Supplementary Data

DNA Polymerase Beta Participates in DNA End-joining

Sreerupa Ray, Gregory Breuer, Michelle DeVeaux, Daniel Zelterman, Ranjit Bindra, Joann B. Sweasy

Supplementary Figures

Figure S1

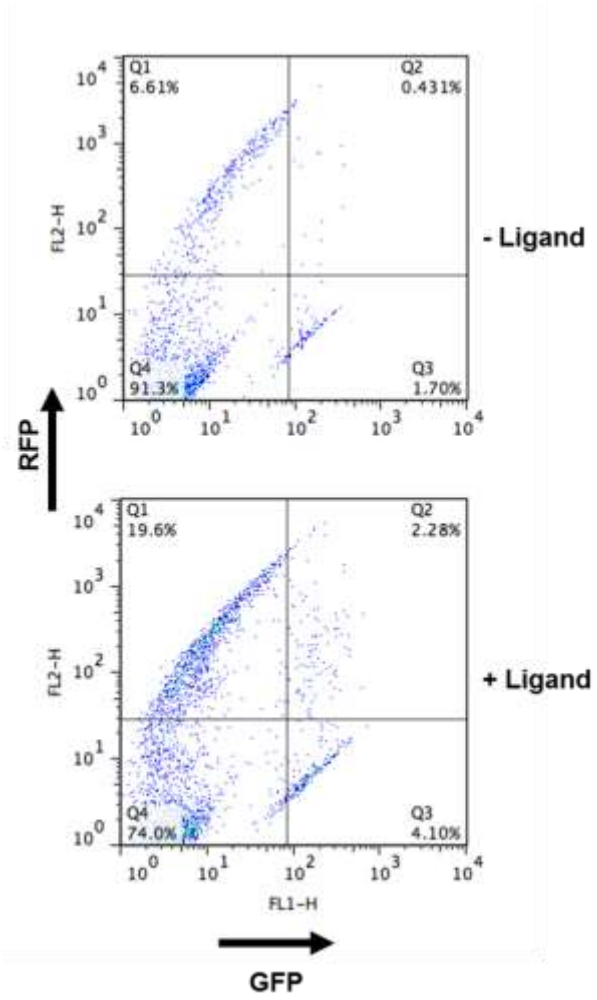

Figure S1. Flow cytometry plots showing RFP and GFP positive cells in FL2 and FL1. Representative flow cytometry plot for U2OS EJ-DRs cells showing low levels of RFP+ and GFP+ cells in the absence of Shield1 and triamcinolone (TA) ligands (top panel), which is highly inducible after a 96-h exposure to the ligands (bottom panel).

Figure S2

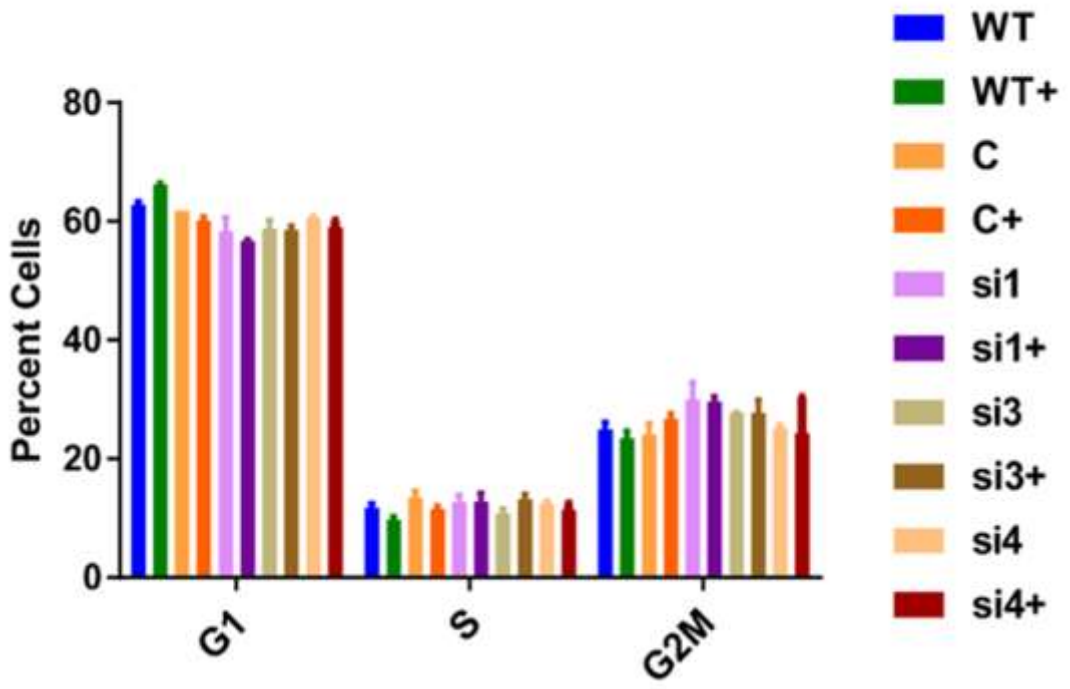

Figure S2. No change cell cycle profile for Pol β-depleted cells. Overall cell cycle profile of the U2OS EJ-DR cells in presence of Pol β siRNA. Quantification of the flow cytometry plots for U2OS EJ-DR cells stained with propidium iodide. WT, C, si1, si3 and si4 indicate cells without ligands. WT+, C+, si1+, si3+ and si4+ indicates cells with ligands to induce DSBs.

Figure S3

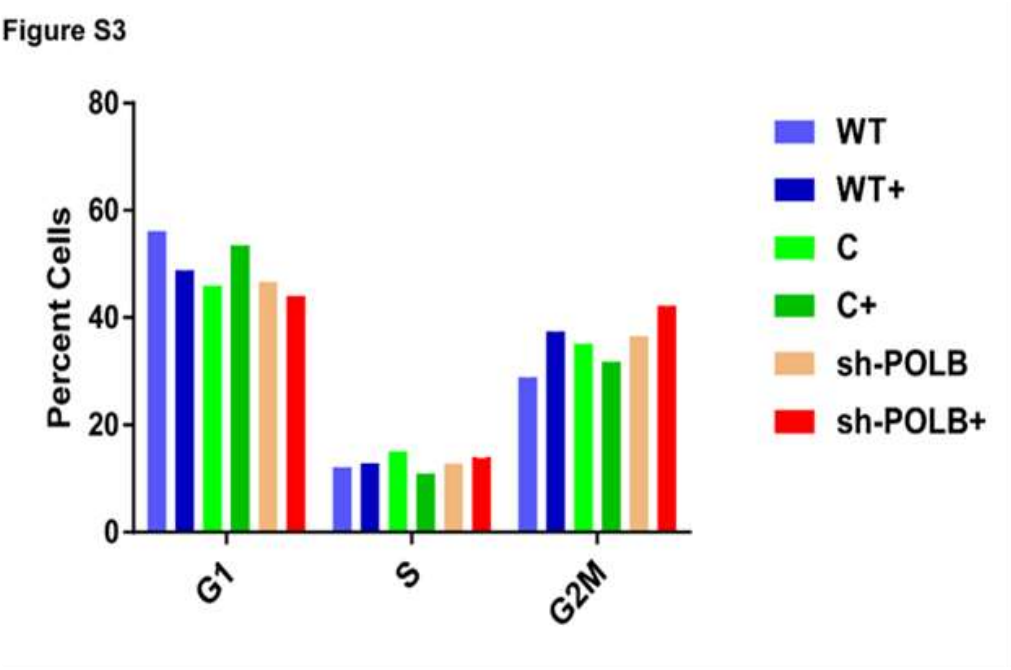

Figure S3. No change in overall cell cycle profile of U2OS EJ-DR cells in presence of Pol β shRNA. WT, C and sh-POLB indicates levels of cells without ligands. WT+, C+ and sh-POLB+ indicate cells with the ligands to induce DSBs.

Figure S4

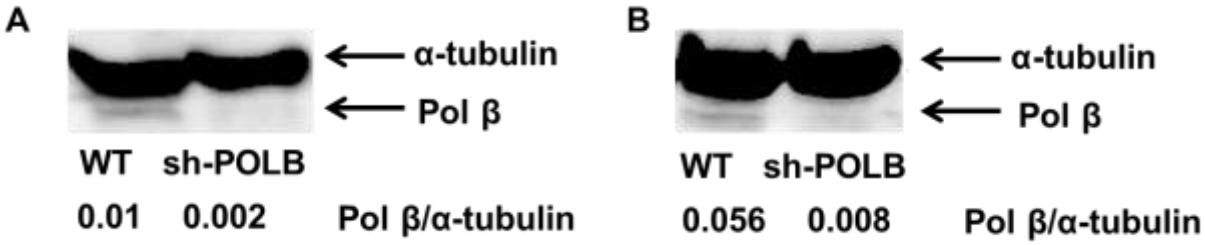

Figure S4. Western blot showing depletion of Pol β using shRNA targeting of the *POLB* gene. A. Depletion of Pol β in MCF7 cells. **B.** Depletion of Pol β in U2OS cells. α-tubulin is used as a loading control. Pol β/α-tubulin indicates the ratio of the intensity of the Pol β band over the α-tubulin band and is a reflection of the relative amount of Pol β that is depleted from the cells.

Figure S5

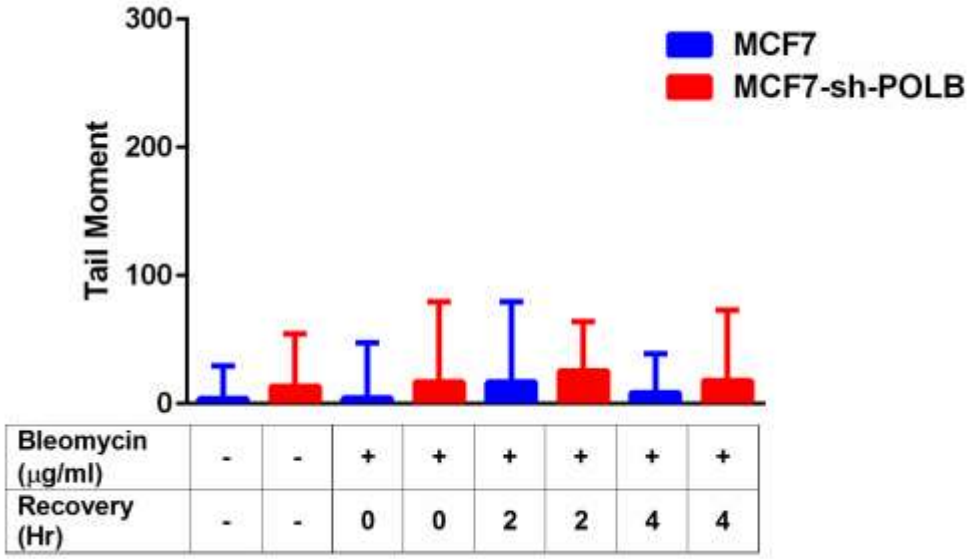

Figure S5. Few SSBs are observed in Pol β-depleted cells. MCF7 and MCF7-sh-POLB cells were treated with 50 μg/ml bleomycin for 1 h and allowed to recover for 0, 2 and 4 h. SSBs were analyzed by the alkaline comet assay. The tail moment is plotted on the Y-axis.

Figure S6.

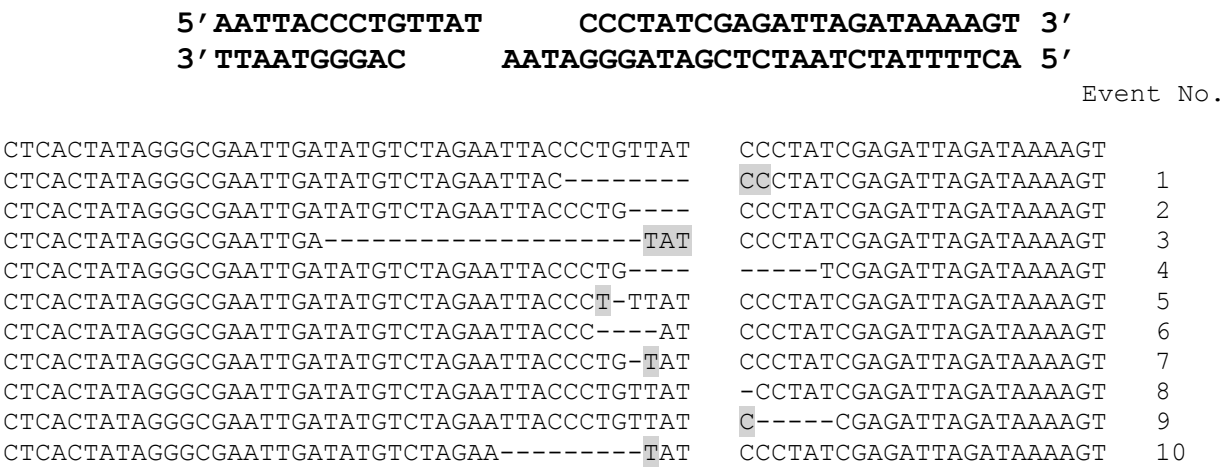

Figure S6. Small deletion events observed more often in the control cells. The sequence in bold indicates the I-SceI recognition site and surrounding sequence after induction of the DSB by I-SceI, which results in 4 base overhangs. The dashes indicate the bases deleted. The microhomologies at the junction are highlighted in gray boxes.

Figure S7.

Figure S7. Small deletion events observed more often in the siPOLB cells. The sequence in bold indicates the I-SceI recognition site and surrounding sequence after induction of the DSB by I-SceI, which results in 4 base overhangs. The dashes indicate the bases deleted. The microhomologies at the junction are highlighted in gray boxes.

Figure S8A. Direct DNA end-joins for which no μH could be identified.

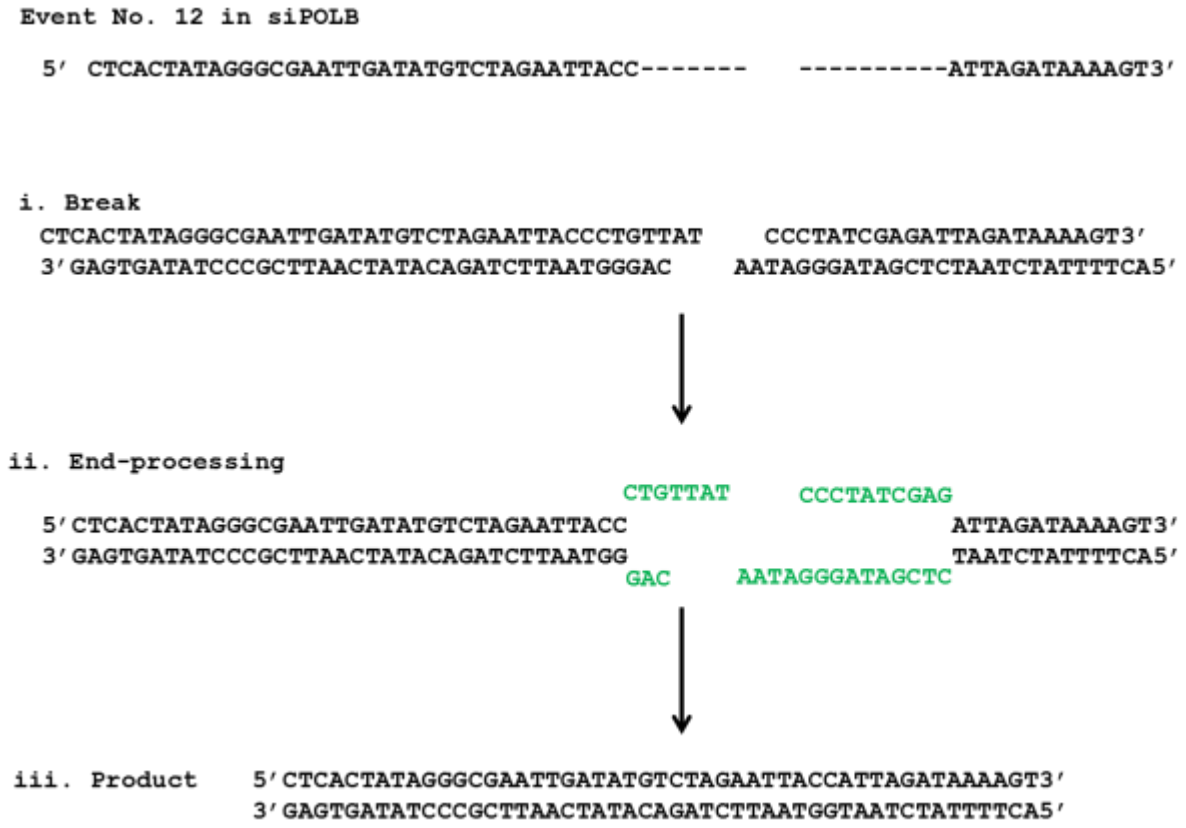

Figure S8B. A DNA polymerase μ (Pol μ) signature in which DNA synthesis is directed across strand breaks.

Figure S8C. A signature of μH-mediated end-joining involving 3' and 5'end-processing activity and gap filling.

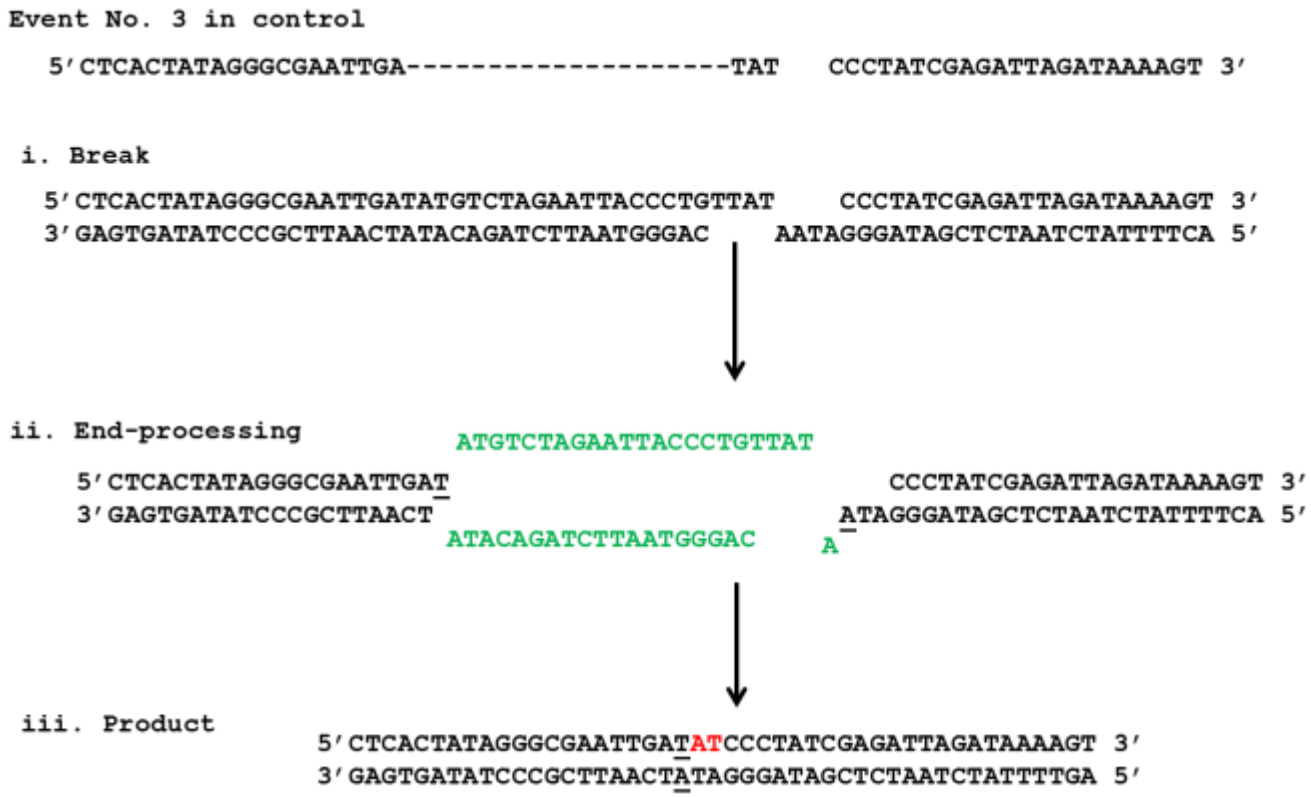

Figure S8D. A signature of μH-mediated end-joining involving both 3'and 5' end-processing activity.

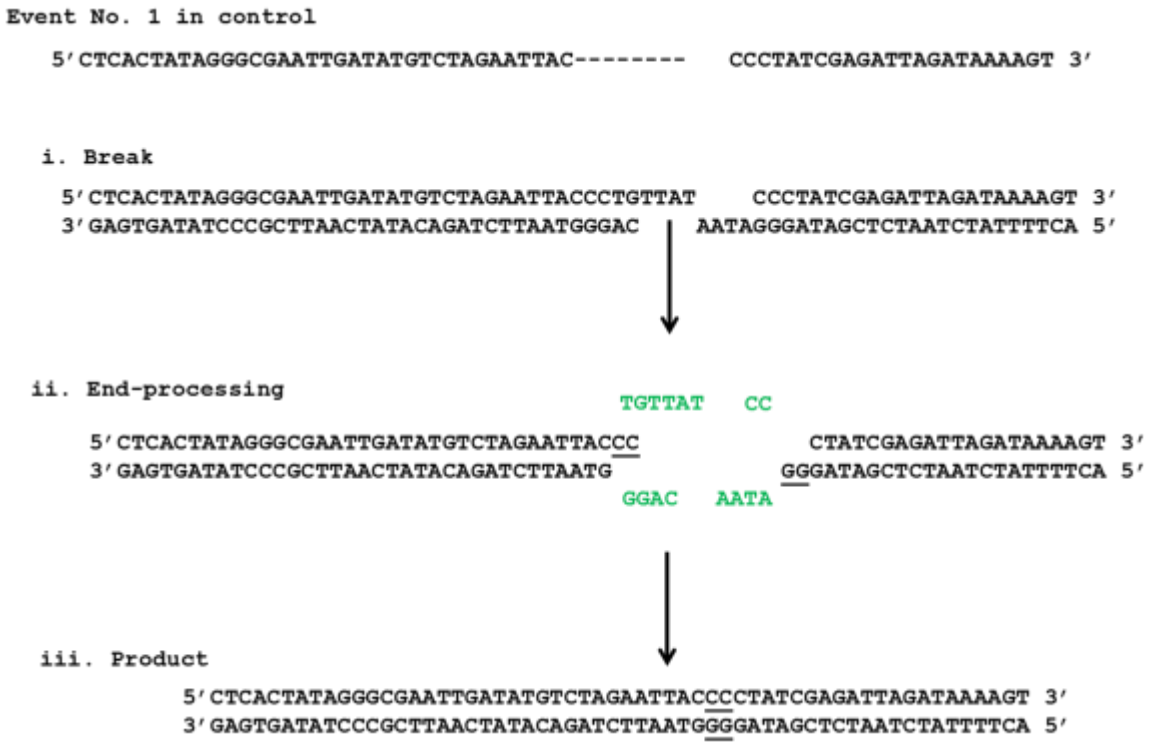

Figure S8E. A signature of μH-mediated end-joining involving only 3'end-processing activity and gap filling.

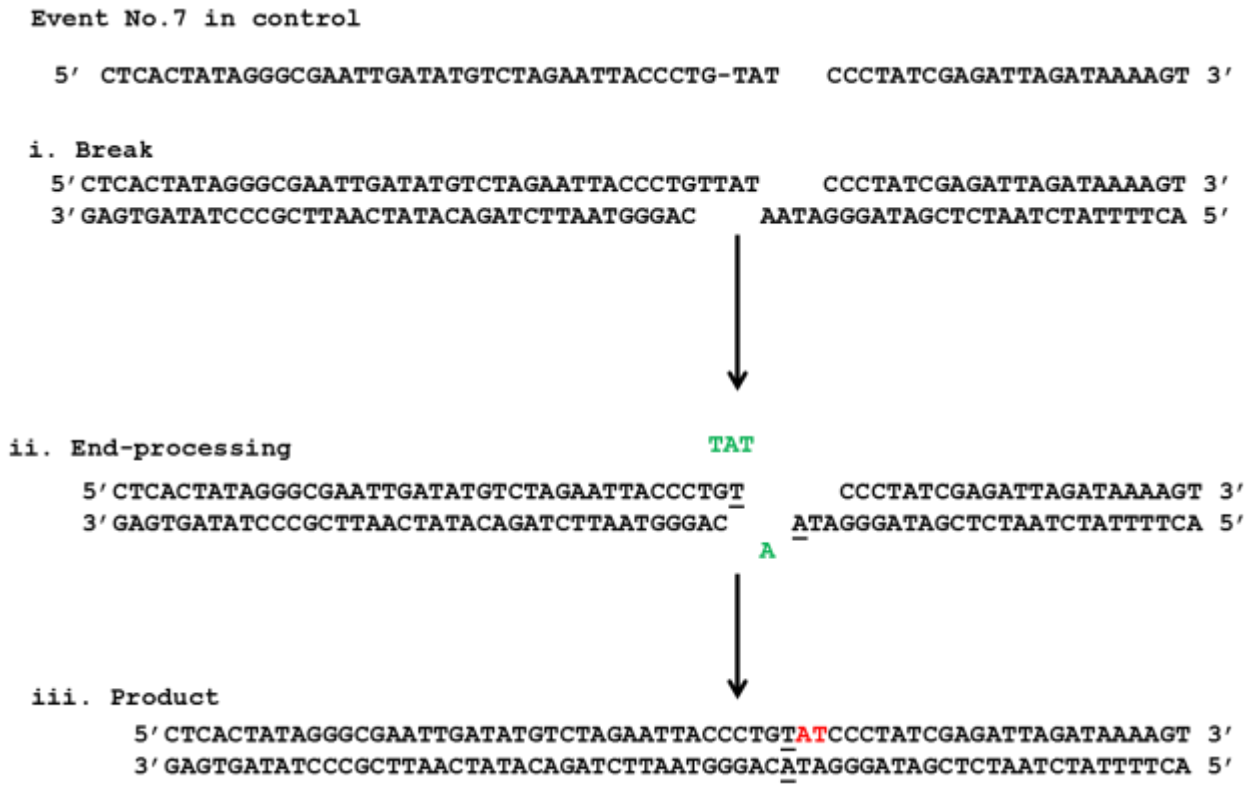

Figure S8F. A signature of μH-mediated end-joining involving only 3’ end-processing activity.

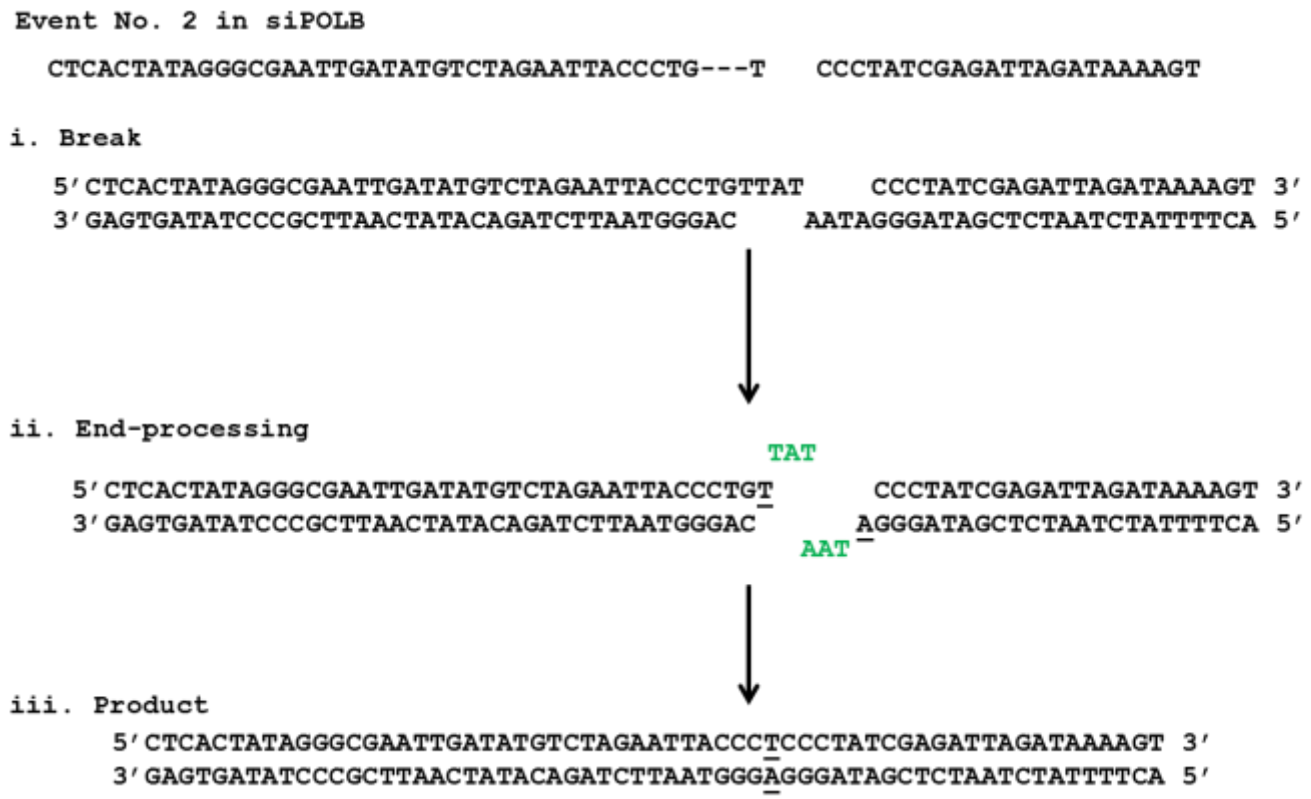

Figure S8. Modeling of the deletions uncovers at least six deletion signatures
A. Model displaying direct DNA end joining for which no μH could be identified. The event number 12 (Figure S7) in siPOLB is an example of direct DNA end-joining. The top most sequence indicates the deletion event. i. The break indicates the endonuclease activity of I-SceI results in 4 base overhangs. ii. The end-processing indicates the bases removed and are highlighted in green. iii. The final product after the bases are deleted. **B.** Model for a DNA polymerase μ (Pol μ) signature in which DNA synthesis is directed across strand breaks. The event number 14 (Figure S7) in siPOLB is an example of a Pol μ signature. The top most sequence indicates the deletion event. i. The break indicates the endonuclease activity of I-SceI results in 4 base overhangs. ii. The end-processing indicates the bases removed and are highlighted in green. iii. The final product after the bases are deleted. The base fill is highlighted in red. **C.** Model displaying a signature of μH-mediated end-joining involving either end-resection including 3' and 5'end-processing activity and gap filling. The event number 3 (Figure S6) in the control is an example μH-mediated end-joining involving end-processing and gap filling. The top most sequence indicates the deletion event. i. The break indicates the endonuclease activity of I-SceI results in 4 base overhangs. ii. The end-processing indicates the bases removed and are highlighted in green. The μH bases are underlined. iii. The final product after the bases are deleted and gap filling occurs (red). **D.** Model displaying a signature of μH-mediated end-joining involving 3' and 5' end-processing activity. Event number 1 (Figure S6) in the control is an example of μH-mediated end-joining and end-processing. The top most sequence indicates the deletion event. i. The break indicates the endonuclease activity of I-SceI results in 4 base overhangs. ii. The end-processing indicates the bases removed and are highlighted in green. The μH bases are underlined. iii. The final product after the bases are deleted. **E.** Model displaying a signature of μH-mediated end-joining involving only 3' end-processing activity and gap filling. Event number 7 (Figure S6) in the control is an example μH-mediated end-joining and gap filling. The top most sequence indicates the deletion event. i. The break indicates the endonuclease activity of I-SceI results in 4 base overhangs. ii. The end-processing indicates the bases removed and are highlighted in green. The μH bases are underlined. iii. The final product after the bases are deleted. The base fill is highlighted in red. **F.** Model displaying a signature of μH-mediated end-joining involving only 3' end-processing activity. Event number 2 (Figure S7) in siPOLB is an example μH-mediated end-joining. The top most sequence indicates the deletion event. i. The break indicates the endonuclease activity of I-SceI results in 4 base overhangs. ii. The end-processing indicates the bases removed and are highlighted in green. The μH bases are underlined. iii. The final product after the bases are deleted.

Figure S9.

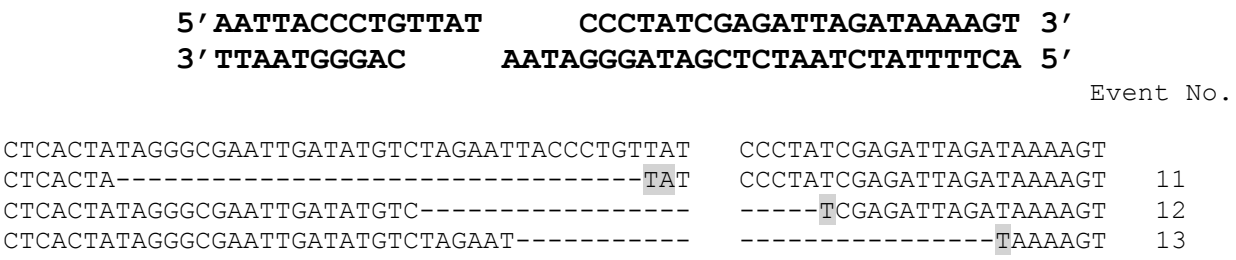

Figure S9. Large deletion events observed more often in the control cells.
The sequence in bold indicates the I-SceI and surrounding sequence after induction of the DSB by I-SceI, which results in 4 base overhangs. The dashes indicate the bases deleted. The microhomologies at the junction are highlighted in gray boxes.

Figure S10.

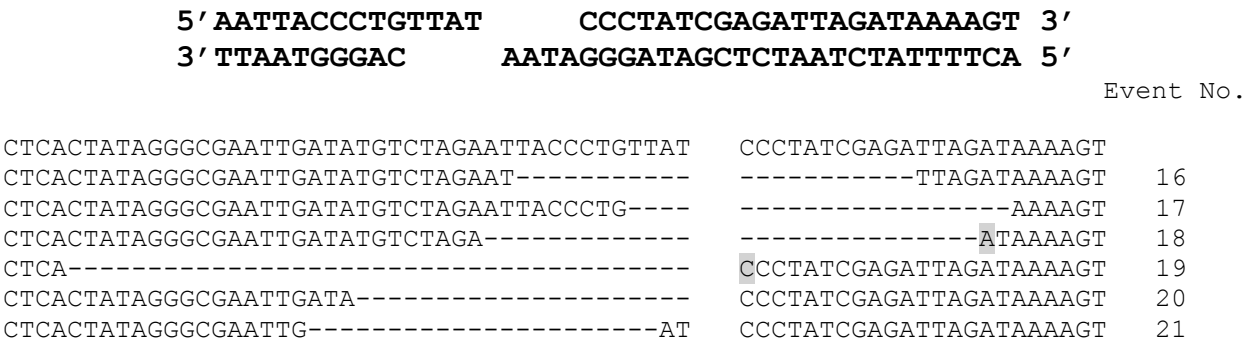

Figure S10. Large deletion events observed more often in the siPOLB cells.
The sequence in bold indicates the I-SceI recognition site and surrounding sequence after induction of the DSB by I-SceI, which results in 4 base overhangs. The dashes indicate the bases deleted. The microhomologies at the junction are highlighted in gray boxes.

Figure S11.

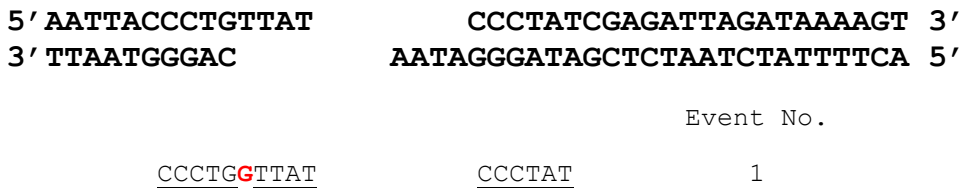

Figure S11. Summary of the insertion events in the control cells.
The sequence in bold indicates the I-SceI and surrounding sequence after induction of the DSB by I-SceI, which results in 4 base overhangs. The parental sequence is underlined and insertion is in red.

Figure S12.

Figure S12. Summary of the insertion events in the siPOLB cells.
The sequence in bold indicates the I-SceI and surrounding sequence after induction of the DSB by I-SceI, which results in 4 base overhangs. The underlined bases indicate inserted bases. The parental sequence is underlined and insertion is in red.

Figure S13A. Templated insertions.

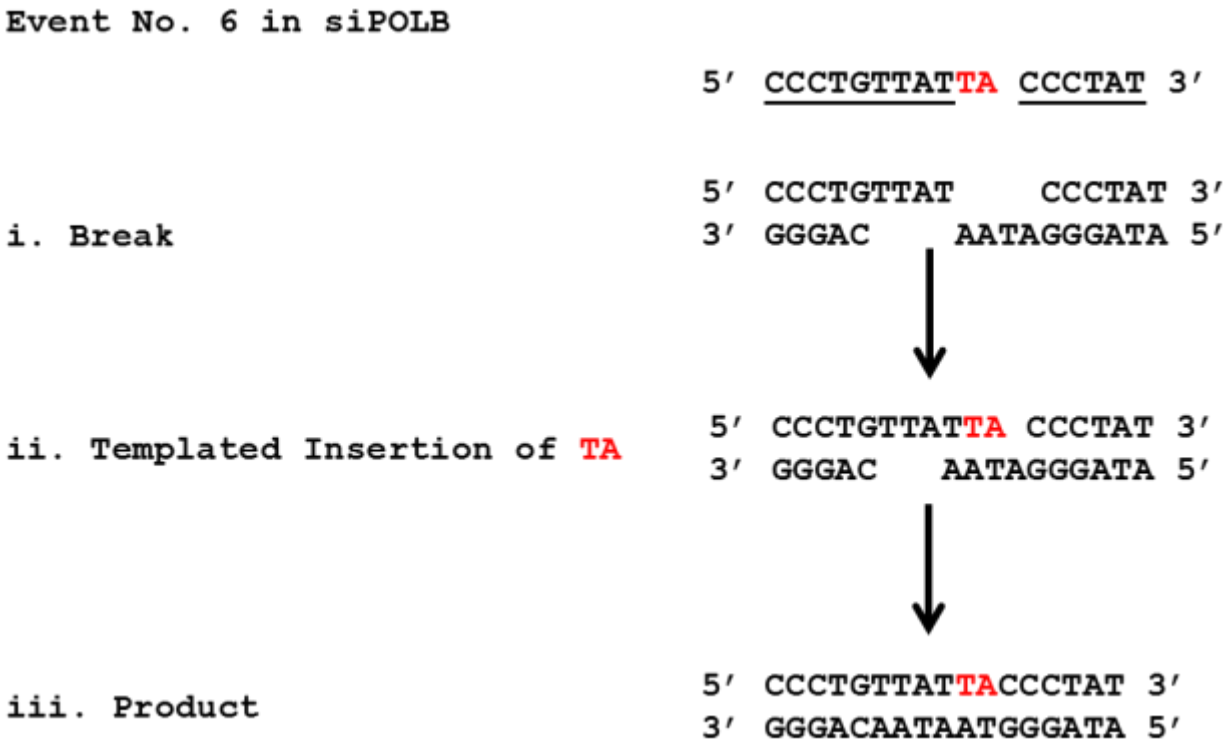

Figure S13B. Templated slippage, a DNA polymerase theta (Pol θ) signature.

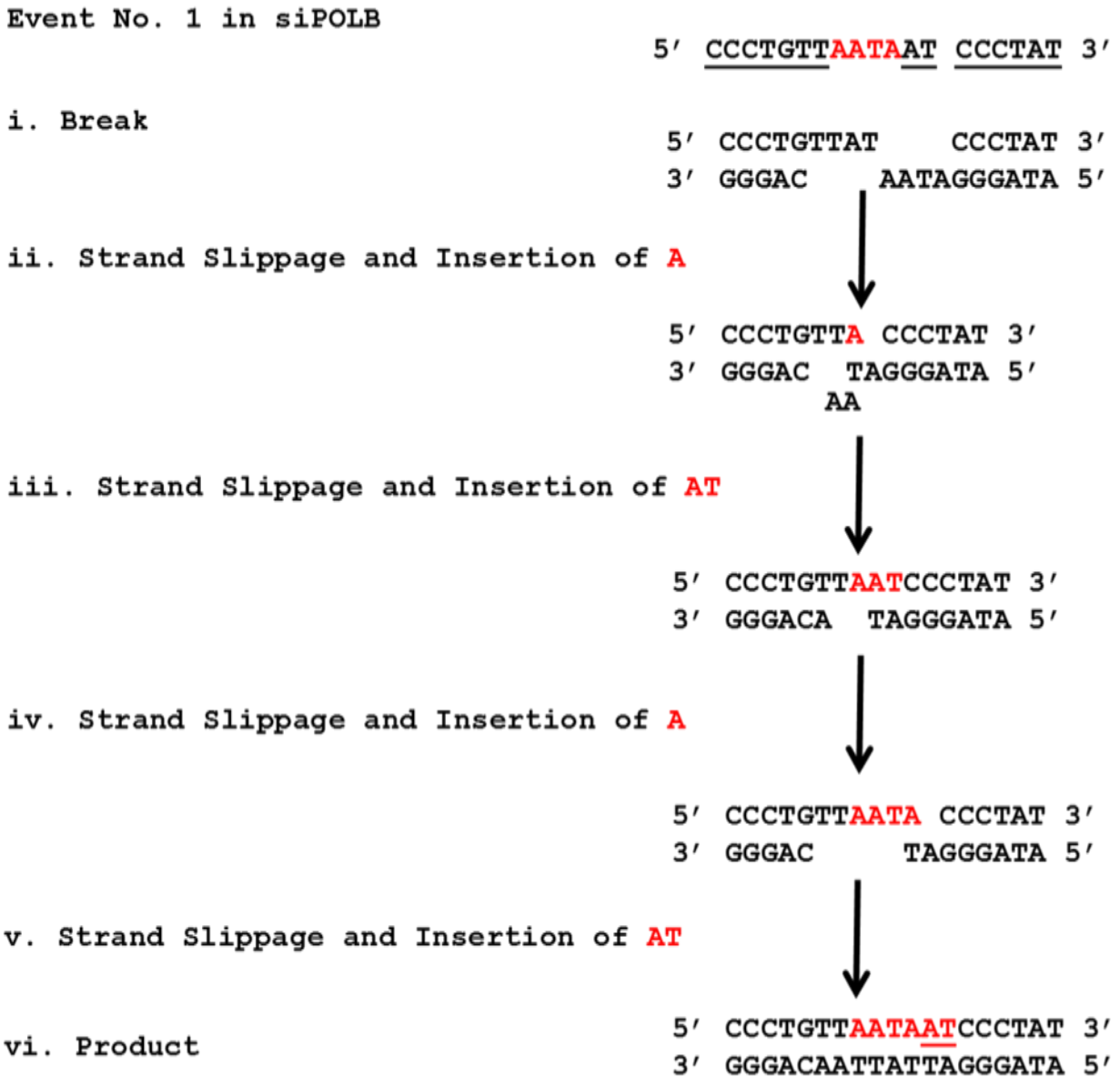

Figure S13C.

Misincorporation and Synthesis Across Break.

Figure S13. Modeling of the insertions uncovers at least two insertion signatures.

A. Model displaying template insertions. The event number 6 (Figure S12) in siPOLB is an example for templated insertions. The top most sequence indicates the insertion event. The parental sequence is underlined and the insertion sequence is indicated in red. i. The break indicates the endonuclease activity of I-SceI results in 4 base overhangs. ii. Templated insertion of TA. iii. The final product after the bases are inserted. Base insertions are highlighted in red. **B.** Model displaying templated slippage, a Pol θ signature. The event number 1 (Figure S12) in siPOLB is an example for templated slippage. The top most sequence indicates the insertion event. The parental sequence is underlined and the insertion sequence is indicated in red. i. The break indicates the endonuclease activity of I-SceI results in 4 base overhangs. ii. Strand slippage and insertion of A. iii. Strand slippage and insertion of AT. iv. Strand slippage and insertion of A. v. The final product after the bases are inserted. Base insertions are highlighted in red. **C.** Model displaying misincorporation and synthesis across break. The event number 1 (Figure S11) in the control is an example for misincorporation and synthesis across a break. The top most sequence indicates the insertion event. The parental sequence is underlined and the insertion sequence is indicated in red. i. The break indicates the endonuclease activity of I-SceI results in 4 base overhangs. ii. The end-processing indicates the bases removed and are highlighted in green. iii. Misincorporation and synthesis across break or slippage. iv. Realignment. v. Gap filling. vi. The final product after the bases are inserted. Base insertions are highlighted in red

Figure S14.

Reconstruction Experiment

Synthesized oligonucleotides that comprise the TetR sequence itself and also oligonucleotides that incorporated siRNA deletion event 1 were ordered from Integrated DNA Technologies. Based on our sequencing results, we estimated that the ratio of uncut: siRNA deletion event 1 was ~1:104. We performed a PCR experiment, using conditions and primers similar to the ones we used in the Ion Torrent experiment described in the manuscript, in which we mixed various ratios of uncut full length template and siPOLB event 1 sequences at 1:1, 10:1, 100:1, and 200:1 (uncut to siPOLB event 1). We also PCR amplified each oligonucleotide separately. We used 6-FAM conjugated primers and analyzed our results using capillary electrophoresis. Each capillary electrophoresis sample distribution was analyzed as a linear combination of two distributions of fragment lengths obtained by conducting the described PCR protocol on pure samples of one oligo or the other. The resulting fraction is the average percent contribution of one oligo to the distribution observed within each mixing sample. The results are shown below.

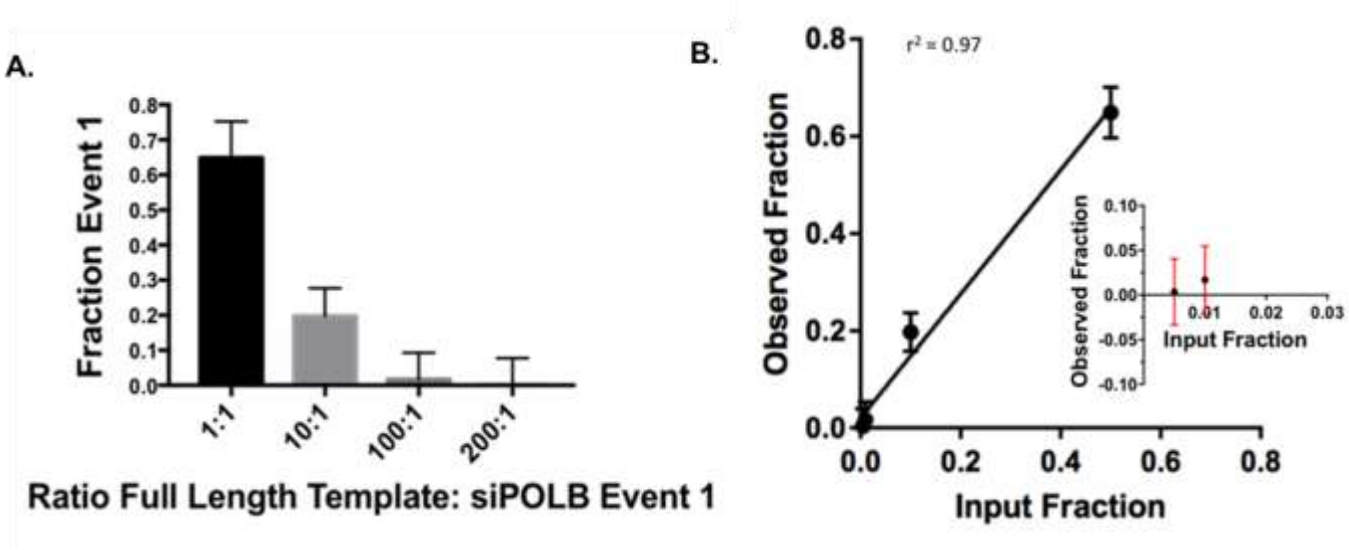

Figure S14. Graphs of the expected and observed amount of each sample. A. The input ratios are graphed against the product of siPOLB event 1 that we obtained in the reconstruction experiments (n=4). As we decrease the template representing siPOLB event 1, we observed decreased fraction of product. Note that for 1:100 the mean fraction we observed is 0.016 and for 1:200 the mean observed fraction is 0.003. **B.** The input fraction is graphed against the observed fraction for siPOLB event 1. The inset displays the 1:100 and 1:200 input fraction (siPOLB event 1: full length template) results. The observed ratios are linear with respect to the expected ratios in the higher percentages, which indicates that if anything the bias is fairly minimal. The output products continue to be linear with respect to the input product down to percentages that represent low-frequency events. Therefore, using the mixing approach we were able to start with a specific proportion of siPOLB and uncut sequences, and observed similar proportions after PCR amplification. This suggests that we are able to use the PCR and Ion Torrent approach to detect sequences with insertions and deletions that may be represented at low levels in our sample in comparison to the uncut (or precisely rejoined) sequences.

Importantly these experiments demonstrate: (1) that the potential bias is not great enough over the course of 30 PCR cycles to make a very minimally observed product into a major observed product; (2) any potential bias is not dependent on starting concentration of products, because the correlation between expected/observed sequence ratios is linear down to very low quantities. This indicates that any bias in template preference would be equal across samples whether the initial repair product occurred at a frequency of 0.5% to 50%.

Supplementary Tables

Table S1. siRNA used for Pol β

si Name	Dharmacon-catalogue number	Sequence	Annotation
siPOLB	D-005164-01	GAAUUGGGCUGAAAUAUUU	si1
siPOLB	D-005164-03	CAAGGAAGUUUGUAGAUGA	si3
siPOLB	D-005164-04	GAUACGAGUUCAUCCAUCA	si4

Table S2: Summary of Small Deletions Observed Predominantly in the Control Cells

Event No.[1]	Control# (%)[2]	siPOLB# (%)[3]	Odds Ratio Control/siPOLB	p-value[4]	# of Bases of μH	Base Fill	Total Bases Removed
1	271 (0.50)	42 (0.17)	2.9	$6.35X10^{-9}$	2	0	16
2	175 (0.33)	30 (0.12)	2.6	0.0001	0	0	8
3	216 (0.40)	38 (0.15)	2.5	$1.32X10^{-5}$	1	2	42
4	324 (0.60)	60 (0.24)	2.4	$2.38X10^{-8}$	1	0	18
5	314 (0.58)	67 (0.27)	2.1	$5.05X10^{-6}$	1	3	7
6	1867 (3.48)	514 (2.11)	1.6	$4.87X10^{-22}$	0	2	10
7	1923 (3.58)	607 (2.49)	1.4	$9.44X10^{-13}$	1	2	4
8	1443 (2.69)	512 (2.10)	1.3	0.0006	0	4	6
9	102 (0.19)	10 (0.04)	4.6	0.0002	2	0	10
10	121 (0.23)	8 (0.03)	6.8	$6.27X10^{-7}$	1	2	18

[1] Event number as indicated in Figure S6. [2] Percentage of these events in the control.
[3] Percentage of these events in the siPOLB. [4] p-value indicates the test of significance for the odds ratio, adjusted for multiple testing.

Table S3: Summary of Small Deletions Observed Predominantly in the Pol β-Depleted Cells

Event No. [1]	Control # (%)[2]	siPOLB# (%)[3]	Odds Ratio siPOLB/Control	p-value[4]	# of Bases of μH	Base Fill	Total Bases Removed
1	1811 (3.37)	1037 (4.26)	1.2	$3.90X10^{-7}$	0	4	8
2	767 (1.43)	476 (1.97)	1.3	$2.33X10^{-5}$	1	0	6
3	1495 (2.78)	1206 (4.96)	1.8	$1.36X10^{-50}$	0	4	32
4	306 (0.57)	262 (1.07)	1.8	$6.87X10^{-12}$	0	2	26
5	112 (0.20)	111 (0.45)	2.1	$1.21X10^{-6}$	0	3	17
6	43 (0.08)	52 (0.21)	2.6	0.00049	0	4	20
7	44 (0.08)	84 (0.34)	4.2	$3.42X10^{-14}$	1	0	36
8	7 (0.01)	56 (0.23)	17.6	$6.99X10^{-20}$	0	4	28
9	26 (0.04)	63 (0.25)	5.3	$7.12X10^{-13}$	1	0	22
10	0	12 (0.04)	55.1	0.0003	1	2	34
11	1 (0.001)	21 (0.08)	46.3	$1.36X10^{-7}$	0	1	21
12	1 (0.001)	21 (0.08)	46.3	$1.36X10^{-7}$	0	0	34
13	9 (0.01)	55 (0.22)	13.5	$4.44X10^{-18}$	2	0	30

14	19 (0.03)	53 (0.21)	6.1	$8.14X10^{-12}$	0	4	22
15	12 (0.02)	27 (0.11)	4.9	0.0002	1	0	40

[1] Event number as indicated in Figure S7. [2] Percentage of these events in the control. [3] Percentage of these events in the siPOLB. [4] p-value indicates the test of significance for the odds ratio, adjusted for multiple testing.

Table S4: Summary of Large Deletions Observed Predominantly in the Control Cells

Event No. [1]	Control# (%)[2]	siPOLB# (%)[3]	Odds Ratio Control/siPOLB	p-value[4]	# of Bases of μH	Base Fill	Total Bases Removed
11	167 (0.31)	8 (0.03)	9.4	$2.05X10^{-11}$	2	1	67
12	58 (0.10)	1 (0.004)	26.3	0.0008	1	0	44
13	71 (0.13)	1 (0.004)	32.2	$3.82X10^{-5}$	1	0	55

[1] Event number as indicated in Figure S9. [2] Percentage of these events in the control. [3] Percentage of these events in the siPOLB. [4] p-value indicates the test of significance for the odds ratio.

Table S5: Summary of Large Deletions Observed Predominantly in the Pol β-Depleted Cells

Event No. [1]	Control# (%)[2]	siPOLB# (%)[3]	Odds Ratio siPOLB/Control	p-value[4]	# of Bases of μH	Base Fill	Total Bases Removed
16	0	16 (0.06)	72.8	$3.16X10^{-6}$	0	0	44
17	0	13 (0.05)	59.5	0.0001	0	0	42
18	1 (0.001)	19 (0.07)	41.9	$1.28X10^{-6}$	1	0	56
19	1 (0.001)	14 (0.05)	30.8	0.0003	1	0	79
20	2 (0.003)	15 (0.06)	16.5	0.0005	0	0	42
21	4 (0.007)	24 (0.09)	13.2	$6.67X10^{-7}$	0	2	46

[1] Event number as indicated in Figure S10. [2] Percentage of these events in the control. [3]Percentage of these events in the siPOLB. [4] p-value indicates the test of significance for the odds ratio.

Table S6: Insertion Events Observed Predominantly in the Control Cells.

Event No. [1]	Control# (%)[2]	siPOLB# (%)[3]	Odds Ratio Control/siPOLB	p-value[4]
1	108 (0.20)	13 (0.05)	3.7	0.0007

[1] Event number as indicated in Figure S11. [2] Percentage of these events in the control. [3] Percentage of these events in the siPOLB. [4]p-value indicates the test of significance for the odds ratio.

Table S7: Insertion Events Observed Predominantly in Pol β-Depleted Cells

Event No.[1]	Control# (%)[2]	siPOLB# (%)[3]	Odds Ratio siPOLB/Control	p-value[4]
1	0	19 (0.07)	86.0	$1.82X10^{-7}$
2	2 (0.003)	16 (0.06)	17.6	0.000193
3	5 (0.009)	21 (0.08)	9.2	$6.13X10^{-5}$
4	19 (0.03)	37 (0.15)	4.3	$1.56X10^{-5}$
5	36 (0.06)	58 (0.23)	3.5	$1.35X10^{-7}$
6	56 (0.10)	77 (0.31)	3.0	$2.12X10^{-8}$
7	587 (1.09)	367 (1.51)	1.3	0.000487
8	3654 (6.81)	2043 (8.40)	1.2	$1.14X10^{-12}$

[1] Event number as indicated in Figure S12. [2] Percentage of these events in the control. [3] Percentage of these events in the siPOLB. [4] p-value indicates the test of significance for the odds ratio.