SUPPLEMENTARY INFORMATION

A High-throughput and Quantitative Method to Assess the Mutagenic Potential of Translesion DNA Synthesis

David J. Taggart, Terry L. Camerlengo, Jason K. Harrison, Shanen M. Sherrer, Ajay K. Kshetry, John-Stephen Taylor, Kun Huang, and Zucai Suo

Supplementary Table S1. Primers used for the addition of unique barcodes and Illumina adapter sequences.

hPolη-Control forward[a]	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTATCTCGGCATCAGCAATGTTG-3′
hPolη-Damaged forward[a]	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTATAGCGGCATCAGCAATGTTG-3′
hPolκ-Control forward[a]	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACTCGGCATCAGCAATGTTG-3′
hPolκ-Damaged forward[a]	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAAGCGGCATCAGCAATGTTG-3′
hPolι-Control forward[a]	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTCCGGCATCAGCAATGTTG-3′
hPolι-Damaged forward[a]	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGACGGCATCAGCAATGTTG-3′
HT-SOSA reverse	5′-CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCCTATCCGCTCGCACAGCACG-3′
Illumina PCR 1[b]	5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′
Illumina PCR 2[b]	5′-CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT-3′
Illumina sequencing[b]	5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′

[a]Underlined nucleotides represent the unique barcode sequences that were used to identify each lesion bypass product origin.

Supplementary Table S2. Error rates of human Y-family DNA polymerases calculated by HT-SOSA analysis.

Enzyme	DNA	Event	Substitution Error Rate[a]	Substitution Error Ratio[b]	Insertion Error Rate[a]	Insertion Error Ratio[b]	Deletion Error Rate[a]	Deletion Error Ratio[b]
hPolη	17-mer/ 77-mer-ctl	Opposite TT[c]	6.6×10^{-2}		5.9×10^{-4}		3.5×10^{-3}	
		Upstream[d]	2.7×10^{-2}		7.6×10^{-4}		2.2×10^{-3}	
		Downstream[e]	2.0×10^{-2}		3.5×10^{-4}		2.4×10^{-3}	
		Total[f]	2.8×10^{-2}		5.6×10^{-4}		2.4×10^{-3}	
hPolη	17-mer/ 77-mer-TT	Opposite TT[c]	4.4×10^{-2}	0.7	1.6×10^{-3}	2.7	1.3×10^{-1}	37.1
		Upstream[d]	2.1×10^{-2}	0.8	5.4×10^{-4}	0.7	7.4×10^{-3}	3.4
		Downstream[e]	2.8×10^{-2}	1.4	5.0×10^{-4}	1.4	8.5×10^{-3}	3.5
		Total[f]	2.7×10^{-2}	1.0	6.3×10^{-4}	1.1	2.0×10^{-2}	8.3
hPolκ	17-mer/ 77-mer-ctl	Opposite TT[c]	1.7×10^{-2}		3.2×10^{-4}		1.9×10^{-3}	
		Upstream[d]	1.0×10^{-2}		4.5×10^{-4}		4.6×10^{-3}	
		Downstream[e]	1.5×10^{-2}		4.2×10^{-4}		2.7×10^{-3}	
		Total[f]	1.3×10^{-2}		4.2×10^{-4}		3.5×10^{-3}	
hPolκ	17-mer/ 77-mer-TT	Opposite TT[c]	1.8×10^{-1}	10.6	4.2×10^{-4}	1.3	1.3×10^{-1}	68.4
		Upstream[d]	7.4×10^{-3}	0.7	4.4×10^{-4}	1.0	8.2×10^{-3}	1.8
		Downstream[e]	1.7×10^{-2}	1.1	6.4×10^{-4}	1.5	5.2×10^{-3}	1.9
		Total[f]	2.9×10^{-2}	2.2	5.3×10^{-4}	1.3	1.9×10^{-2}	5.4
hPolι	17-mer/ 77-mer-ctl	Opposite TT[c]	4.3×10^{-1}		2.7×10^{-3}		7.5×10^{-2}	
		Upstream[d]	2.1×10^{-1}		2.0×10^{-3}		5.9×10^{-3}	
		Downstream[e]	1.1×10^{-1}		1.6×10^{-3}		8.7×10^{-3}	
		Total[f]	1.8×10^{-1}		1.9×10^{-3}		1.4×10^{-2}	
hPolι	17-mer/ 77-mer-TT	Opposite TT[c]	4.3×10^{-1}	1.0	1.1×10^{-3}	0.4	2.6×10^{-1}	3.5
		Upstream[d]	2.1×10^{-1}	1.0	1.2×10^{-3}	0.6	8.7×10^{-3}	1.5
		Downstream[e]	1.1×10^{-1}	1.0	1.7×10^{-3}	1.0	8.0×10^{-3}	0.9
		Total[f]	1.9×10^{-1}	1.0	1.4×10^{-3}	0.7	3.3×10^{-2}	2.4

[a]Calculated using Σ(specific mutation type)/[(number of sequences)×(number of bases in event)].

[b]Calculated using [Σ(specific mutation type)/[(number of sequences)×(number of bases in event)]]$_{77\text{-mer-TT}}$/[Σ(specific mutation type)/[(number of sequences)×(number of bases in event)]]$_{77\text{-mer-ctl}}$.

[c]Opposite TT events include all events at position -1 and +1 in Fig. 2.

[d]Upstream events include all events that occurred at positions -10 to -2, before an enzyme encountered the lesion.

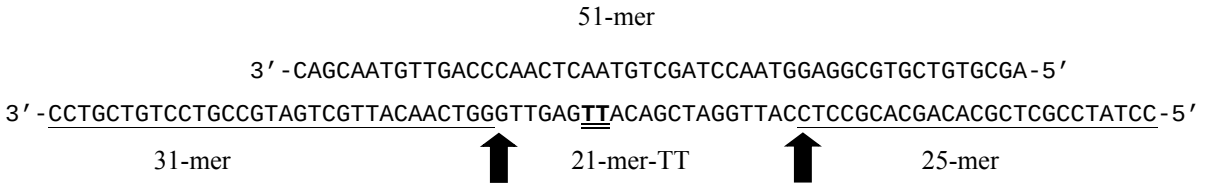
[e]Downstream events include all events that occurred at positions +2 to +10, after an enzyme encountered the lesion.

[f]Total events include all events from position -10 to +10 in Fig. 2.

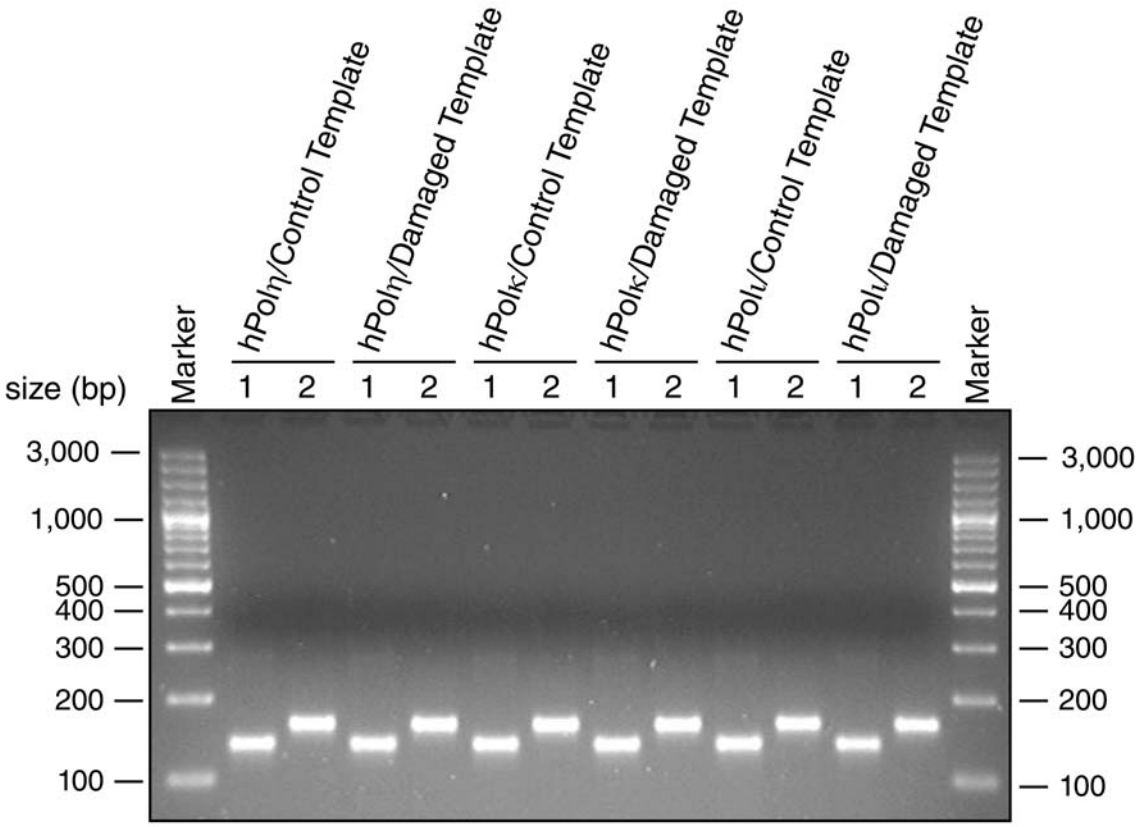
Supplementary Table S3. Number of sequences that contain select mutations opposite the *cis-syn* TT dimer site and the percent of the total sequences analyzed

Sequence[a]	hPolη		hPolκ		hPolι	
	Control template	Damage template	Control template	Damage template	Control template	Damage template
ACCCAACTC**AA**TGTCGATCC	1444163 (55.3%)	864828 (46.0%)	2664307 (72.7%)	1143058 (40.0%)	153639 (7.9%)	38124 (1.6%)
ACCCAACTC**cA**TGTCGATCC	2276 (0.1%)	13070 (0.7%)	3952 (0.1%)	50689 (1.8%)	12 (<0.1%)	21 (<0.1%)
ACCCAACTC**gA**TGTCGATCC	16718 (0.6%)	12535 (0.7%)	12642 (0.3%)	332141 (11.6%)	975 (<0.1%)	2443 (0.1%)
ACCCAACTC**tA**TGTCGATCC	2676 (0.1%)	12737 (0.7%)	2269 (<0.1%)	243830 (8.5%)	527 (<0.1%)	542 (<0.1%)
ACCCAACTC**Ac**TGTCGATCC	11795 (0.5%)	12039 (0.6%)	36656 (1.0%)	25039 (0.9%)	24 (<0.1%)	8 (<0.1%)
ACCCAACTC**Ag**TGTCGATCC	170416 (6.5%)	6642 (0.4%)	32188 (0.9%)	5475 (0.2%)	149 (<0.1%)	43 (<0.1%)
ACCCAACTC**At**TGTCGATCC	16151 (0.6%)	7572 (0.4%)	1957 (<0.1%)	5760 (0.2%)	49 (<0.1%)	140 (<0.1%)
ACCCAACTC**cc**TGTCGATCC	8 (<0.1%)	533 (<0.1%)	27 (<0.1%)	4202 (0.2%)	0 (0.0%)	0 (0.0%)
ACCCAACTC**gg**TGTCGATCC	196 (<0.1%)	259 (<0.1%)	8 (<0.1%)	382 (<0.1%)	14 (<0.1%)	32 (<0.1%)
ACCCAACTC**tt**TGTCGATCC	52 (<0.1%)	73 (<0.1%)	2 (<0.1%)	5624 (0.2%)	132 (<0.1%)	465 (<0.1%)
ACCCAACTC**AA**aGTCGATCC	14104 (0.5%)	40476 (2.1%)	20136 (0.6%)	3397 (0.1%)	15 (<0.1%)	33 (<0.1%)
ACCCAACTC**AA**cGTCGATCC	3156 (0.1%)	6121 (0.3%)	5197 (0.2%)	3557 (0.1%)	16 (<0.1%)	1 (<0.1%)
ACCCAACTC**AA**gGTCGATCC	4564 (0.2%)	17464 (0.9%)	7879 (0.3%)	9227 (0.3%)	7 (<0.1%)	17 (<0.1%)
ACCCAACTC**AA**-GTCGATCC	4687 (0.2%)	31291 (1.7%)	5898 (0.2%)	15625 (0.6%)	152 (<0.1%)	280 (<0.1%)
gCCCggCTC**gg**TGTCGgTCC	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (<0.1%)	9 (<0.1%)

[a]Incorporations opposite from the *cis-syn* TT dimer site are in bold. Correct incorporations as shown as upper case. Substitution mutations are shown as lower case. Deletion mutations are shown as a dash "-".



Supplementary Figure S1. Generation of DNA template 77-mer-TT. The arrows indicate the ligation sites of individual DNA oligomers. The *cis-syn* TT dimer is denoted in bold and double underlined as **TT**. The ligation oligomers, 31-mer and 25-mer, are underlined. The ligation annealing template, 51-mer, is on top. After annealing these four DNA oligomers together, T4 DNA ligase (Fermentas) was used to ligate the 31-mer and radiolabeled 25-mer to 21-mer-TT. The ligation products were then separated from the 51-mer via denaturing PAGE (8 M urea and 10% polyacrylamide). The desired 77-mer-TT was extracted from gel pieces, and desalted by using an activated 1-mL C18 column (Waters).



Supplementary Figure S2. Creation of next-generation sequencing libraries of the *cis-syn* TT-dimer bypass products. After PAGE purification, full-length products generated by the indicated polymerase and DNA substrate combination were PCR amplified with bar-coded primers (1) bearing the majority of the adapter sequences necessary for Illumina sequencing. These PCR products were gel purified and subsequently PCR amplified with primers containing the remaining adapter sequences (2). These PCR products were separated by using a 3% agarose gel and stained with ethidium bromide prior to imaging.

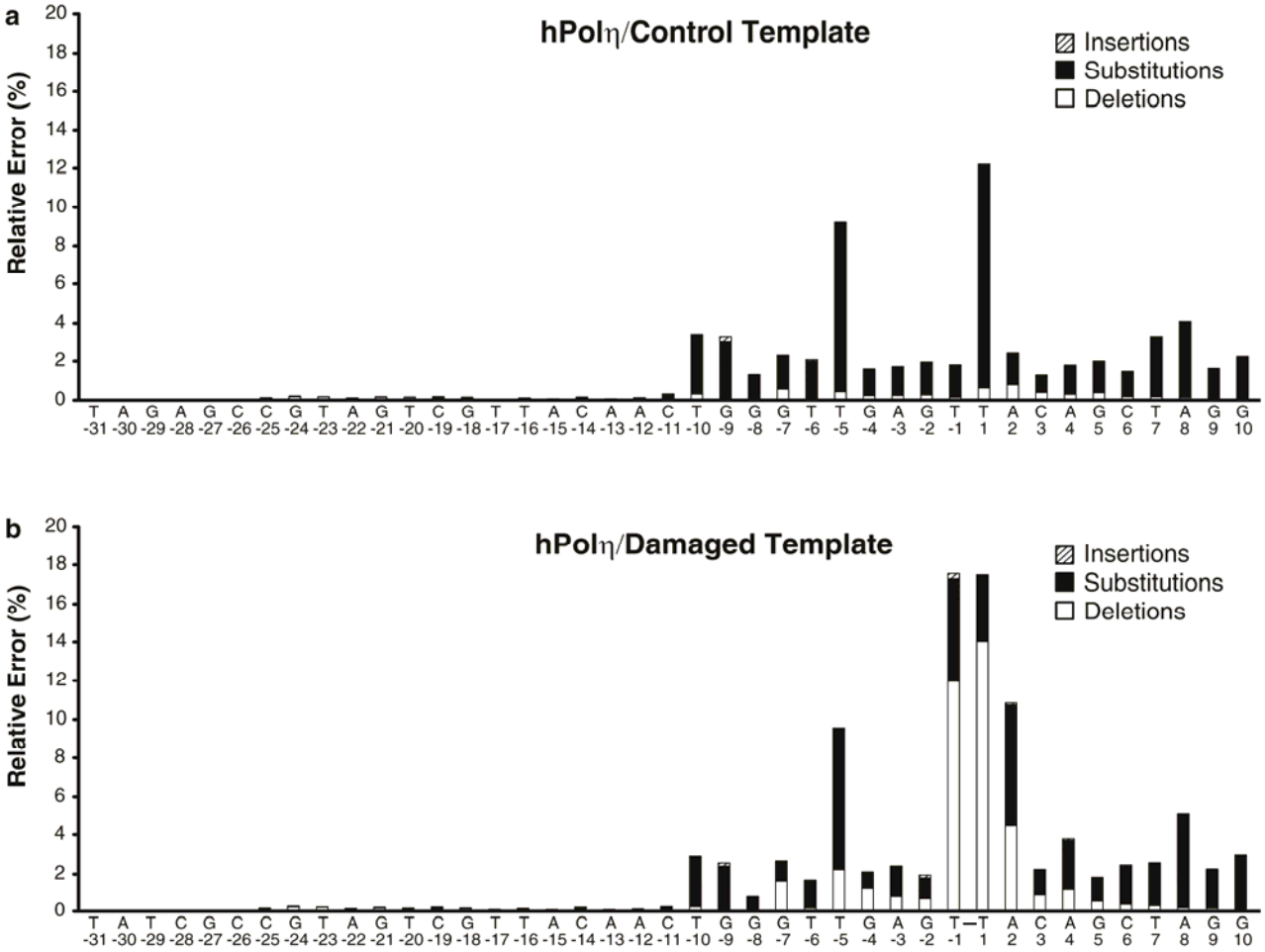
Barcode **21 bp synthesized by a Y-family polymerase**

XXXXCGGCATCAGCAATGTTGACCCAACTC**??**TGTCGATCCA

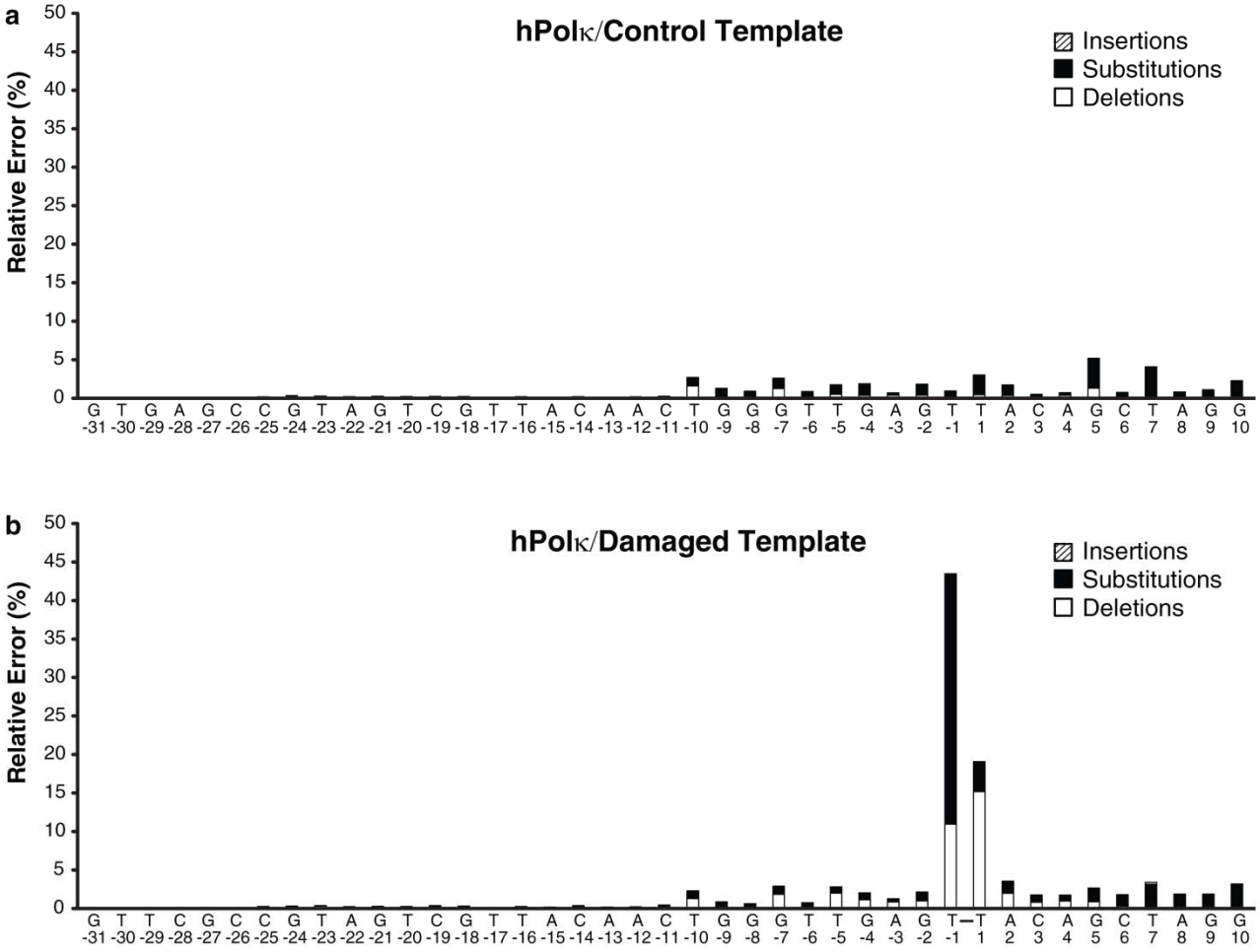
Derived from 17-mer primer

Enzyme	Template	Barcode	Sequences analyzed	% of total
hPolη	**17-mer/77-mer-ctl**	**ATCT**	**2,610,324**	**17.0**
hPolη	**17-mer/77-mer-TT**	**ATAG**	**1,883,244**	**12.3**
hPolκ	**17-mer/77-mer-ctl**	**CACT**	**3,663,552**	**23.8**
hPolκ	**17-mer/77-mer-TT**	**CAAG**	**2,863,446**	**18.6**
hPolι	**17-mer/77-mer-ctl**	**GCTC**	**1,949,239**	**12.7**
hPolι	**17-mer/77-mer-TT**	**TGGA**	**2,392,388**	**15.6**

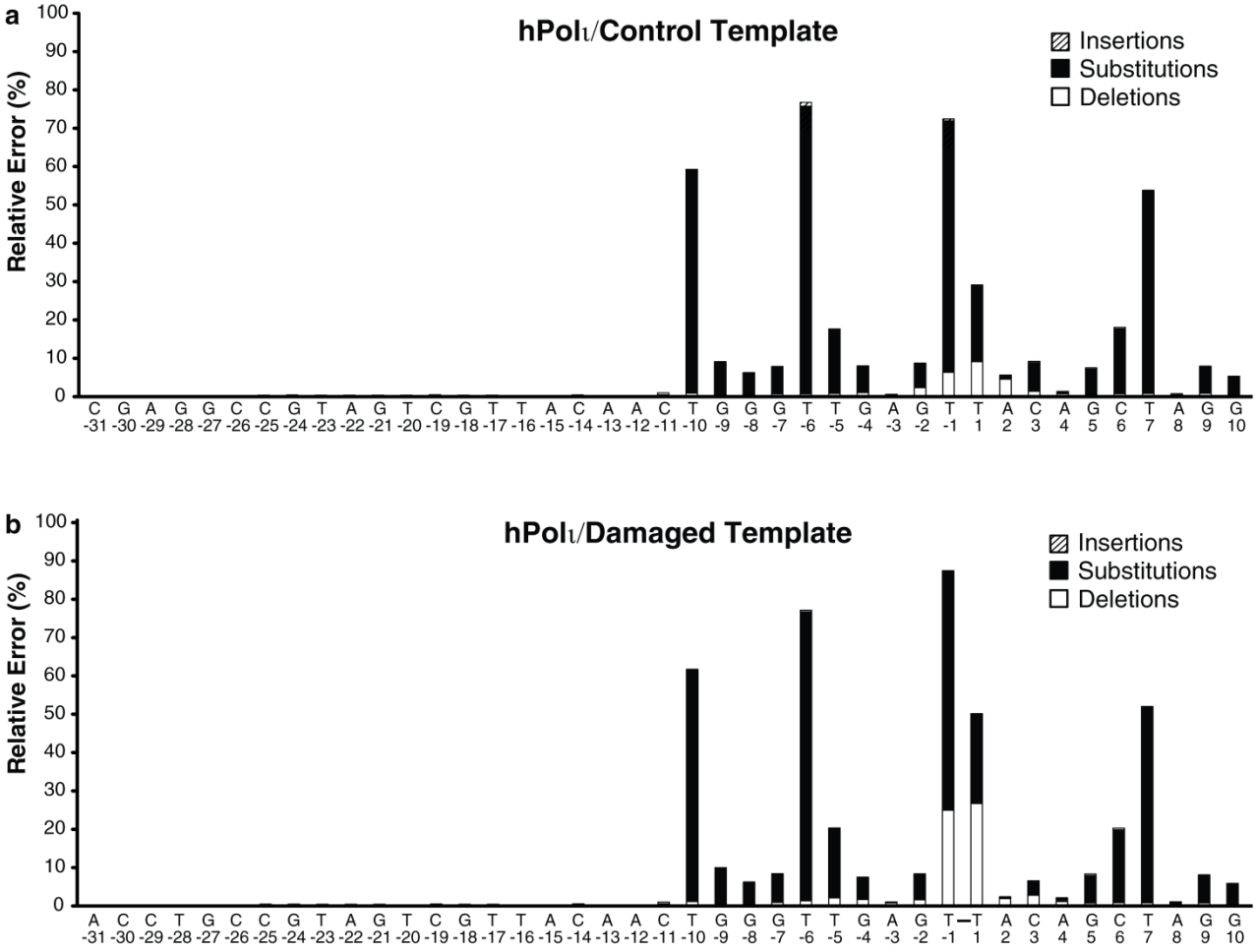
Supplementary Figure S3. The HT-SOSA sequences. The reference sequence for correct incorporation at all positions is depicted above. The location of the barcoded nucleotides are indicated by the bold **XXXX**. The sequence derived from the 17-mer primer and the 21 nucleotides originally synthesized by each Y-family DNA polymerase are also indicated. The locations of nucleotides incorporated opposite the *cis*-*syn* TT dimer are indicated by "??" in bold. The barcodes, total number of sequences analyzed and the percent of the total sequences analyzed are also shown below.



Supplementary Figure S4. Expanded histogram of the relative percent error of hPolη with (a) the undamaged control DNA substrate or (b) the *cis-syn* TT dimer-containing DNA substrate. This figure is an extension of Figures 2a and 2b. The relative number of base insertion mutations (striped bar), substitution mutations (black bar), and deletion mutations (white bar) as a percentage of the total dNTP incorporations are shown at each template position. The indicated template position is relative to the *cis-syn* TT dimer site within the 77-mer-TT template. The template bases are indicated and the *cis-syn* TT dimer is represented as T-T.



Supplementary Figure S5. Expanded histogram of the relative percent error of hPolκ with (a) the undamaged control DNA substrate or (b) the *cis-syn* TT dimer-containing DNA substrate. This figure is an extension of Figures 2c and 2d. The relative number of base insertion mutations (striped bar), substitution mutations (black bar), and deletion mutations (white bar) as a percentage of the total dNTP incorporations are shown at each template position. The indicated template position is relative to the *cis-syn* TT dimer site within the 77-mer-TT template. The template bases are indicated and the *cis-syn* TT dimer is represented as T-T.



Supplementary Figure S6. Expanded histogram of the relative percent error of hPolι with (a) the undamaged control DNA substrate or (b) the *cis-syn* TT dimer-containing DNA substrate. This figure is an extension of Figures 2e and 2f. The relative number of base insertion mutations (striped bar), substitution mutations (black bar), and deletion mutations (white bar) as a percentage of the total dNTP incorporations are shown at each template position. The indicated template position is relative to the *cis-syn* TT dimer site within the 77-mer-TT template. The template bases are indicated and the *cis-syn* TT dimer is represented as T-T.

```
Sequence Read ID: HWI-EAS435_100:5:1:2469:1040

Reference: CACTCGGCATCAGCAATGTTGACCCAACTCAA-TGTCGATCCA
Query:     CACTCGGCATCAGCAATGTTGACGCAACTCAACTGTCGATCC-

S@24G, I@33C, D@42
```

Scoring

- S@24G: Substitution of dG at base 24
- I@33C: Insertion of dC at base 33
- D@42: Deletion at base 42 – disregarded due to previous insertion

Supplementary Figure S7. Example of an audit file produced by the Next-Generation Sequencing Position Counter program. This example compares the reference sequence to a query sequence after alignment. The unique sequence read ID is displayed above each alignment. Initially the sequence is aligned with one substitution mutation at base 24, one insertion mutation at base 33 and one deletion mutation at base 42. However, because each sequence contains only 42 nucleotides of sequencing information, each insertion mutation within a quarry sequence generates a compensatory deletion at the 3′ end of the sequence after alignment. Therefore, the numbering used for tabulation purposes is adjusted for each insertion mutation and the resulting compensatory deletions at the 3′ end of the quarry sequences are disregarded.