms of CRISPR gene editing and 3D genome folding [4, 10, 11, 13]. The advantage of using Cas9 with dual sgRNAs over single sgRNAs is that the repair outcomes of chromosomal rearrangements cannot be recut whereas there is repeated cutting and re-ligation for Cas9 with single sgRNAs [19, 23]. Consequently, repair outcomes of Cas9-mediated nucleotide insertions at ligation junctions of chromosomal rearrangements are more precise and predictable than those at editing sites with single sgRNAs [18, 19, 24–30]. These precise insertions of predictable nucleotides at editing sites are thought to be resulted from fill-in and ligation of staggered DSB ends of Cas9 cleavages. However, the underlying DNA polymerase(s) remains unknown. Here we systematically analyzed the role of DNA polymerases in

CRISPR/Cas9-mediated chromosomal rearrangements and found prominent roles of Pol λ and Pol θ in processing DSB ends during DNA-fragment editing yet with unexpected specificities.

Results

Reporter assay systems for large resections

In contrast to the small insertions from staggered or cohesive Cas9 cleavages, there is little known about either small or large deletions. To provide mechanistic details into these processes, we first developed reporter assays using *HPRT1* (hypoxanthine phosphoribosyltransferase 1) and *DCK* (deoxycytidine kinase) systems. The *HPRT1* gene functions in the purine synthesis pathway and the encoded enzyme converts 6-thioguanine (6-TG) into a toxic product of thioguanine nucleotides. Thus, only *HPRT1*-defective cells can survive in 6-TG supplemented medium (Additional file 1: Fig. S2A and B). The *DCK* gene encodes an essential enzyme for DNA synthesis. Therefore, *DCK*-defective cells cannot grow in normal medium (Additional file 1: Fig. S2C and D). Thus, successful DSB events induced by Cas9 programmed with single sgRNAs targeting exons or with dual sgRNAs targeting introns of these genes can be readily assayed by cell growth in these two systems (Additional file 1: Fig. S2).

More specifically, if we design single sgRNAs targeting exonic sequences (Additional file 1: Fig. S2A and C), these two reporter systems can assay the efficiency of Cas9-induced DSB repair (Additional file 1: Fig. S2B and D). If we design dual sgRNAs targeting exon-proximal intronic sequences (Additional file 1: Fig. S2E and G), these reporter systems can be used to assay large resections (long deletions) into the flanking exonic sequences (Additional file 1: Fig. S2F and H) because, with no large resection into the flanking exons, pre-mRNA splicing will not disrupt the normal function of *HPRT1* or *DCK*. We first examined the efficiency and sensitivity of these reporter systems (Additional file 1: Fig. S2B, D, E, H) and indeed both the *HPRT1* and *DCK* reporter assay systems indicated that there exist large resections into the flanking exons for Cas9 programmed with dual sgRNAs (Additional file 1: Fig. S2F and H).

Polymerases in Cas9-induced large resections or long deletions

We then used these reporter assay systems to investigate the roles of DNA polymerase genes (*PolM*, *PolD*, *PolL*, *PolQ*, and *PolK*) in Cas9-induced large resections via knockdown of each of these five polymerases in HEC-1-B cells (Additional file 1: Fig. S3). For the *HPRT1* reporter assay, upon *PolL* (encoding Pol λ) or *PolQ* knockdown, especially *PolQ* knockdown, there is more cell growth

despite adding 6-TG (Fig. 1A and B), which suggested that Pol θ and Pol λ play a role in Cas9-induced large resections into the flanking exons of *HPRT1*. As a control, RT-PCR experiments demonstrated normal splicing in wild-type cells upon dual Cas9 cleavages within the intron 2 of *HPRT1* (Fig. 1C). However, there are observable decreases of spliced *HPRT1* mRNA upon *PolL* or *PolQ* knockdown (Fig. 1C), suggesting that there are increased large resections into the flanking *HPRT1* exons upon perturbation of *PolL* or *PolQ*. We made similar observations of increased large resections upon *PolL* or *PolQ* knockdown using the DCK reporter system (Fig. 1D–F). DNA sequencing confirmed large resections from the second targeting site within intron 2 into the downstream exon 3 of *HPRT1* during DNA-fragment deletion (Fig. 1G). In addition, we also confirmed large resections during DNA-fragment inversion (Fig. 1H). Finally, there exist large resections at the upstream cleavage junction (Fig. 1I). Several examples of large resections at these junctions in the *HPRT1* and *DCK* loci are shown in the additional file (Additional file 1: Fig. S4). Together, these data suggest that both Pol λ and Pol θ play a role in Cas9-induced large resections.

We then adopted LAM-HTGTS [31] to assay large resections during chromosomal rearrangements induced by Cas9 with dual sgRNAs. LAM-HTGTS can assay both DNA-fragment deletions and inversions simultaneously (Fig. 1J). We observed a higher frequency of DNA-fragment deletions compared with inversions during DNA-fragment editing (Fig. 1K). We also found that knockdown of *PolQ* or *PolL* results in significant decreases in repaired DNAs, suggesting again that Pol θ and Pol λ are required for DSB repairs during DNA-fragment editing (Fig. 1L).

Computational analyses of the next-generation sequencing (NGS) data with a customized computer program (see reads processing in “Methods” and Additional file 2: Notes S1–S3) identified rare but significant portion of high-throughput sequencing reads for the large resections during DNA-fragment deletions (Fig. 1M).

The customized computer program uses two consecutive dynamic programming to map the query sequences to the upstream and downstream of the cleavage site, leaving the middle insertion unmapped at the cleavage site. We also identified a large number of reads for the large resections that occurred during DNA-fragment inversions (Fig. 1N). Finally, there exists a large number of reads of large resections at the upstream cleavage junctions (Fig. 1O). These data demonstrated that there are asymmetrical large resections at the Cas9 cleavage site. Importantly, *PolQ* knockdown results in significant increases of large resections at all of these chromosomal rearrangement junctions (Fig. 1P–R), in line with observed increases of large resections by the *HPRT1* and *DCK* reporter assay systems (Fig. 1A–F). We also observed that the vast majorities of sequencing reads at junctions of chromosomal rearrangements have small indels and that *PolQ* knockdown exhibits a more pronounced effect on chromosomal rearrangements programmed with dual sgRNAs (Additional file 1: Fig. S5A, B, D, E) than on editing outcomes from single cleavages (Additional file 1: Fig. S5C and F).

Pol θ in small deletions at editing sites

In addition to rare large resections, small deletions are more frequently observed at junctions of chromosomal rearrangements during DNA-fragment editing. We found that upon *PolQ* but not *PolD* or *PolK* knockdown, there are consistent and significant decreases of small deletions at both upstream and downstream junctions of fragment inversions (Fig. 2A, B) as well as at the junctions of fragment deletions (Fig. 2C) and duplications (Fig. 2D) at the *MeCP2* locus. We then examined the role of *PolQ* at four additional loci (namely, *MAZ*, *PRDM5*, *PARP1*, and *YY1*). These five loci encode important epigenetic regulators or chromatin architectural proteins which we are interested in their mechanistic role in 3D genome folding. We used our DNA-fragment editing systems with double cutting guided by two sgRNAs. *PolQ* knockdown results in significant decreases of small deletions at junctions

(See figure on next page.)

Fig. 1 DNA polymerases in Cas9-induced large resections. Significant increases of resistance to 6-thioguanine (A, B) and decreases of normal splicing (C) by Cas9-induced large resections of *HPRT1* upon knockdown of *PolQ* or *PolL*. Significant increases of sensitivity (D, E) and decreases of normal splicing (F) by Cas9-induced large resections of DCK upon knockdown of *PolQ* or *PolL* (see Additional file 5: Table S3, $n = 2$ replicates, mean \pm SEM). Spliced *HPRT1* (C) and *DCK* (F) cDNAs are TA cloned and confirmed by Sanger sequencing in both orientations. Confirmation of large resections by DNA sequencing during DNA-fragment deletion (G) and inversion (H) as well as at the upstream cleavage junction (I) programmed by Cas9 with dual sgRNAs. Schematic (J) of LAM-HTGTS with biotinylated and nested primers and simultaneous assessment of DNA-fragment deletion and inversion (K) by next-generation sequencing (NGS). Significant decreases of repaired DNA, measured as ratio of reads with junctions to the total number of reads, upon knockdown of *PolQ* or *PolL* (L) (see Additional file 5: Table S3, $n = 2$ replicates, mean \pm SEM). Significant increases of Cas9-induced large resections during DNA-fragment deletion (M, P) and inversion (N, Q) as well as at the upstream cleavage junction (O, R) assayed by NGS upon knockdown of *PolQ* (P–R). Percentages of rare resection products were quantified as ratio of large resection reads to the total number of reads

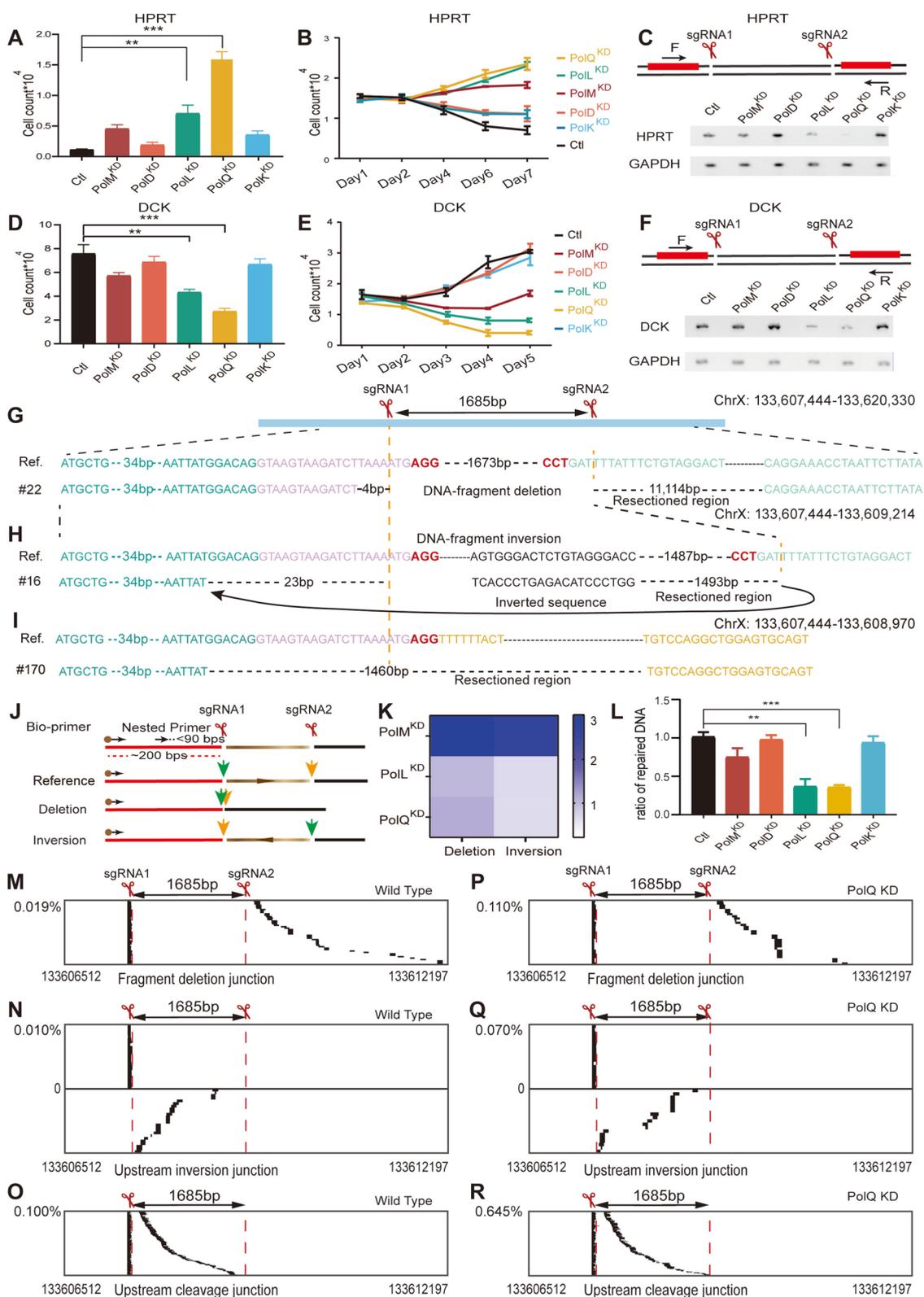


Fig. 1 (See legend on previous page.)

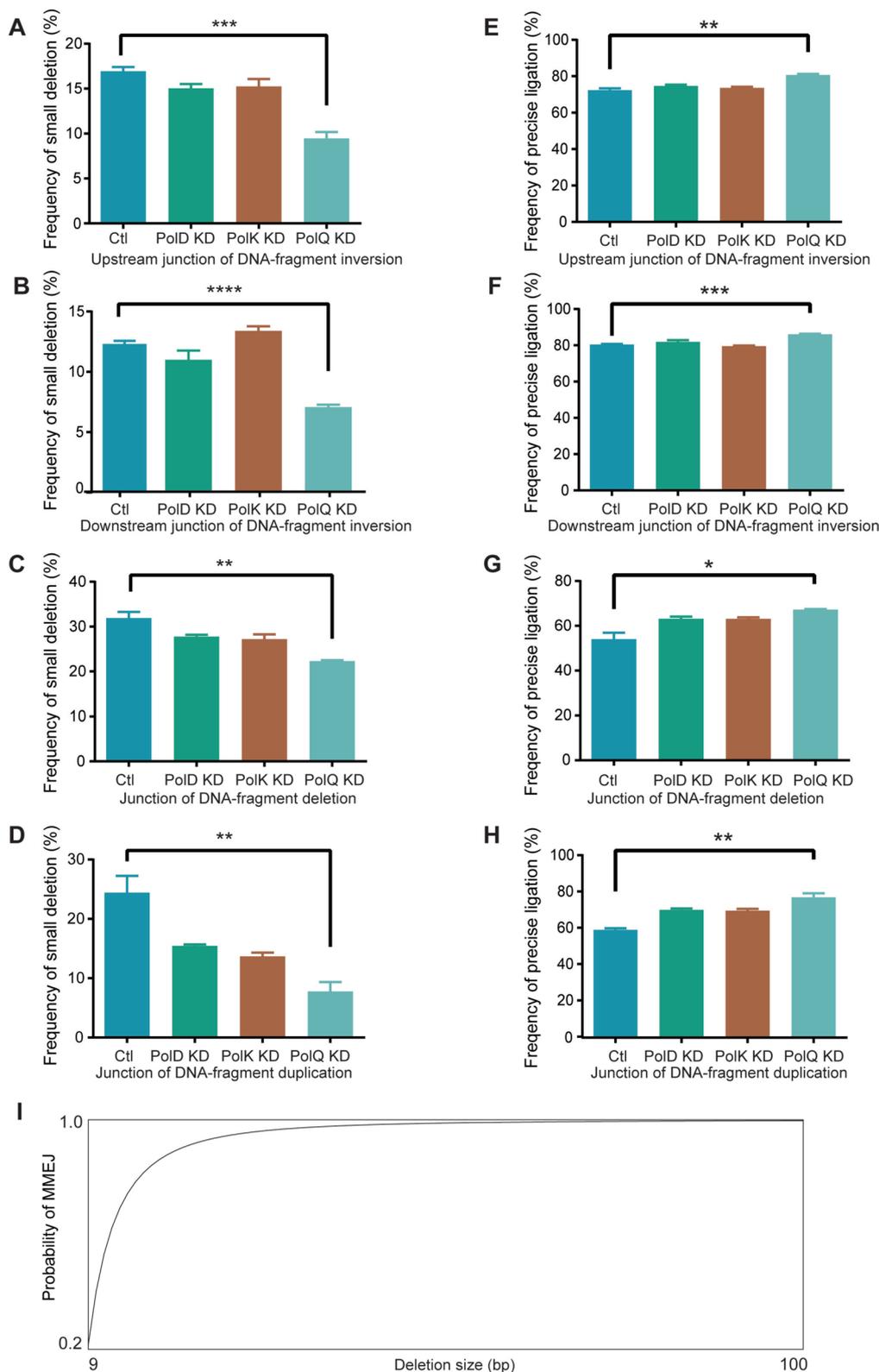


Fig. 2 Polθ promotes small deletions of chromosomal rearrangements. Significant decreases of small deletions and increases of precise ligations at the upstream (**A**, **E**) and downstream (**B**, **F**) junctions of DNA-fragment inversions as well as at the junctions of DNA-fragment deletions (**C**, **G**) and duplications (**D**, **H**) upon knockdown of *PolQ* (encoding Polθ polymerase) (see Additional file 5: Table S3, *n* = 3 replicates, mean ± SEM). **I** Estimation of the probability of MMEJ with increasing deletion size

of DNA-fragment editing in all of these four loci (Additional file 1: Fig. S6). Taken together, these data suggest that Pol θ is essential for generating small deletions during DNA-fragment editing by Cas9 with dual sgRNAs.

Disruption of *CtIP* or *FANCD2*, two DNA repair genes required in the Alt-NHEJ or MMEJ pathway, results in increased precise ligations at junctions of chromosomal rearrangements during DNA-fragment editing [18]. Thus, cNHEJ, which functions in precise ligations of this editing, competes with Alt-NHEJ for repair substrates. Interestingly, *PolQ* knockdown results in a consistent increase of precise ligations at both upstream and downstream junctions of inversions (Fig. 2E, F) as well as junctions of deletions (Fig. 2G) and duplications (Fig. 2H) at the *MeCP2* locus. In addition, *PolQ* knockdown also results in increased precise ligations in the *MAZ*, *PRDM5*, and *PARP1* loci (Additional file 1: Fig. S7). This is in line with the competition of Alt-NHEJ and cNHEJ for repairing Cas9-induced DSB ends during DNA-fragment editing.

Small deletions are editing outcomes of the MMEJ repair pathway upon Cas9 cleavages. The size of MMEJ deletions is determined by the distance from embedded microhomology to the Cas9 cleavage site. Computational analysis revealed that the conservative estimation of the probability of finding at least 2bp microhomology increases rapidly to 99.7% as the deletion size reaches 100bp (Fig. 2I). Accordingly, we analyzed small deletions of less than 100bp in detail below.

Distinct mechanisms for 1bp and >1bp deletions

Recent gene editing using Cas9-Pol I fusion proteins revealed that 1bp and >1bp deletions may be generated differently [32], but the underlying mechanism is unknown. To this end, we separately analyzed 1bp and 2–100bp deletions. Remarkably, upon *PolQ* knockdown, there is a significant increase in 1bp deletions during DNA-fragment editing at the *MAZ* locus (Fig. 3A). In contrast, *PolQ* knockdown results in a significant decrease of 2–100bp deletions at the *MAZ* locus (Fig. 3B, C). We performed *PolQ* knockdown experiments for four additional loci (namely, *MeCP2*, *PARP1*, *PRDM5*, and *YY1*), and observed similar increases of 1bp deletions and decreases of 2–100bp deletions (Fig. 3D–O). These data suggest that Pol θ is essential for the generation of small deletions of 2–100bp, which most likely resulted from the processing by the MMEJ or TMEJ pathways. By contrast, the mechanism for generating 1bp deletions is different, most likely resulting from the processing by the cNHEJ pathway. We also investigated the role of *PolD*, the polymerase for DNA replication, and of *PolK*, the translesion DNA polymerase, the data are not conclusive (Fig. 3).

Pol λ in 1bp deletions

To provide further insight into the molecular mechanisms underlying 1bp and >1bp deletions, we knocked down two members of the DNA polymerase X family, *PolL* and *PolM*, separately or together. Interestingly, we found that *PolL* knockdown results in a significant decrease of 1bp deletion frequency in two cell lines of HEC-1-B and HEK293T (Fig. 4; Additional file 1: Figs. S8 and S9), suggesting an essential role of Pol λ in the 1bp deletions. In contrast, *PolL* knockdown leads to a significant increase of >1bp deletion frequency (Fig. 4; Additional file 1: Fig. S9). This again suggests that 1bp and >1bp deletions are generated by the different repair pathways of cNHEJ and MMEJ, respectively. Further knockdown of *PolM*, another member of the polymerase X family, in combination with the *PolL* knockdown only produced minimal effects (Fig. 4; Additional file 1: Fig. S9). This suggests that Pol μ , in contrast to Pol λ , plays a limited role in the generation of small deletions. As a positive control, we knocked down Ku70/80, two known members of the cNHEJ pathway (Fig. 4; Additional file 1: Figs. S8 and S9). In conjunction with the data from the *PolQ* knockdown (Fig. 3), we conclude that the mechanism of generating 1bp deletions is fundamentally different from that of generating >1bp deletions and that they are generated by cNHEJ and MMEJ pathways, respectively.

Pol λ fill-in of 5' overhangs or cohesive ends of Cas9 staggered cleavages

Recent studies have revealed that Cas9 endonucleolytic cleavage generates staggered DSB ends with 1–3bp 5' overhangs in addition to blunt ends during chromosomal rearrangements induced by Cas9 with dual sgRNAs [18, 19]. Consistent with staggered endonucleolytic Cas9 cleavage, we found that 1–3bp deletions at junctions of chromosomal rearrangements are strongly biased toward the –4, –5, and –6 positions upstream of the PAM site (Additional file 1: Fig. S10). However, the mammalian polymerase(s) responsible for the fill-in of the staggered Cas9 DSB ends is presently unclear.

We thus analyzed 1–3bp templated insertions from fill-in of staggered DSB ends upon polymerase knockdown. However, the available software to characterize Cas9 editing does not take its staggered cleavage into account [33–36]. To this end, we developed a customized computer program to specifically enable this analysis (see reads processing in “Methods” and Additional file 2: Notes S1–S3).

We first analyzed the 1–3bp templated insertions from staggered Cas9 cleavages with sgRNA1 and sgRNA2 at the *MAZ* locus and found that templated 1–3bp

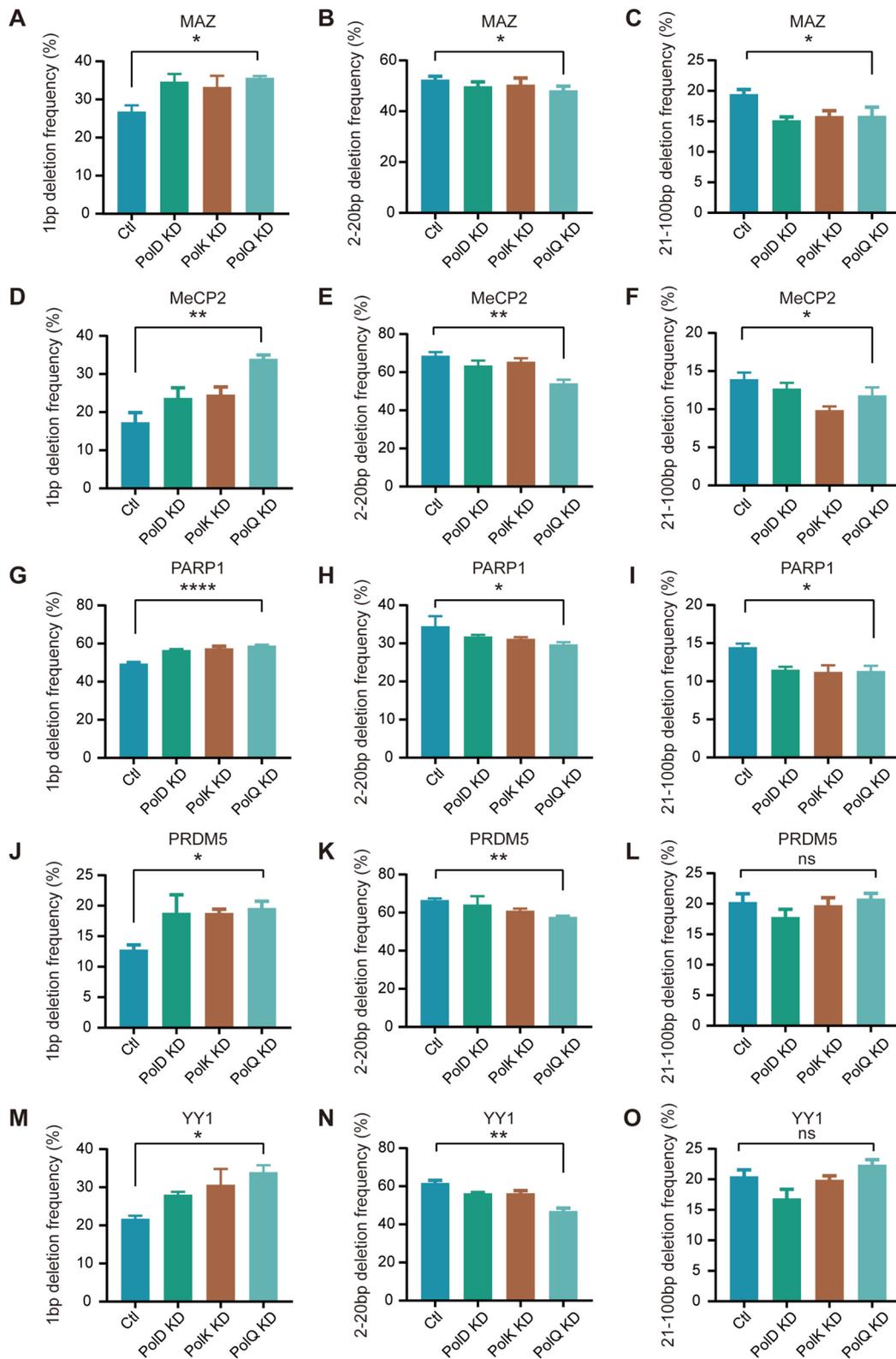


Fig. 3 Mechanistic differences for 1bp and 2–100bp deletions. Significant increases of the frequency of 1bp deletions (A, D, G, J, M) and decreases of the frequency of 2–100bp deletions (B, E, H, K, N and C, F, I, L, O) upon *PolQ* knockdown at the *MAZ* (A–C), *MeCP2* (D–F), *PARP1* (G–I), *PRDM5* (J–L), and *YY1* (M–O) loci (see Additional file 5: Table S3, $n = 3$ replicates, mean \pm SEM). The data for *PolD* or *PolK* knockdown were not conclusive

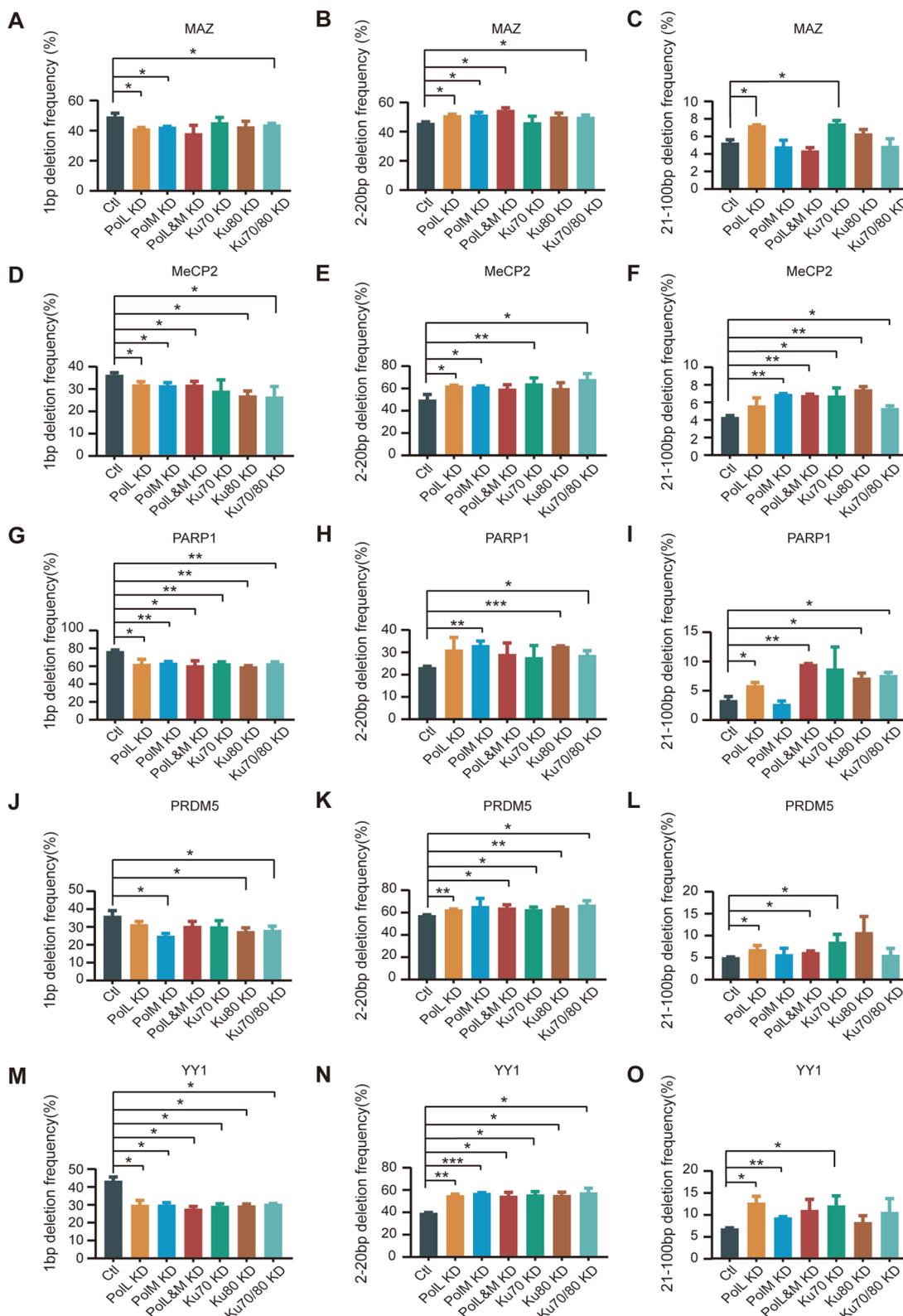


Fig. 4 Pol λ enhances editing outcomes of 1bp deletions and suppresses the generation of >1bp deletions in HEC-1-B cells. There is a significant decrease of frequencies of 1bp deletions but an increase of frequencies of 2–20bp and 21–100bp deletions upon *PolI* knockdown at the MAZ (A–C), *MeCP2* (D–F), *PARP1* (G–I), *PRDM5* (J–L), and *YY1* (M–O) loci. As positive controls, the trends of small deletions upon *Ku70/80*-knockdown are similar as those of *PolI*-knockdown (see Additional file 5: Table S3, $n = 2$ replicates, mean \pm SEM)

insertions generated by both sgRNA1 and sgRNA2 are significantly decreased upon knockdown of *PolL*, and to a lesser extent upon knockdown of *PolM*, for both HEC-1-B and HEK293T cells (Fig. 5A and B; Additional file 1: Fig. S11A and B). However, knockdown of both *PolL* and *PolM* together does not lead to further decreases in templated 1–3bp insertions compared to the knockdown of *PolL* only (Fig. 5A, B). This suggests that Pol λ has a dominant role in the fill-in of staggered Cas9 cleavages, consistent with its role in promoting mutagenesis observed in a recent CRISPR large-scale analysis [37]. We then performed these knockdown experiments for four more loci (*MeCP2*, *PARP1*, *PRDM5*, and *YY1*) and found significant decreases of templated 1–3bp insertions for *MeCP2*, *PARP1*, *PRDM5*, and *YY1* upon knockdown of *PolL* (Fig. 5C–J). Finally, we found no significant difference in frequency of 1–3bp templated insertions upon knockdown of *PolD*, *PolK*, or *PolQ* (Additional file 1: Fig. S12). Altogether, we conclude that Pol λ is the main polymerase responsible for the cellular fill-in of staggered Cas9 endonucleolytic cleavages in vivo.

Discussion

DNA polymerases are thought to counteract nuclease activities of the MRN complex during CRISPR/Cas9 gene editing [32]. Here we have identified for the first time the polymerases involved in CRISPR gene editing events, providing important mechanistic details about various deletions and insertions during gene editing. Overall, we find that Pol λ , and to a lesser extent Pol μ , are the long-sought DNA polymerases that fill-in the staggered DSB ends from programmed Cas9 cleavage. Surprisingly, we find prominent large resections at junctions of chromosomal rearrangements. In addition, we find that Pol λ and Pol θ are very important for these large resections. We also find that mechanisms for 1bp and >1bp deletions are distinct because of opposite dependencies on Pol λ and Pol θ in generating small deletions. Hence, there appears to be fundamentally different pathways enlisted in these Cas9-dependent genome modifications.

Although we find a role of Pol λ in large resections induced by Cas9 with dual sgRNAs (Fig. 1A–F), the underlying mechanism is still not clear. In particular, since Pol λ is a member of family X polymerases, and it participates in cNHEJ via BRCT interaction with the Ku-XRCC4-LIG IV complex [5], it may be able to block large resections. However, how Pol λ precisely blocks large resection remains to be investigated in the future.

We find prominent numbers of sequencing reads indicating complex large resections at junctions of chromosomal rearrangements during CRISPR DNA-fragment editing by Cas9 programmed with dual sgRNAs. Previous studies revealed that Pol θ can mediate the joining of

two 3' overhangs with 2–20bp microhomology exposed after MRN resection [38]. Specifically, Pol θ facilitates microhomology search and stabilizes annealing of microhomologous sequences via its complex activities such as dNTP-dependent 3'-end trimming and template-dependent DNA synthesis [39–43]. Here, we find that large resections are increased and small deletions are decreased upon *PolQ* knockdown, which suggests the essential role of Pol θ in suppressing Cas9-mediated large resections and in inducing TMEJ or MMEJ with small deletions. It is possible that Pol θ perturbation impairs MMEJ which may permit continuous resection into flanking regions thus resulting in large resections.

It is puzzling that there is a significant decrease of 1bp deletion upon perturbation of *PolL*. Here we find that knockdown of *PolL* compromises the fill-in of Cas9-induced cohesive ends and thus the ligation efficiency of the cNHEJ pathway (Fig. 5). Because cNHEJ and MMEJ are competing repair pathways for ligation of DSB ends, compromised cNHEJ should result in shifting to the alternative ligation of the MMEJ pathway from the cNHEJ pathway. Since the cNHEJ pathway also results in 1bp deletion, this explains the decrease of 1bp deletion upon *PolL* knockdown.

It is interesting that knockdown of *PolQ* and *PolL* reveals mechanistic differences of generating 1bp and >1bp deletions. It is consistent with that cNHEJ is responsible for generating 1bp deletion and MMEJ using microhomology sequences embedded in the flanking region results small deletions of 2–100bp. However, the exact locations of microhomology are sequence context dependent. In addition, the cNHEJ repair pathway may generate deletions of 1bp or very few base pairs and deletions of 2bp or 3bp may not be generated by the MMEJ pathway. The exact turning point between cNHEJ and MMEJ may be dependent on sequence or chromatin contexts. Finally, we find that Pol λ is the main polymerase responsible for the fill-in of staggered Cas9 cleavages in vivo. Taken together, our data reveal the crucial role of Pol λ and Pol θ in the repair of Cas9-induced DSBs (Fig. 5K) and should be conducive to the development of controllable CRISPR chromosomal rearrangements.

Conclusions

Using our DNA-fragment editing system to induce chromosomal rearrangements programmed by dual sgRNAs, we found prevalent large resections at Cas9 cleavage junctions. In addition, we found that Pol θ and Pol λ play an inhibitory role in large resections, suggesting that Pol θ and Pol λ are required for ligation of processed DSB ends via the NHEJ (cNHEJ and MMEJ) pathway. This is consistent with the fact that large resections by EXO1 facilitated by MRN complex are essential for the

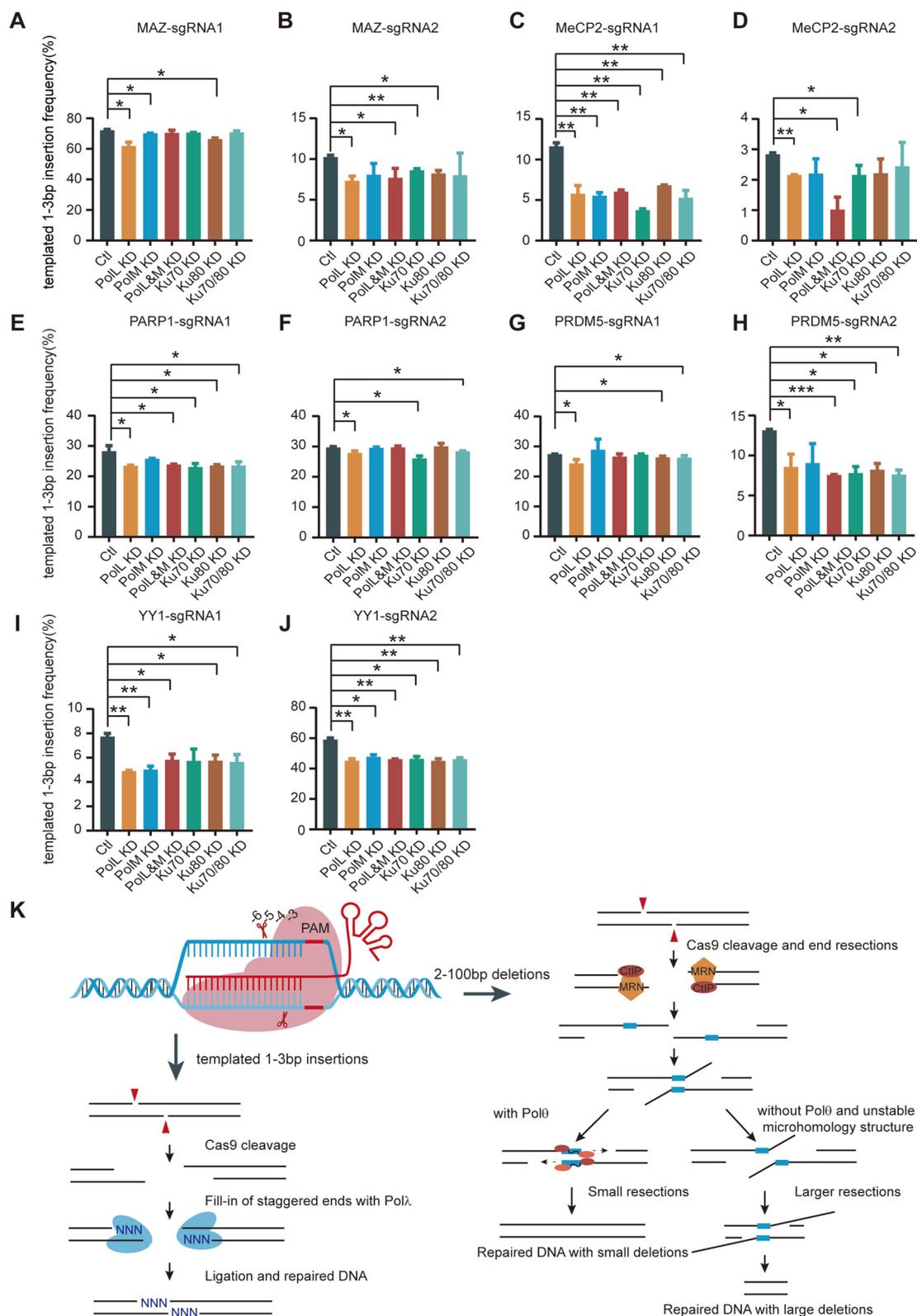


Fig. 5 Fill-in of staggered Cas9 DSB ends by PolI. Significant decreases of templated 1-3bp insertions at Cas9 cleavage junctions programmed with dual sgRNAs at the MAZ (A, B), MeCP2 (C, D), PARP1 (E, F), PRDM5 (G, H), and YY1 (I, J) loci upon PolI knockdown in HEC-1-B cells. As positive controls, the trends of templated insertions upon Ku70/80-knockdown are similar as those of PolI-knockdown (see Additional file 5: Table S3, n = 2 replicates, mean ± SEM). **K** Schematic of the role of PolI and PolII in repairing Cas9-induced DSB ends

HR repair pathway and that HR and NHEJ are competing pathways for repair of DSB ends. Further analyses provide strong evidence for the Pol λ polymerase to fill in the staggered or cohesive ends of Cas9 cleavage, resulting in predictable insertions (predominantly 1bp) for gene editing. Furthermore, our data are consistent with the proposal that Pol λ is associated with the cNHEJ pathway. Finally, we report that 1bp and >1bp deletions are generated by cNHEJ and MMEJ/TMEJ, respectively. These findings have interesting implications not only for mechanistic understanding of the essential roles of DNA polymerases in distinct DSB repair pathways, but also for future development of therapeutic drugs for diseases such as cancers via manipulation of DNA polymerases especially Pol θ .

Methods

Cell culture

The human endometrial carcinoma HEC-1-B cells were cultured in the modified Eagle's medium (MEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin at 37°C in a 5% (v/v) CO₂ incubator.

The human embryonic kidney HEK293T cells were cultured in the Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin at 37°C in a 5% (v/v) CO₂ incubator.

SgRNA design and plasmid construction

We designed sgRNA sequences using CRISPOR, most of which were located at DNase I hypersensitive sites. The plasmid construction was performed as previously described [18]. In brief, the pGL3-U6-sgRNA-PGK-puro was linearized with BsaI (NEB) at 37°C for 1.5h. The resulting plasmid backbone of the linearized vector was run on the 0.8% agarose gel and purified by Monarch DNA Gel Extraction Kit (NEB). The oligos for the inserted sgRNA targeting sequences were synthesized with two overhangs compatible with the linearized vector and complementary to each other. For example, we annealed two pairs of oligos for the double cutting in the MeCP2 locus (MeCP2-1-Fw: 5'-ACCGC ATACA TGGGT CCCC GTCA-3', Rv: 5'-AAACT GACCG GGGAC CCATG TATG-3' for the first cut; MeCP2-2-Fw: 5'-ACCGT TGAAG TGCGA CTCAT GCTG-3', Rv: 5' -AAACC AGCAT GAGTC GCACT TCAA-3' for the second cut). After annealing, the duplexes were ligated with the purified vector with T4 DNA ligase (NEB). The ligation products were transformed into DH5 α bacteria for amplification. All plasmids were confirmed by Sanger sequencing.

DNA polymerase and Ku70/80 knockdown

Knockdown experiments were performed as previously described [18, 27]. Briefly, we designed two sgRNAs for each polymerase and Ku70/80, both targeting coding regions to achieve an efficient knockdown. For example, we constructed two sgRNAs plasmids targeting *PolD* with the oligos listed below (PD1-Fw: 5'-ACCGG TATGG GAAGT AGACC TGGG-3', Rv: 5'-AAACC CCAGG TCTAC TTCCC ATAC-3'; PD2-Fw: 5'-ACCGT GATGA TCACG TAGGG GACG-3', Rv: 5'-AAACC GTCCT CTACG TGATC ATCA-3'). HEC-1-B cells and HEK293T cells were plated in 6-well plates with around 30–40% cell confluence 1 day before transfection. When cells reached more than 80% confluence, they were co-transfected with Cas9 and two sgRNAs plasmids using Lipofectamine 3000 (Thermo Fisher) according to the manufacturer's instructions. After 12 h of culturing with 5% FBS, the culture medium was changed back to the normal condition with 10% FBS. Culturing continued for an additional 12 to 24 h, then cell growth was assayed with the *HPRT1* and *DCK* reporter systems and target site cleavage.

HPRT1 and DCK assay systems

We used the two reporter systems of *HPRT1* or *DCK* to detect large resections of the flanking exons induced by intronic targeting sites by the CRISPR/Cas9 system with dual sgRNAs. For the *HPRT1* assay system, cells with functional *HPRT1* are very sensitive to the 6-TG (6-thioguanine) chemical, and convert it into toxic thioguanosine monophosphate. By contrast, cells with deficient or non-functional *HPRT1* are resistant to this lethal drug and can survive. Cells without *DCK*, a housekeeping gene that plays an important role in DNA synthesis, are not able to accomplish DNA synthesis and will end with cell apoptosis. We designed dual sgRNAs 70–100 bp away from the splicing site within the intron 2 and intron 4 of *HPRT1* and *DCK*, respectively. If there are large resections into the flanking exons of *HPRT1* or *DCK* induced by Cas9 with dual sgRNAs, cells will survive in the *HPRT1* or die in the *DCK* assay systems.

For the *HPRT1* cell growth assay, we plated HEC-1-B cells on the 6-well plate with a cell confluence of 30–40% 1 day before transfection. The number of cells plated in each well was kept consistent. When cell confluence reached 80%, we transfected the cells with plasmids targeting different polymerases to obtain knockdown cell populations. Two days later, cells were transfected again with sgRNAs targeting intron 2 of *HPRT1* in the low serum medium. The medium was changed back to normal serum and continued culturing for one more day. Finally, we selected cells with 6-TG at a concentration of

10 µg/ml for 7 consecutive days. The cells were collected to count the numbers on day 1, day 2, day 4, day 6, and day 7. For the *DCK* cell growth assay, the procedures were similar but without the use of 6-TG, and cells were collected on day 1, day 2, day 3, day 4, and day 5.

Genomic DNA extraction

We extracted genomic DNA from transfected cells to obtain purified DNA templates for further analysis. Briefly, DPBS was used to collect cells when cell confluence reached 70–80%. After centrifugation and discarding the supernatant, the cell pellets were resuspended in the lysis buffer (200 mM NaCl, 10 mM Tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), and 0.2% (wt/vol) SDS) and incubated at 37°C with 750 rpm overnight. The genomic DNA was precipitated with 0.7× volume of isopropyl alcohol after centrifuging at a high speed of 14,000g for 0.5h. Finally, the pellet was washed with 80% ethanol and DNA was dissolved with TE. The genomic DNA can be stored at –20°C for at least half a year.

Preparation of junctional amplicon libraries

Chromosomal rearrangements including fragment deletion, inversion, and duplication can be induced by Cas9 with two sgRNAs. We used high-throughput sequencing to assay various junctional repair outcomes of different chromosomal rearrangements. Since the sizes of the amplified reads within an amplicon library are roughly the same, the bias of PCR amplification efficiency should be negligible. In addition, PCR modeling-based analysis showed that amplified sequences within an amplicon library have similar amplification efficiency (Additional file 1: Fig. S13). Considering the limitation of the read length, the primers used here were all near the junctional site and the length of the final amplified products was less than 290 bp. The experiments were performed as previously described with modifications [19]. Briefly, the primers were designed to be compatible with the Illumina sequencing platform. The PCR conditions were as follows: initial denaturation at 95°C for 3 min, 30 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 15 s, and extension at 72°C for 30 s, followed by a final extension at 72°C for 3 min. The PCR products were purified with the High-Pure PCR Product Purification kit (Roche) and then sequenced by the X ten platform.

Multiplex high-throughput sequencing

For assessing the junctional repair outcomes of each chromosomal rearrangement upon perturbing DNA polymerases and Ku70/80, we constructed libraries using Illumina P5/P7 primers with unique barcodes and indexes. For cost-effective sequencing, we constructed libraries for the same experiment but different replicates

with the same index and barcode, but split samples of replicates into different lanes for efficient sequencing. After library construction, we quantified libraries with Qubit dsDNA HS assay and pooled samples of different experiments with equal mole for efficient sequencing. We performed each polymerase and Ku70/80 knockdown experiment with three replicates. In total, we constructed 829 libraries for high-throughput sequencing.

RNA extraction and RT-PCR

We used the TRIzol reagents (Invitrogen) to obtain the total mRNA for the RT-PCR test. In detail, we used 1ml TRIzol reagent for each well of six-well plates with cell confluence of more than 80%. After homogenization, the samples were incubated for 5 min at room temperature to complete the dissociation of nucleoprotein complexes. Then 200 µl of chloroform was added to the samples, which were then shaken continuously vigorously for 15 s. After shaking, samples were left at room temperature for 5 min, then spun at 12,000g for 15 min at 4°C. After centrifugation, RNA was precipitated with 500 µl isopropyl alcohol. Finally, the pellets were washed with 75% ethanol twice and dissolved with RNase-free water. The RNA can be stored at –20°C for up to a year. For RT-PCR, we used HiScript III RT SuperMix (Vazyme) for reverse transcription according to the manufacturer's instructions followed by PCR with targeting primers. Primer sequences are listed in Additional file 3: Table S1.

Simultaneous sequencing of deletion and inversion junctions

LAM-HTGTS (linear amplification-mediated high-throughput genomic translocation sequencing) was first introduced to detect translocations [31]. We used this method with a few modifications to assay large resections at junctional sites of chromosomal rearrangements induced by CRISPR/Cas9 systems with dual sgRNAs. Briefly, HEC-1-B cells were plated on the 6-well plate with a cell confluence of around 30% 1 day before transfection. When cell confluence reached 70%, we added fresh medium with 5% FBS and performed transfection with Lipofectamine 3000 (Thermo Fisher) according to the manufacturer's instructions. The medium was changed back to the normal medium 24h later and continued culturing for another day to obtain total genomic DNA. We dissolved genomic DNA at a final concentration of 250 ng/µl for sonication. The sonication conditions were 8 trains of 30 s ON and 90 s OFF with low intensity. After sonication, the fragmented DNA was analyzed on 1.5% agarose gel and the ideal size should be 400–600 bp.

To acquire junctional repair outcomes of inversion and deletion simultaneously, we used primers targeting the

left side of bait DSB and performed linear amplification to acquire prey sequences. Briefly, we used 5 µg sonicated DNA as input and amplified the target with Super-Fidelity DNA Polymerase (Vazyme) using 5'-biotinylated primers, which can be captured efficiently by streptavidin beads and ease downstream enrichment. The linear amplification conditions are as follows: initial denaturation at 98°C for 3 min, 85 cycles of denaturation at 98°C for 30 s, annealing at 58°C for 30 s, and extension at 72°C for 90 s, followed by a final extension at 72°C for 5 min. The linear amplification products were enriched with streptavidin beads. To get rid of free primers, we used BW buffer (5mM Tris-HCl, 0.5mM EDTA, 1M NaCl) to wash the beads. Finally, we resuspended the beads with ddH₂O.

Considering various amplification 3' ends, we ligated linear amplification products from the last step with annealed partial double-strand adaptors which have six random nucleotides at the 3' end of one strand. After adaptor ligation, we proceeded with on-bead PCR using Super-Fidelity DNA Polymerase (Vazyme) with P5/P7 adaptors. The PCR conditions were as follows: initial denaturation at 95°C for 5 min; 19 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s, extension at 72°C for 60 s; followed by a final extension at 72°C for 5 min. The PCR products were purified with the High-Pure PCR Product Purification kit (Roche) and the library was sequenced by Illumina X ten platform.

Customized computer program for reads processing

Although Cas9 has been reported to have staggered cleavage activity, up to now, there has not been any alignment software that takes this into account. We developed an alignment program that considers the complexity and diversity of Cas9 cleavage activity. With this program, we can obtain more precise alignments and thus ease downstream analyses.

CRISPR-related insertions and deletions are frequently consecutive nucleotides. Software such as CrisprVariants [36] and AmpliconDIVider [35] maps next-generation sequencing (NGS) reads by traditional aligners like BWA-MEM [44] and NovoAlign (<http://www.novocraft.com>). The software often reports CRISPR-unrelated short non-consecutive insertions and deletions. To solve this problem, Labun et al. developed ampliCan [34] by removing the gap-extension penalty and by modifying other scoring parameters of the Needleman-Wunsch algorithm. Thus, ampliCan tends to report consecutive long deletions and/or insertions. However, the indels reported by ampliCan are not considered to be at the Cas9 cleavage sites. Clement et al. proposed a partial solution to this problem in CRISPResso2 [33] by introducing a reward or bonus at the cleavage site to incentivize indels there.

Nevertheless, this does not completely solve the problem because the Cas9 cleavage may be staggered [18, 19]. In particular, it is not proper to treat the diverse profiles of Cas9 endonucleolytic cleavages as a single position of the -3 nucleotide upstream of PAM.

We develop a new program to solve this conundrum. It aligns each NGS input read to the junctional reference by two levels of optimization. Each NGS input read is separated into three parts before being mapped to the junctional reference. At the lower level, the program searches the optimal alignments of the left and right parts to the junctional reference, and the possibly empty middle part is the unmapped random insertion. At the upper level, the program searches the optimal separation of the three parts. The two levels of optimization are technically integrated into dynamic programming. We permit the left and right parts of each NGS input read to overlap to capture the overhang of the staggered Cas9 cleavage ends. The detailed mathematical design and generalization as well as computational dynamic programming and source code (main.cpp) including its usage (Additional file 2: Notes S1-S4) are available on the GitHub Platform (<https://github.com/ljw20180420/lierlib>).

Calling for insertions and small deletions

Insertions and small deletions are called as previously described with optimizations [18, 19]. We designed PCR primer pairs near the junctional sites (Additional file 3: Table S1) for generating amplicon libraries (PCR products not more than 290bp in size) to assay small indels at junctions of chromosomal rearrangements. Therefore, paired end sequences can be merged for each read. In total, we obtained about 1.6 billion reads for this assay. After demultiplexing raw data of the FASTQ format with the index and barcode, we trimmed the sequences with Cutadapt [45]. For each member of the amplicon library, we then merged the two paired end sequence reads (read1 and read2) using PANDAseq [46]. We divided each junctional repair outcome of chromosomal rearrangements into the four groups of deletions, insertions, indels, and precise ligations, and calculated their respective frequencies.

Large resection analysis

Reads were mapped to the hg19 genome in both strands with our customized program. For reads covering large resections during DNA-fragment deletion, we required that the second segment maps strictly downstream of the first segment and that both are mapped to the forward strand. If the second segment maps to the reverse strand, then this is the case of large resections during DNA-fragment inversion.

Mathematical estimation for MMEJ probability of small deletions

The region of the length n around a Cas9 cleavage has microhomology if and only if the length M of the longest common sequences flanking the cleavage site in the region is larger than a certain artificial threshold L defined by biological experiments. Although it is not easy to obtain the explicit cumulative probability distribution of M , an estimation is available [47] by transforming this microhomology problem within a region of DNA sequences into the problem of tossing a coin for a specific number of times equal to the length of DNA. However, each position of a DNA sequences can have any of the four bases of G,C,A,T in contrast to that each coin-tossing only has either head (obverse) or tail (reverse). The length R of the longest run of heads in the first n tosses of a coin is approximately $\log_2(n)$ [48]. More strictly, $R/\log_2(n)$ converges to 1 almost everywhere as n tends to infinity.

By generalizing, Richard Arratia and Michael S. Waterman prove that $M/\log_{1/p}(n)$ converges to 2 almost everywhere as n tends to infinity [47], where $p = 1/4$ is the probability that two random nucleotides in the corresponding position of the microhomology flanking the cleavage site are the same. For small n , they estimate the probability of $M \leq L$ with an upper bound $(1+p)/(1-p) \left(1 - \left(\frac{2}{n}\right) \log_{1/p}(n)\right)^{-2} p^{2\log_{1/p}(n)-L-1}$ and a lower bound $1 - p^{L-2\log_{(1/p)}(n)+1}$ [47].

We generated the curve of the lower bound estimations of MMEJ probabilities $P(M \geq 2)$ with increasing deletion sizes n for the panel of Fig. 2I with a customized MATLAB script (Additional file 2: Note S5).

Statistical analysis

All high-throughput sequencing libraries are constructed with at least two replicates. The significance tests are performed using the GraphPad software with two-tailed t -tests, with one, two, three, and four asterisks indicating P -values less than 0.05, 0.01, 0.001, and 0.0001, respectively.

Abbreviations

6-TG	6-thioguanine
alt-NHEJ	Alternative NHEJ
cNHEJ	Canonical NHEJ
Cas9	CRISPR-associated protein 9
CRISPR	Clustered regularly interspaced short palindromic repeats
DCK	Deoxycytidine kinase
DSB	Double-strand break
HPRT1	Hypoxanthine phosphoribosyltransferase 1
HR	Homologous recombination
LAM-HTGTS	Linear amplification-mediated high-throughput genomic translocation sequencing
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining

MMEJ	Microhomology-mediated end joining
MRN	MRE11-RAD50-NBS1 complex
sgRNA	Single-guide RNA
TMEJ	Theta-mediated end joining

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-023-01784-y>.

Additional file 1: Fig. S1. DNA-fragment editing by Cas9 with dual sgRNAs. **Fig. S2.** *HPRT1* and *DCK* reporter assay systems for Cas9-induced large resections. **Fig. S3.** Quantitative RT-PCR at two time points in HEK-1-B cells. **Fig. S4.** Additional examples of large resections by DNA sequencing during DNA-fragment editing. **Fig. S5.** High-throughput NGS of junctional sequences of DNA-fragment editing. **Fig. S6.** Significant decreases in the frequency of small deletions. **Fig. S7.** Significant increases in the frequency of precise ligations. **Fig. S8.** Quantitative RT-PCR at two time points in HEK293T cells. **Fig. S9.** Pol λ enhances editing outcomes of 1bp deletions and suppresses the generation of >1bp deletions in HEK293T cells. **Fig. S10.** Biased deletion of nucleotides at junctional sites of chromosomal rearrangements confirms the staggered or cohesive Cas9 cleavages. **Fig. S11.** Fill-in of cohesive Cas9 DSB ends by Pol λ in HEK293T cells. **Fig. S12.** Pol δ , Pol κ , and Pol θ are not engaged in the fill-in of cohesive Cas9 cleavage ends. **Fig. S13.** PCR modelling-based analysis by the pcrEfficiency software.

Additional file 2: Note S1. Sequence alignments designed specifically for CRISPR cleavages. **Note S2.** Dynamic programming of alignment algorithm for CRISPR cleavages. **Note S3.** Generalization of the alignment algorithm. **Note S4.** The basic usage of customized computer program and its source code in C++. **Note S5.** MATLAB script.

Additional file 3: Table S1. Oligonucleotide sequences used in this study.

Additional file 4: Table S2. SRA metadata.

Additional file 5: Table S3. Supporting data values.

Acknowledgements

We thank Prof. Dan Czajkowsky for great improvements on the manuscript and all members of our laboratory for helpful discussion.

Conflict of interest statement

None declared.

Authors' contributions

M.M.M. and X.S. performed and Q.W. designed the experiments. J.L. developed the alignment program. X.S. and J.L. analyzed and interpreted data. Q.W. supervised the project. All authors read and approved the final manuscript.

Funding

National Key R&D Program of China (2022YFC3400200), the National Natural Science Foundation of China (31630039), and the Science and Technology Commission of Shanghai Municipality (21DZ2210200).

Availability of data and materials

All data generated or analyzed during this study are included in this published article, its supplementary information files and publicly available repositories. High-throughput sequencing files have been submitted into to SRA with accession number SRP405576 [49]. SRA metadata are in Additional file 4: Table S2. The computer codes used to generate the results are available at <https://github.com/Jw20180420/lierlib>. Please cite the Zenodo doi: <https://doi.org/10.5281/zenodo.10205238> when using the raw code [50]. Supporting data values for figures are in Additional file 5: Table S3.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 16 January 2023 Accepted: 27 November 2023

Published online: 08 December 2023

References

- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337(6096):816–21.
- Yeh CD, Richardson CD, Corn JE. Advances in genome editing through control of DNA repair pathways. *Nat Cell Biol*. 2019;21(12):1468–78.
- Anzalone AV, Koblan LW, Liu DR. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol*. 2020;38(7):824–44.
- Nambiar TS, Baudrier L, Billon P, Ciccio A. CRISPR-based genome editing through the lens of DNA repair. *Mol Cell*. 2022;82(2):348–88.
- Yang W, Gao Y. Translesion and repair DNA polymerases: diverse structure and mechanism. *Annu Rev Biochem*. 2018;87:239–61.
- Stinson BM, Loparo JJ. Repair of DNA double-strand breaks by the nonhomologous end joining pathway. *Annu Rev Biochem*. 2021;90:137–64.
- Cisneros-Aguirre M, Lopezcolorado FW, Tsai LJ, Bhargava R, Stark JM. The importance of DNAPKcs for blunt DNA end joining is magnified when XLF is weakened. *Nat Commun*. 2022;13(1):1–17.
- Kosicki M, Allen F, Steward F, Tomberg K, Pan Y, Bradley A. Cas9-induced large deletions and small indels are controlled in a convergent fashion. *Nat Commun*. 2022;13(1):1–11.
- Porteus MH, Pavel-Dinu M, Pai SY. A curative DNA code for hematopoietic defects: novel cell therapies for monogenic diseases of the blood and immune system. *Hematol Oncol Clin North Am*. 2022;36(4):647–65.
- Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, et al. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell*. 2015;162(4):900–10.
- Huang H, Wu Q. CRISPR double cutting through the labyrinthine architecture of 3D genomes. *J Genet Genomics = Yi chuan xue bao*. 2016;43(5):273–88.
- Wu Q, Shou J. Toward precise CRISPR DNA fragment editing and predictable 3D genome engineering. *J Mol Cell Biol*. 2020;12(11):828–56.
- Wang H, Han M, Qi LS. Engineering 3D genome organization. *Nat Rev Genet*. 2021;22(6):343–60.
- Canver MC, Bauer DE, Dass A, Yien YY, Chung J, Masuda T, Maeda T, Paw BH, Orkin SH. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J Biol Chem*. 2014;289(31):21312–24.
- Kraft K, Geuer S, Will AJ, Chan WL, Paliou C, Borschiwer M, Harabula I, Wittler L, Franke M, Ibrahim DM, et al. Deletions, inversions, duplications: engineering of structural variants using CRISPR/Cas in mice. *Cell Rep*. 2015;10:833–9.
- Li J, Shou J, Guo Y, Tang Y, Wu Y, Jia Z, Zhai Y, Chen Z, Xu Q, Wu Q. Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. *J Mol Cell Biol*. 2015;7(4):284–98.
- Shin HY, Wang C, Lee HK, Yoo KH, Zeng X, Kuhns T, Yang CM, Mohr T, Liu C, Hennighausen L. CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat Commun*. 2017;8:1–10.
- Shou J, Li J, Liu Y, Wu Q. Precise and predictable CRISPR chromosomal rearrangements reveal principles of Cas9-mediated nucleotide insertion. *Mol Cell*. 2018;71:498–509.
- Shi X, Shou J, Mehryar MM, Li J, Wang L, Zhang M, Huang H, Sun X, Wu Q. Cas9 has no exonuclease activity resulting in staggered cleavage with overhangs and predictable di- and tri-nucleotide CRISPR insertions without template donor. *Cell Discov*. 2019;5(1):1–4.
- Dahiya R, Hu Q, Ly P. Mechanistic origins of diverse genome rearrangements in cancer. *Semin Cell Dev Biol*. 2022;123:100–9.
- Aguilera A, Garcia-Muse T. Causes of genome instability. *Annu Rev Genet*. 2013;47:1–32.
- Tonegawa S. Somatic generation of antibody diversity. *Nature*. 1983;302(5909):575–81.
- Bodai Z, Bishop AL, Gantz VM, Komor AC. Targeting double-strand break indel byproducts with secondary guide RNAs improves Cas9 HDR-mediated genome editing efficiencies. *Nat Commun*. 2022;13(1):1–15.
- Shen MW, Arbab M, Hsu JY, Worstell D, Culbertson SJ, Krabbe O, Cassa CA, Liu DR, Gifford DK, Sherwood RI. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature*. 2018;563:646–51.
- Taheri-Ghahfarokhi A, Taylor BJM, Nitsch R, Lundin A, Cavallo AL, Madeyski-Bengtson K, Karlsson F, Clausen M, Hicks R, Mayr LM, et al. Decoding non-random mutational signatures at Cas9 targeted sites. *Nucleic Acids Res*. 2018;46(16):8417–34.
- Allen F, Crepaldi L, Alsinet C, Strong AJ, Kleshchevnikov V, De Angeli P, Palenikova P, Khodak A, Kiselev V, Kosicki M, et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol*. 2019;37(1):64–72.
- Chen W, McKenna A, Schreiber J, Haeussler M, Yin Y, Agarwal V, Noble WS, Shendure J. Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Res*. 2019;47(15):7989–8003.
- Chakrabarti AM, Henser-Brownhill T, Monserrat J, Poetsch AR, Luscombe NM, Scaffidi P. Target-specific precision of CRISPR-mediated genome editing. *Mol Cell*. 2019;73(4):699–713.
- Leenay RT, Aghazadeh A, Hiatt J, Tse D, Roth TL, Apathy R, Shifrut E, Hultquist JF, Krogan N, Wu Z, et al. Large dataset enables prediction of repair after CRISPR-Cas9 editing in primary T cells. *Nat Biotechnol*. 2019;37(9):1034–7.
- Bennett EP, Petersen BL, Johansen IE, Niu Y, Yang Z, Chamberlain CA, Met O, Wandall HH, Frodin M. INDEL detection, the “Achilles heel” of precise genome editing: a survey of methods for accurate profiling of gene editing induced indels. *Nucleic Acids Res*. 2020;48(21):11958–81.
- Hu J, Meyers RM, Dong J, Panchakshari RA, Alt FW, Frock RL. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat Protoc*. 2016;11(5):853–71.
- Yoo KW, Yadav MK, Song Q, Atala A, Lu B. Targeting DNA polymerase to DNA double-strand breaks reduces DNA deletion size and increases templated insertions generated by CRISPR/Cas9. *Nucleic Acids Res*. 2022;50(7):3944–57.
- Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, Cole MA, Liu DR, Joung JK, Bauer DE, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol*. 2019;37(3):224–6.
- Labun K, Guo X, Chavez A, Church G, Gagnon JA, Valen E. Accurate analysis of genuine CRISPR editing events with ampliCan. *Genome Res*. 2019;29(5):843–7.
- Varshney GK, Pei W, LaFave MC, Idol J, Xu L, Gallardo V, Carrington B, Bishop K, Jones M, Li M, et al. High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res*. 2015;25(7):1030–42.
- Lindsay H, Burger A, Biyong B, Felker A, Hess C, Zaugg J, Chiavacci E, Anders C, Jinek M, Mosimann C, et al. CrispRvariants charts the mutation spectrum of genome engineering experiments. *Nat Biotechnol*. 2016;34(7):701–2.
- Hussmann JA, Ling J, Ravisankar P, Yan J, Cirincione A, Xu A, Simpson D, Yang D, Bothmer A, Cotta-Ramusino C, et al. Mapping the genetic landscape of DNA double-strand break repair. *Cell*. 2021;184:5653–69.
- Seol JH, Shim EY, Lee SE. Microhomology-mediated end joining: Good, bad and ugly. *Mutat Res*. 2018;809:81–7.
- van Schendel R, Roerink SF, Portegijs V, van den Heuvel S, Tijsterman M. Polymerase Theta is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis. *Nat Commun*. 2015;6:7394.
- Wyatt DW, Feng W, Conlin MP, Yousefzadeh MJ, Roberts SA, Mieczkowski P, Wood RD, Gupta GP, Ramsden DA. Essential roles for polymerase theta-mediated end joining in the repair of chromosome breaks. *Mol Cell*. 2016;63(4):662–73.
- Schimmel J, Kool H, van Schendel R, Tijsterman M. Mutational signatures of non-homologous and polymerase theta-mediated end-joining in embryonic stem cells. *EMBO J*. 2017;36(24):3634–49.

42. Saito S, Maeda R, Adachi N. Dual loss of human POLQ and LIG4 abolishes random integration. *Nat Commun.* 2017;8:16112.
43. Zahn KE, Jensen RB. Polymerase theta coordinates multiple intrinsic enzymatic activities during DNA repair. *Genes.* 2021;12(9):1310.
44. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013; <https://doi.org/10.48550/arXiv.1303.3997>.
45. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10–2.
46. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PAN-DAseq: paired-end assembler for Illumina sequences. *BMC Bioinformatics.* 2012;13:1–7.
47. Arratia R, Waterman MS. An Erdős-Rényi law with shifts. *Adv Math.* 1985;55(1):13–23.
48. Erdős P, Rényi A. On a new law of large numbers. *J Anal Math.* 1970;22:103–11.
49. Mehryar MM, Shi X, Li J, Wu Q. DNA polymerases in precise predictable CRISPR/Cas9-mediated chromosomal rearrangements. *Sequence Read Archive (SRA)*. 2023. <https://www.ncbi.nlm.nih.gov/sra/?term=SRP405576>.
50. Mehryar MM, Shi X, Li J, Wu Q. DNA polymerases in precise predictable CRISPR/Cas9-mediated chromosomal rearrangements. 2023. Zenodo. <https://doi.org/10.5281/zenodo.10205238>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

