PNAS
www.pnas.org

Supplementary Information for:

Mechanistic Basis for Microhomology Identification and Genome Scarring by Polymerase Theta

Juan Carvajal-Garcia[a], Jang-Eun Cho[b], Pablo Carvajal-Garcia[c], Wanjuan Feng[b], Richard D. Wood[d], Jeff Sekelsky[b,e], Gaorav P. Gupta[b,f,g], Steven A. Roberts[h], Dale A. Ramsden[b,f,*]

* To whom correspondence should be addressed, Dale A. Ramsden

Email: dale_ramsden@med.unc.edu

This PDF file includes:

Supplementary text

Figures S1 to S4

Tables S1 to S6

SI References

Supplementary materials and methods

Materials

Cells were incubated at 37 °C, 5% CO_2 and cultured in DMEM (Gibco) with 10% Fetal Bovine Serum (VWR Life Science Seradigm) and Penicillin (5 U/ml, Sigma). These lines and variants described below were confirmed to be free of mycoplasma contamination by a qPCR (1) with a detection limit below 10 genomes/1ml. Cell lines were additionally selected at random for third party validation of PCR results using Hoechst staining (2).

www.pnas.org/cgi/doi/10.1073/pnas.1921791117

Methods

Extrachromosomal Substrate generation

A core fragment generated by PCR (primers in table S3) was digested with BsaI, and extrachromosomal substrate assembled by ligation of this BsaI digested core fragment with a 5-fold excess of head and tail caps generated by annealing oligos described in table S2 with T7 ligase (NEB). Excess cap was removed using QIAquick PCR Purification Kit (QIAGEN), and efficient substrate assembly and removal of excess cap confirmed by analytical gel electrophoresis.

High throughput sequencing junction characterization

The total number of sequence reads for each biological replicate is reported in supplemental Table S6. Junctions were characterized by independently identifying within each read the least-deleted 10 nucleotide match to sequence upstream of the break site, then the least-deleted 10 nucleotide match to sequence downstream of the break site. When present, microhomologies were defined as the overlap between upstream and downstream matches, and insertions were defined as non-matching sequences separating upstream and downstream matches. Further analysis of junctions employed a "reconstructed junction", consisting of a concatenation of the 10 bp upstream flank match (including microhomology, when present), inserted sequence (when present), and the 10 bp downstream flank (excluding microhomology, when present). We report in supplemental datasets S1-S5 the frequencies of these reconstructed junctions for each of the 3 biological replicates for complemented *Polq*$^{-/-}$ cells ("PlusPOLQ") as well as the parental line ("MinusPOLQ") for all 5 break sites (R26A-E). The location of microhomologies is defined as the distance relative to the break point after excluding microhomologous sequence in these supplemental tables (columns titled upstream and downstream deletion), as well as in results and discussion. Additional tables that describe the further analysis of these junctions can be provided upon request.

For Figs. 3 and 4, insertions were characterized as templated direct repeats if the first 5 nucleotides of the inserted sequence could be mapped to sequence within 50 nt downstream of the break site ("downstream direct repeats") or the last 5 nucleotides of the inserted sequence could be mapped to sequence 50 nt upstream of the break site ("upstream direct repeats"). Insertions were characterized as templated inverse repeats if the 5 last nucleotides of the inserted sequence could

be mapped to the reverse complement of sequence within 50 nt downstream of the break ("downstream inverse repeats") or if the 5 first nt of the inserted sequence could be mapped to the reverse complement of the 50 nt upstream of the break ("upstream inverse repeats"). We included cases of insertions < 5 nt as templated if additional templated insertions could be inferred due to involvement of 2° microhomologies in resolution (i.e. when sequence downstream of the insertion extended the identity that was detected in the proposed template to a total of 5 nt or more). When multiple flanking sequences with 5 nt or more of identity to the insert were identified, we selected as template the flanking sequence with the largest match. We excluded insertions where the first inserted nucleotide was substituted, relative to the reference, but subsequent inserted sequence was identical to reference; such products could be identified in control experiments as substitutions made during sample amplification.

Direct repeat 1° microhomologies in TINS with template in downstream DNA were determined by assessing the extent of identity between nucleotides in flanking DNA upstream of the insert when compared to nucleotides upstream of the identified template. Conversely, direct repeat 1° microhomologies in TINS with template in upstream DNA were determined by assessing the extent of identity between nucleotides in flanking DNA downstream of the insert when compared to nucleotides downstream of the identified template. 2° microhomologies were assessed similarly, except we assessed the extent of identity when comparing sequences on the opposite sides of the insert and identified template. As noted in the legend to Fig. 4F, we assessed 2° microhomologies only for TINS less than 5 nucleotides, to exclude significant contribution of TINS where there was more than one round of synthesis. (e.g. Fig. S3B).

For Fig. 6 and S5, mutations from 569 whole genome sequenced breast cancers were obtained from the International Cancer Genome Consortium (ICGC) data portal (https://dcc.icgc.org/api/v1/download?fn=/release_26/Projects/BRCA-EU/simple_somatic_mutation.open.BRCA-EU.tsv.gz). Mutation lists were sorted by sample, chromosome and position prior to removing duplicate mutation entries. All insertion mutations and multiple nucleotide variants (MNV, which involve deletion of a sequence and insertion of a new, non-reference sequence) were extracted. For MNVs, the inserted sequences were compared to

the deleted reference sequences to confirm accurate determination of the position the MNV event. For a small portion of MNVs, several nucleotides at either the 5' or 3' portions of the inserted sequences matched the sequence of the reference. These matching sequences were removed from the insertion sequence and the position of the MNV corrected prior to subsequent analysis. 100 nt of DNA sequence flanking each mutation was retrieved from the hg19 reference sequence for the human genome. Templated insertions (TINS) events among the breast cancer mutations were defined as above, except we considered only 30 nt of flanking DNA sequence as possible template. We also excluded templated insertions in which the inserted sequence immediately followed the template (tandem repeats). Breast cancer mutations were additionally stratified by *BRCA1* and/or *BRCA2* deficiency as determined by a germline mutation or hyper methylation in either gene as reported (3). Original data files describing the *BRCA1/2* status of ICGC characterized tumors can be found at https://github.com/wenweixiong/BRCA2018. We obtained gene expression data for the evaluated tumors as log(2) transformed fragments per kilobase per million reads (FPKM) RNA-seq from Supplementary Table 7 of (4). We normalized *POLQ* log(2) FPKM counts according to housekeeping gene expression data as follows: for each of three housekeeping genes (*TBP*, *HPRT*, and *GAPDH*), we determined the difference between the log(2) FPKM counts for each tumor from the mean counts for the whole set, then averaged these three numbers for each tumor to determine normalization factors, which were then subtracted from the *POLQ* log(2) FPKM counts for each tumor.

Figure S1

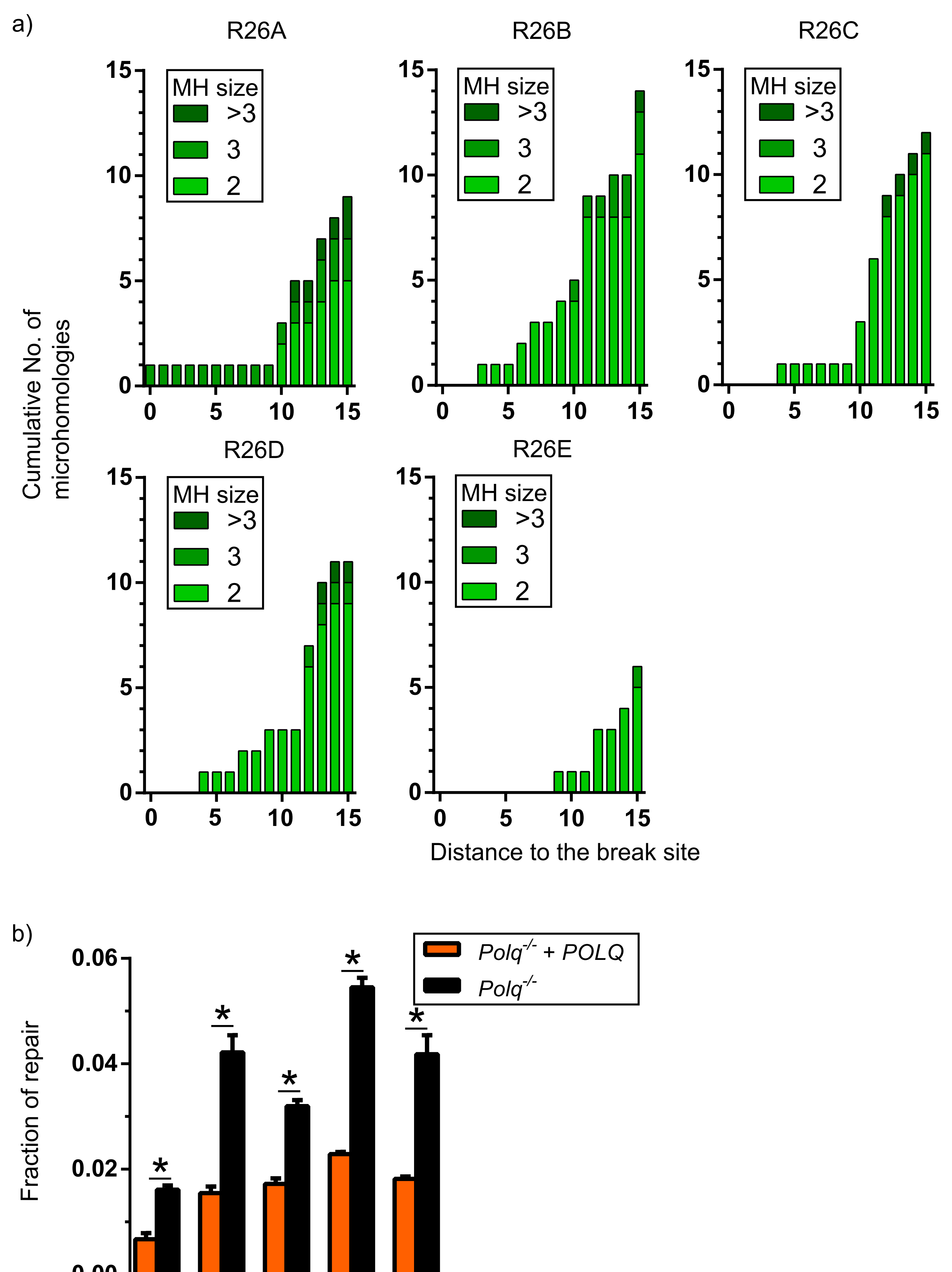


Fig. S1. A) Cumulative number of microhomologies larger than 1 bp at the noted distance to the break site. Distance to the break site corresponds to the largest of the two distances (upstream and downstream) from the beginning of the microhomology to the break terminus. B) Fraction of repair corresponding to deletions with microhomologies 2 bp or more and located more than 15 nt from both sides of the break site in *Polq*$^{-/-}$ cells that expressed POLQ (orange) or not (black). Bar represents the means and error bars SEM for 3 biological replicates. Statistical significance was assessed by one-way ANOVA with Bonferroni correction to account for multiple comparisons; *, $p<0.05$.

Figure S2

a)

dT(2) unpaired tail

dT(n) unpaired tails
(n)=2 or 15 nt

Joining efficiency

1.25
1.00
0.75
0.50
0.25
0.00

*
*

2/2 2/15 2/2

+*POLQ* + + -

b)

R26D

9 nt
2 nt
TA
TA
TA
AT
4 nt

or

2 nt
TA
TA
AT
4 nt

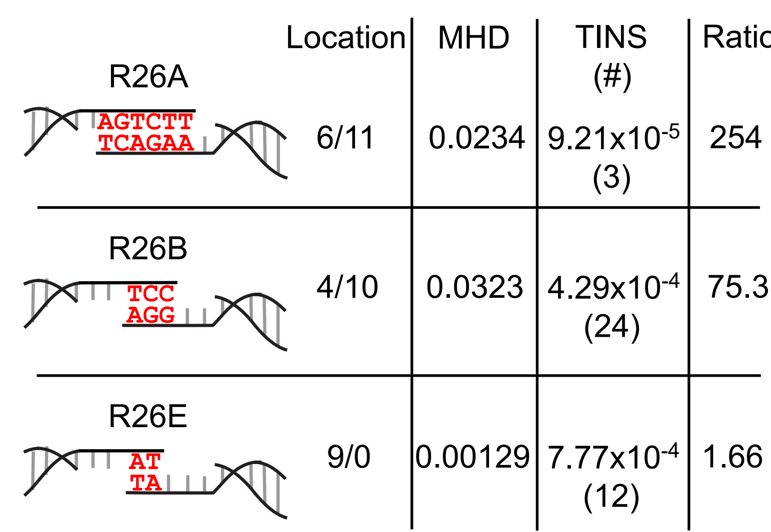
9 nt
TA
TA
AT
4 nt

Average (SEM) 0.0141 (0.0008) 3-fold 0.00430 (0.0001)

Fig. S2. A) Joining efficiency was calculated as in Fig. 2 for a substrate with a microhomology located 2 nt away from the 3' terminus for both head and tails ends (2/2), as well as a substrate with a microhomology 2 nt from the head 3' terminus and 15 nt from the tail 3' terminus (2/15). Bars represent the mean and error bars SEM for 3 biological replicates. Statistical significance was assessed by one-way ANOVA with Bonferroni correction to account for multiple comparisons B) Schematic of two Pol θ dependent MHD identified in chromosomal repair products at the break site R26D. The fraction of repair products enriched by *POLQ* expression was calculated as in Fig 1. SEM is shown in parenthesis for 3 biological replicates.

Figure S3

a)



	Location	MHD	TINS (#)	Ratio
R26A (AGTCTT / TCAGAA)	6/11	0.0234	9.21×10^{-5} (3)	254
R26B (TCC / AGG)	4/10	0.0323	4.29×10^{-4} (24)	75.3
R26E (AT / TA)	9/0	0.00129	7.77×10^{-4} (12)	1.66

b)

R26E

AACCAGACAGATTAGTTACATACACCACAA ATCGAGGCTGTAGCTGGGGCCTCAACACTG

TTGGTCTGTCTAATCAATGTATGTGGTGCC TAGCTCCGACATCGACCCCGGTGTTGTGAC

Insert: GTTAGATTAGT

TGTAATCA: 1° synthesis

AATCAA: 2° synthesis

CAGACAGATTAGTTA

TCTAATCAAT

1° microhomology

CAGACAGATTAGTTA

AATCAATCTAATCA AT

2° microhomology

CAGACAGATTAGTTA

CTAATCAATCAATCTAATCA AT

3° microhomology

Insert

CAGACAGATTAGTTAGTTACATTAGT

CTAATCAATCAATCTAATCAAT

Fig. S3. A) The fraction of repair products enriched by *POLQ* expression comparing MHD vs. TINS and number of different TINS identified (#) for each of three 1° microhomologies of differing length; a 6 bp AGTCTT microhomology in R26A, a 3 bp TCC microhomology in R26B, and a 2 bp AT microhomology in R26E). The location of the 1° microhomology with respect to the break site is indicated as upstream deletion/downstream deletions). B) Generation of a repair product consistent with three microhomology primed synthesis events. Primary (1°) microhomology is shown in red, 1° round of synthesis in dark blue, and secondary (2°) round of synthesis in cyan.

Figure S4

a)

R26A AGTCTTTCTAGA AGATGGGCGGGA

AGTCTTTCTAGA---TGGGCGGGA

3 bp MHD

wt → 17%

Polq-/- → 11%

Ku70-/- → 1.5%

b)

Polq-/- + *POLQ*

Polq-/-

Fraction of repair

0.15

0.10

0.05

0.00

1.8 fold

7.2 fold

MHD

TINS

c)

TINS/tumor

35

30

25

20

15

10

5

0

*

Low

High

POLQ expression

d)

% AT

70

65

60

55

50

*

*

TINS

Flanks

wt

mutant

BRCA1/2 germline status

e)

Normalized *POLQ* mRNA

4

2

0

-2

-4

*

wt

mutant

BRCA1/2 germline status

Fig. S4. A) Structure of a repair event in R26A characterized as deletion of a terminal 3 bp microhomology (in orange), with the fraction of repair represented by this product noted for wild type, *Polq*$^{-/-}$ and *Ku70*$^{-/-}$ cells. Data for *Ku70*$^{-/-}$ was obtained from (5). B) Fraction of repair corresponding to MHD and TINS averaged across the 5 break sites tested in cells expressing *POLQ* (orange) or not (black). The fold difference is shown and was calculated as the fraction in *POLQ* expressing cells divided by the fraction in parental *Polq*$^{-/-}$ cells. C) The frequency of TINS/tumor genome was determined for tumors with high or low levels of *POLQ* expression determined as in Fig. 6A, except tumors with germline *BRCA* mutations were excluded. D) Average percent AT content in TINS found in tumor genomes defined as in Fig. 6 (blue), compared to the average AT content in the 100 bp surrounding the insert (Flanks, white). Statistical significance was assessed by a two-tailed t-test; *, $p<0.05$. E) Levels of *POLQ* mRNA were normalized as in Fig. 6A, and compared in tumors with wild type (wt) germline *BRCA* genes vs. tumors with germline mutations in *BRCA* genes. Statistical significance was assessed with a two-tailed Mann-Whitney test, *$p<0.05$.

Table S1: Pol θ dependent repair events

Locus	Upstream deletion	Downstream deletion	Microhomology size	Fraction enriched by POLQ
R26A	0	0	3	0.057557
	6	11	6	0.023439
	10	0	2	0.006968
	3	5	0	0.000259
R26B	4	10	3	0.032344
	3	0	2	0.017555
	1	11	2	0.008442
	6	6	2	0.006746
	3	4	1	0.004777
	7	2	2	0.004391
	8	4	1	0.00411
	6	9	1	0.003844
	2	6	0	0.001395
R26C	2	4	2	0.036814
	9	9	0	0.000588
	11	6	0	0.00058
	3	7	0	0.000381
	15	3	0	0.000305
R26D	4	7	2	0.021654
	2	4	2	0.014109
	1	4	1	0.008736
	2	1	0	0.006699
	0	1	0	0.005915
	2	1	1	0.004951
	9	4	2	0.004299
	4	0	1	0.004095
	8	4	1	0.002299
	1	1	0	0.0018
	9	1	0	0.001155
	2	41	3	9.37E-05
R26E	6	6	0	0.00366

Upstream and downstream deletion, microhomology size, and fraction of repair enriched by POLQ expression for all the repair events significantly enriched in cells expressing wt POLQ vs. Polq-/- in triplicate experiments. Statistical significance was identified using a two tailed t-test and the Benjamini-Hochberg procedure to adjust p values for multiple comparisons, with a false discovery rate of 0.05.

Table S2. Substrate preparation oligonucleotides

Substrate	Side	Sequence
All (bottom strand)	Head	CATCGCTTAGCTGTATA
	Tail	CTCACACCCATCTCA
2/2	Head	5'/phos/TGACTATACAGCTAAGCGATGCTCTCACCGAGCGTATCTGCTGGGTTGTGGATGAATTACATATGCTGGGAGAACCAAGATTGGGCAGTT
	Tail	5'/phos/AGTCTGAGATGGGTGTGAGAGTGAAGATCCTCACCTTCGGAGTACTCCTTCTTTTGACCATTGATACGATACTTCTCAGCCGAGCTGCTT
10/10	Head	5'/phos/TGACTATACAGCTAAGCGATGCTCTCACCGAGCGTATCTGCTGGGTTGTGGATGAATTACATATGCTGGGAGAACCGCAGTTTTTTTTTT
	Tail	5'/phos/AGTCTGAGATGGGTGTGAGAGTGAAGATCCTCACCTTCGGAGTACTCCTTCTTTTGACCATTGATACGATACTTCTCTGCTTTTTTTTTT
15/15 (4)	Head	5'/phos/TGACTATACAGCTAAGCGATGCTCTCACCGAGCGTATCTGCTGGGTTGTGGATGAATTACATATGCTGGGCGCAGTTTTTTTTTTTTTTT
	Tail	5'/phos/AGTCTGAGATGGGTGTGAGAGTGAAGATCCTCACCTTCGGAGTACTCCTTCTTTTGACCATTGATACGATCCTGCTTTTTTTTTTTTTTT
14/14 (6)	Head	5'/phos/TGACTATACAGCTAAGCGATGCTCTCACCGAGCGTATCTGCTGGGTTGTGGATGAATTACATATGCTGGGAGCAGTTTTTTTTTTTTTTT
	Tail	5'/phos/AGTCTGAGATGGGTGTGAGAGTGAAGATCCTCACCTTCGGAGTACTCCTTCTTTTGACCATTGATACGATACTGCTTTTTTTTTTTTTTT
30/30	Head	5'/phos/TGACTATACAGCTAAGCGATGCTCTCACCGAGCGTATCTGCTGGGTTGTGGATGAGCAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
	Tail	5'/phos/AGTCTGAGATGGGTGTGAGAGTGAAGATCCTCACCTTCGGAGTACTCCTTCTTTACTGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
2/15	Head	5'/phos/TGACTATACAGCTAAGCGATGCTCTCACCGAGCGTATCTGCTGGGTTGTGGATGAATTACATATGCTGGGAGAACCAAGATTGGGCAGTT
	Tail	5'/phos/AGTCTGAGATGGGTGTGAGAGTGAAGATCCTCACCTTCGGAGTACTCCTTCTTTTGACCATTGATACGATCCTGCTTTTTTTTTTTTTTT
2+10/ 2+10	Head	5'/phos/TGACTATACAGCTAAGCGATGCTCTCACCGAGCGTATCTGCTGGGTTGTGGATGAATTACATATGCTATCCACGACGCAGTACGGCAGTT
	Tail	5'/phos/AGTCTGAGATGGGTGTGAGAGTGAAGATCCTCACCTTCGGAGTACTCCTTCTTTTGACCATTGATACGAGGTCACCCTGCTTGGCTGCTT
3bp-AT	Head	5'/phos/TGACTATACAGCTAAGCGATGCTCTCACCGAGCGTATCTGCTGGGTTGTGGATGAATTACATATCATGTAAGAACATTATTTCATTAGTT
	Tail	5'/phos/AGTCTGAGATGGGTGTGAGAGTGAAGATCCTCACCTTCGGAGTACTCCTTCTTTTGACCATTGATTAGTATAGTTATTCATTCATCTATT
3bp-GC	Head	5'/phos/TGACTATACAGCTAAGCGATGCTCTCACCGAGCGTATCTGCTGGGTTGTGGATGAATTACATATGCTGGGACGGCGACGACGGCGTAGTT
	Tail	5'/phos/AGTCTGAGATGGGTGTGAGAGTGAAGATCCTCACCTTCGGAGTACTCCTTCTTTTGACCATTGATGCGCGGACCAGGCGGAGTGGCTATT
6bp-AT	Head	5'/phos/TGACTATACAGCTAAGCGATGCTCTCACCGAGCGTATCTGCTGGGTTGTGGATGAATTACATATCATGTAAGAACATTATTTTAGATGTT
	Tail	5'/phos/AGTCTGAGATGGGTGTGAGAGTGAAGATCCTCACCTTCGGAGTACTCCTTCTTTTGACCATTGATTAGTATAGTTATTCATTCATCTATT

Top and bottom strands of the caps (head and tail) used to make extrachromosomal substrates (bottom is common to all substrates, preferentially used microhomology in top strands underlined). All oligos were PAGE purified (IDT).

Table S3. PCR primer sequences

PCR	Primer	Sequence
Substrate core PCR	Fwd	CAAGTGGTCTCAGACTGGCTACCCTGCT
	Rev	GCCGAGGTCTCCGTCAGGGAAATCAAACGTAA
Substrate quantification PCR	Fwd	TAAGCGATGCTCTCACCGAG
	Rev	GATGGGTGTGAGAGTGAAGATC

Forward and reverse primers used to amplify the substrate core and to quantify and characterize substrate repair products

Table S4. Break site sequences

Locus	gRNA sequence + PAM	Chr. 6 location
R26A	ACTCCAGTCTTTCTAGAAGATGG	113,068,731
R26B	CCGCCTCGGAGTATTTTCCATCG	113,068,958
R26C	CATGGATTTCTCCGGTGAATAGG	113,071,870
R26D	CCTATTCACGTAACCAGGTTAGC	113,075,531
R26E	CCACAAATCGAGGCTGTAGCTGG	113,069,295

gRNA sequences + PAM (underlined) and sequence location in the mouse chromosome 6 for all 5 break sites tested.

Table S5. Sequencing library preparation primers

PCR	Sample	Primer	SEQUENCE
R26A	*Polq*$^{-/-}$ + *POLQ* (1)	Fwd	CTACACGACGCTCTTCCGATCTCGATGTCGAAGTTCTCTGCTGCCTCCTGGCTTCT
		Rev	GCTGAACCGCTCTTCCGATCTCAGATCTTCTGTGGGAAGTCTTGTCCCTCCAA
	Polq$^{-/-}$ + *POLQ* (2)	Fwd	CTACACGACGCTCTTCCGATCTATCACGATGAAGTTCTCTGCTGCCTCCTGGCTTCT
		Rev	GCTGAACCGCTCTTCCGATCTCAGATCTTCTGTGGGAAGTCTTGTCCCTCCAA
	Polq$^{-/-}$ + *POLQ* (3)	Fwd	CTACACGACGCTCTTCCGATCTTTAGGCTGCGAAGTTCTCTGCTGCCTCCTGGCTTCT
		Rev	GCTGAACCGCTCTTCCGATCTCAGATCTTCTGTGGGAAGTCTTGTCCCTCCAA
	Polq$^{-/-}$ (1)	Fwd	CTACACGACGCTCTTCCGATCTCGATGTCGAAGTTCTCTGCTGCCTCCTGGCTTCT
		Rev	GCTGAACCGCTCTTCCGATCTTGACCAGTTCTGTGGGAAGTCTTGTCCCTCCAA
	Polq$^{-/-}$ (2)	Fwd	CTACACGACGCTCTTCCGATCTATCACGATGAAGTTCTCTGCTGCCTCCTGGCTTCT
		Rev	GCTGAACCGCTCTTCCGATCTTGACCAGTTCTGTGGGAAGTCTTGTCCCTCCAA
	Polq$^{-/-}$ (3)	Fwd	CTACACGACGCTCTTCCGATCTTTAGGCTGCGAAGTTCTCTGCTGCCTCCTGGCTTCT
		Rev	GCTGAACCGCTCTTCCGATCTTGACCAGTTCTGTGGGAAGTCTTGTCCCTCCAA
R26B	*Polq*$^{-/-}$ + *POLQ* (1)	Fwd	CTACACGACGCTCTTCCGATCTACAGTGGAGTGCAATTGGAGGGACAAGACTTCC
		Rev	GCTGAACCGCTCTTCCGATCTGCCAATCCTGTGCAGGATCTCAAGCAGGAGAGTA
	Polq$^{-/-}$ + *POLQ* (2)	Fwd	CTACACGACGCTCTTCCGATCTGCCAATCCTGTGCAATTGGAGGGACAAGACTTCC
		Rev	GCTGAACCGCTCTTCCGATCTGCCAATCCTGTGCAGGATCTCAAGCAGGAGAGTA
	Polq$^{-/-}$ + *POLQ* (3)	Fwd	CTACACGACGCTCTTCCGATCTTAGCTTAACACAATAATTGGAGGGACAAGACTTCC
		Rev	GCTGAACCGCTCTTCCGATCTGCCAATCCTGTGCAGGATCTCAAGCAGGAGAGTA
	Polq$^{-/-}$ (1)	Fwd	CTACACGACGCTCTTCCGATCTACAGTGGAGTGCAATTGGAGGGACAAGACTTCC
		Rev	GCTGAACCGCTCTTCCGATCTACAGTGGAGTGCAGGATCTCAAGCAGGAGAGTA
	Polq$^{-/-}$ (2)	Fwd	CTACACGACGCTCTTCCGATCTGCCAATCCTGTGCAATTGGAGGGACAAGACTTCC
		Rev	GCTGAACCGCTCTTCCGATCTACAGTGGAGTGCAGGATCTCAAGCAGGAGAGTA
	Polq$^{-/-}$ (3)	Fwd	CTACACGACGCTCTTCCGATCTTAGCTTAACACAATAATTGGAGGGACAAGACTTCC
		Rev	GCTGAACCGCTCTTCCGATCTACAGTGGAGTGCAGGATCTCAAGCAGGAGAGTA
R26C	*Polq*$^{-/-}$ + *POLQ* (1)	Fwd	CTACACGACGCTCTTCCGATCTATCACGCACAAGGCAGACAACCAAGAAAC
		Rev	GCTGAACCGCTCTTCCGATCTTAGCTTAGACCTCGAAATAGCAGCTTTG

	Polq$^{-/-}$ + *POLQ* (2)	Fwd	CTACACGACGCTCTTCCGATCTATCACGCACAAGGCAGACAACCAAGAAAC
		Rev	GCTGAACCGCTCTTCCGATCTACAGTGTAGACCTCGAAATAGCAGCTTTG
	Polq$^{-/-}$ + *POLQ* (3)	Fwd	CTACACGACGCTCTTCCGATCTATCACGCACAAGGCAGACAACCAAGAAAC
		Rev	GCTGAACCGCTCTTCCGATCTGCCAATGCAGACCTCGAAATAGCAGCTTTG
	Polq$^{-/-}$ (1)	Fwd	CTACACGACGCTCTTCCGATCTCTTGTAATACAAGGCAGACAACCAAGAAAC
		Rev	GCTGAACCGCTCTTCCGATCTTAGCTTAGACCTCGAAATAGCAGCTTTG
	Polq$^{-/-}$ (2)	Fwd	CTACACGACGCTCTTCCGATCTCTTGTAATACAAGGCAGACAACCAAGAAAC
		Rev	GCTGAACCGCTCTTCCGATCTACAGTGTAGACCTCGAAATAGCAGCTTTG
	Polq$^{-/-}$ (3)	Fwd	CTACACGACGCTCTTCCGATCTCTTGTAATACAAGGCAGACAACCAAGAAAC
		Rev	GCTGAACCGCTCTTCCGATCTGCCAATGCAGACCTCGAAATAGCAGCTTTG
R26D	*Polq*$^{-/-}$ + *POLQ* (1)	Fwd	CTACACGACGCTCTTCCGATCTACAGTGACCTGGCATGGTATTGCTTATC
		Rev	GCTGAACCGCTCTTCCGATCTTTAGGCAGTCACCAGGTTTGTACTTGGTTCA
	Polq$^{-/-}$ + *POLQ* (2)	Fwd	CTACACGACGCTCTTCCGATCTGCCAATGACCTGGCATGGTATTGCTTATC
		Rev	GCTGAACCGCTCTTCCGATCTTTAGGCAGTCACCAGGTTTGTACTTGGTTCA
	Polq$^{-/-}$ + *POLQ* (3)	Fwd	CTACACGACGCTCTTCCGATCTGCCAATGACCTGGCATGGTATTGCTTATC
		Rev	GCTGAACCGCTCTTCCGATCTCGATGTGCACCAGGTTTGTACTTGGTTCA
	Polq$^{-/-}$ (1)	Fwd	CTACACGACGCTCTTCCGATCTACAGTGACCTGGCATGGTATTGCTTATC
		Rev	GCTGAACCGCTCTTCCGATCTGGCTACCTCACCAGGTTTGTACTTGGTTCA
	Polq$^{-/-}$ (2)	Fwd	CTACACGACGCTCTTCCGATCTGCCAATGACCTGGCATGGTATTGCTTATC
		Rev	GCTGAACCGCTCTTCCGATCTGGCTACCTCACCAGGTTTGTACTTGGTTCA
	Polq$^{-/-}$ (3)	Fwd	CTACACGACGCTCTTCCGATCTACAGTGACCTGGCATGGTATTGCTTATC
		Rev	GCTGAACCGCTCTTCCGATCTCGATGTGCACCAGGTTTGTACTTGGTTCA
R26E	*Polq*$^{-/-}$ + *POLQ* (1)	Fwd	CTACACGACGCTCTTCCGATCTCAGATCTTGGCTTATCCAACCCCTAGA
		Rev	GCTGAACCGCTCTTCCGATCTCTTGTATGTGGAACACCACCTGACG
	Polq$^{-/-}$ + *POLQ* (2)	Fwd	CTACACGACGCTCTTCCGATCTCAGATCTTGGCTTATCCAACCCCTAGA
		Rev	GCTGAACCGCTCTTCCGATCTACTTGACTGTGGAACACCACCTGACG
	Polq$^{-/-}$ + *POLQ* (3)	Fwd	CTACACGACGCTCTTCCGATCTCAGATCTTGGCTTATCCAACCCCTAGA

		Rev	GCTGAACCGCTCTTCCGATCTCAGATCGATGTGGAACACCACCTGACG
	Polq$^{-/-}$ (1)	Fwd	CTACACGACGCTCTTCCGATCTTGACCACATGGCTTATCCAACCCCTAGA
		Rev	GCTGAACCGCTCTTCCGATCTCTTGTATGTGGAACACCACCTGACG
	Polq$^{-/-}$ (2)	Fwd	CTACACGACGCTCTTCCGATCTTGACCACATGGCTTATCCAACCCCTAGA
		Rev	GCTGAACCGCTCTTCCGATCTACTTGACTGTGGAACACCACCTGACG
	Polq$^{-/-}$ (3)	Fwd	CTACACGACGCTCTTCCGATCTTGACCACATGGCTTATCCAACCCCTAGA
		Rev	GCTGAACCGCTCTTCCGATCTCAGATCGATGTGGAACACCACCTGACG
Substrate	3bp-AT	Fwd	CTACACGACGCTCTTCCGATCTCAGATCATCGACAGATCTAAGCGATGCTCTCACCGAG
		Rev	GCTGAACCGCTCTTCCGATCTCAGATCGTACCTTGACCAGATGGGTGTGAGAGTGAAGATC
	3bp-GC	Fwd	CTACACGACGCTCTTCCGATCTCAGATCATCGACAGATCTAAGCGATGCTCTCACCGAG
		Rev	GCTGAACCGCTCTTCCGATCTCAGATCTCGAGACGATGTGATGGGTGTGAGAGTGAAGATC
	6bp-AT	Fwd	CTACACGACGCTCTTCCGATCTCAGATCGTACCTTGACCATAAGCGATGCTCTCACCGAG
		Rev	GCTGAACCGCTCTTCCGATCTCAGATCATCGACAGATCGATGGGTGTGAGAGTGAAGATC
Secondary NGS PCR	All	Fwd	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
		Rev	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

Forward and reverse primers used for high throughput sequencing library preparation.

Table S6. Sequencing reads

Genotype	Break site	Replicate 1	Replicate 2	Replicate 3
Polq$^{-/-}$ + *POLQ*	R26A	31617	26815	19817
	R26B	64880	9713	140496
	R26C	135961	116342	140111
	R26D	172163	173076	175362
	R26E	185531	207476	247170
Polq$^{-/-}$	R26A	25726	18706	18636
	R26B	21176	141187	153290
	R26C	146179	148533	164259
	R26D	146713	138013	151899
	R26E	144012	132219	93086

Number of reads analyzed from each biological replicate for the two genotypes in the five loci tested.

SI references

1. K. Janetzko, *et al.*, A single-tube real-time PCR assay for mycoplasma detection as a routine quality control of cell therapeutics. *Transfus. Med. Hemotherapy* **41**, 83–89 (2014).
2. M. Battaglia, D. Pozzi, S. Crimaldi, T. Parasassi, Hoechst 33258 staining for detecting mycoplasma contamination in cell cultures: A method for reducing fluorescence photobleaching. *Biotech. Histochem.* **69**, 152–156 (1994).
3. W. X. Wen, C. O. Leong, Association of BRCA1- And BRCA2-deficiency with mutation burden, expression of PD-L1/ PD-1, immune infiltrates, and T cell-inflamed signature in breast cancer. *PLoS One* **14** (2019).
4. S. Nik-Zainal, *et al.*, Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
5. W. Feng, *et al.*, Genetic determinants of cellular addiction to DNA polymerase theta. *Nat. Commun.* **10** (2019).