Supporting Information for

"DNA Polymerase η Promotes the Transcriptional Bypass of N^2-Alkyl-2′-deoxyguanosine Adducts in Human Cells"

Ying Tan[a], Su Guo[a], Jun Wu[b], Hua Du[b], Lin Li[b], Changjun You[b] and Yinsheng Wang[a,b,*]

[a]Environmental Toxicology Graduate Program, University of California, Riverside, CA 92521-0403; and [b]Department of Chemistry, University of California, Riverside, CA 92521-0403.

SI Materials and Methods:

Syntheses of N^2-*n*-butyl-2′-deoxyguanosine (N^2-*n*Bu-dG) and [D_9]-N^2-*n*Bu-dG

N^2-*n*Bu-dG was synthesized from 2-fluoro-6-*O*-(trimethylsilylethyl)-2′-deoxyinosine (Scheme S1). Briefly, 2-fluoro-6-*O*-(trimethylsilylethyl)-2′-deoxyinosine (4.0 mg, 0.01 mmol) was dissolved in 80 μL dimethyl sulfoxide (DMSO), to which solution were added *n*-butylamine (5.0 μL, 0.05 mmol) and anhydrous *N*,*N*-diisopropylethylamine (DIEA, 3.8 μL, 0.02 mmol). The reaction mixture was stirred at 55°C for 3 days. The resulting mixture was dried *in vacuo*, reconstituted in 0.1 M acetic acid (100 μL) to deprotect the trimethylsilylethyl group. The solution was neutralized with an equal volume of 0.1 M sodium bicarbonate. The crude product was dried *in vacuo* and redissolved in 1 mL doubly distilled water.

A similar approach was employed for the preparation of [D_9]-N^2-*n*Bu-dG (Scheme S2). In brief, 2-fluoro-6-*O*-(trimethylsilylethyl)-2′-deoxyinosine (5.7 mg, 0.015 mmol) was dissolved in 80 μL DMSO, to which solution were added D_9-*n*-butylamine (Toronto Research Chemicals, 37.9 μL, 0.045 mmol) dissolved in 100 μL DMSO and anhydrous DIEA (5.4 μL, 0.030 mmol). The reaction mixture was allowed to stir at 55 °C for 5 days. The resulting mixture was subsequently dried *in vacuo*. To the mixture was subsequently added 0.1 M acetic acid (100 μL) and the mixture was stirred at room temperature for 24 h. The solution was again neutralized with sodium bicarbonate, and dried *in vacuo*.

HPLC purification of N^2-*n*Bu-dG and [D_9]-N^2-*n*Bu-dG

An ODS-silica column (10 × 250 mm, 10 μm in particle size, 100 Å in pore size, Soochow High Tech Chromatography Co.) was employed to purify N^2-*n*Bu-dG and [D_9]-N^2-*n*Bu-dG. Doubly distilled water and methanol were selected as mobile phases A and B, respectively. The gradient compromised of 10% B at 0-2 min and 10-80% B at 2-75 min, with the flow rate being 2.5 mL/min. The purified N^2-*n*Bu-dG and [D_9]-N^2-*n*Bu-dG were confirmed by ESI-MS and MS/MS analyses (Figure S8).

List of Supporting Schemes and Figures:

Scheme S1. Synthesis of N^2-*n*Bu-dG from 2-fluoro-6-*O*-(trimethylsilylethyl)-2′-deoxyinosine.

Scheme S2. Synthesis of [D_9]-N^2-*n*Bu-dG. Asterisks represent those hydrogens that were replaced with deuterons.

Figure S1. Restriction digestion and post-labeling method for determining the transcriptional bypass efficiencies and mutation frequencies of N^2-alkyl-dG lesions in TLS polymerase-deficient cells.

Figure S2. Higher-resolution "zoom" scan ESI-MS for monitoring the $[M-3H]^{3-}$ ions of the restriction fragments of interest from the transcription of non-lesion control (A) and N^2-alkyl-dG lesions (B-E) in HEK293T cells.

Figure S3. LC-MS/MS for monitoring the restriction digestion products of interest corresponding to wild-type or mutated transcripts arising from N^2-alkyl-dG containing substrates.

Figure S4. The temporal transcriptional alterations of N^2-*n*Pr-dG in parental and TLS polymerase-depleted HEK 293T cells. The N^2-*n*Pr-dG-containing plasmids and competitor plasmids were transfected to HEK 293T cells or CRISPR-Cas9 knockout cells. The cells were then harvested after 3 h, 8 h or 24 h for RNA extraction and CTAB assay. (AB) The representative gel images of restriction fragments arising from RT-PCR products at different timepoints. (C) Relative bypass efficiencies (RBE) of N^2-*n*Pr-dG at 3 h, 8 h and 24 h. (D) Relative mutation frequencies (RMF) of N^2-*n*Pr-dG at 8 h.

Figure S5. Genotyping (a) and Western-blot verification (b) of *REV1*/*POLH* double-knockout cells generated with CRISPR/Cas9-based genome editing of *REV1* gene in Pol η-depleted HEK293T cells. Insertion/deletion mutations of the genes encoding translesion synthesis DNA polymerases (i.e., *POLH* and *REV1*, which encode DNA polymerases η and Rev1, respectively) were identified by Sanger sequencing. The guide sequence for human REV1 gene with the PAM motif being underlined was: ATCAGATGCTGCTATGCAGAAGG. The lack of Rev1 polymerase in *REV1*/*POLH* double-knockout cells was verified by Rev1 antibody.

Figure S6. Agarose (1%) gel electrophoresis for monitoring the qualities of purified gapped vector (A) and the fully ligated supercoiled N^2-*n*Pr-dG-containing vector (B). Lane 1 & 5, 1 kb plus DNA ladder. Lane 2, control vector without lesion. Lane 3, gapped vector generated by Nt.BstNBI restriction digestion on control vector and removal of a 25-mer (details in figure 2b). Lane 4, self-ligation test to ensure completely removal of the 25-mer by incubating gapped vector with T4 DNA ligase and ATP. Lane 6, ligation reaction mixture where the gapped vector was annealed with the 12-mer N^2-*n*Pr-dG-containing insert and a 13-mer ODN and then ligated with T4 DNA ligase

and ATP.

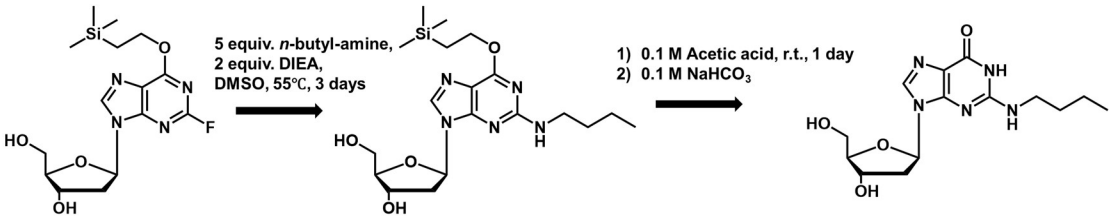
Figure S7. Representative agarose gel image of RT-PCR negative controls (lane 1-15) and normal RT-PCR products (lane 16) in *POLH*$^{-/-}$ cells. The negative controls are done by direct PCR amplification from RNA samples without reverse-transcription process to ensure the RNA samples are free of DNA contamination. Lane 1 & 11, 1 kb plus DNA ladder. Lane 2-4, negative control for 3 biological replicates of control substrate. Lane 5-7, negative control for 3 biological replicates of N^2-Me-dG substrate. Lane 8-10, negative control for 3 biological replicates of N^2-Et-dG substrate. Lane 12-14, negative control for 3 biological replicates of N^2-*n*Pr-dG substrate. Lane 15-17, negative control for 3 biological replicates of N^2-*n*Bu-dG substrate. Lane 18, RT-PCR amplification products of control substrate in *POLH*$^{-/-}$ cells (500 bps).

Figure S8. Positive-ion ESI-MS (insets) and MS^2 of N^2-*n*Bu-dG (A) and [D_9]-N^2-*n*Bu-dG (B).

Figure S9. Representative selected-ion chromatograms (SICs) for monitoring *m/z* 324 → 208 (top panel) and 333 → 217 (bottom panel) transitions for the protonated ions of N^2-*n*BudG and its corresponding isotope-labeled counterpart.

Figure S10. The calibration curve for the quantification of N^2-*n*Bu-dG.

Scheme S1. Synthesis of N^2-*n*Bu-dG from 2-fluoro-6-*O*-(trimethylsilylethyl)-2′-deoxyinosine.



Scheme S2. Synthesis of [D_9]-N^2-*n*Bu-dG. Asterisks represent those hydrogens that were replaced with deuterons.

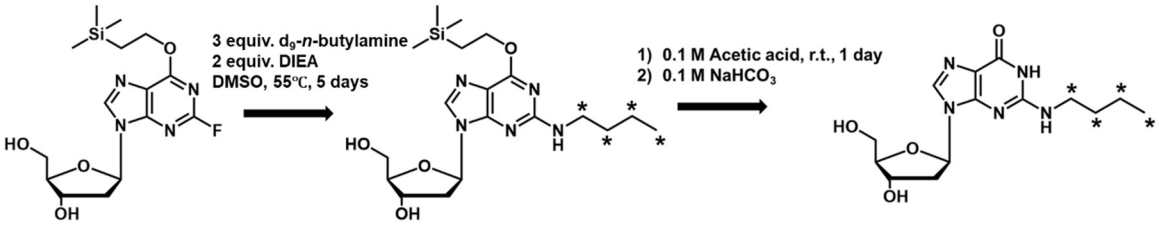


Figure S1. Restriction digestion and post-labeling method for determining the transcriptional bypass efficiencies and mutation frequencies of N^2-alkyl-dG lesions in TLS polymerase-deficient cells. (A) Digestion of RT-PCR products with NcoI and SfaNI restriction endonucleases and post-labeling assay. 'p*' denotes a ^{32}P-labeled phosphate group. 'M' represents the site where the lesion was initially situated. 'P' indicated the nucleobase 5′ to the lesion site. 'N' and 'Q' are the complementary bases paired with 'M' and 'P', respectively. Representative gel images for monitoring the restriction fragments of interest in cells that are deficient in both Pol η and Rev1 (B) or only in Pol ι , Pol κ, Pol ζ or REV1 (C-D). The lesion-containing plasmids were individually premixed with the competitor plasmid at a molar ratio of 2:1 (lesion/competitor) for transfection into HEK293T, *POLK*$^{-/-}$, *POLI*$^{-/-}$ and *REV3L*$^{-/-}$ cells, and 5:1 for *REV1*$^{-/-}$ cells. The restriction fragment arising from competitor vector, i.e., d(CATGGCGATAGGCTAT), is designated as '16mer'; '13 mer G', '13 mer A', '13 mer C', '13 mer T', '13 mer 5′T' and '13 mer TT' represent the standard synthetic ODNs d(CATGGCPMGCTAT), where 'PM' represent 'GG', 'GA', 'GC', 'GT', 'TG' and 'TT', respectively.

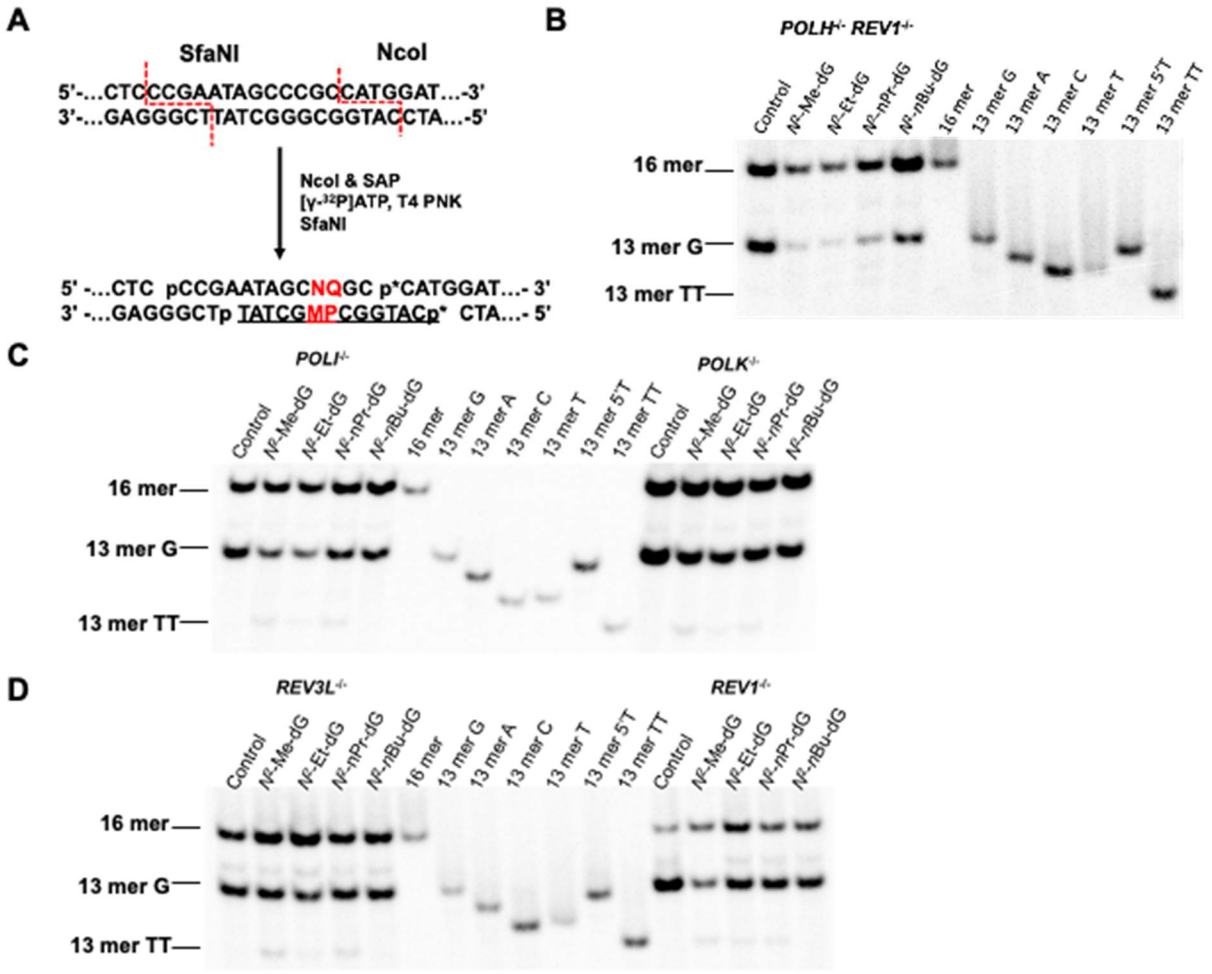


Figure S2. Higher-resolution "zoom" scan ESI-MS for monitoring the $[M-3H]^{3-}$ ions of the restriction fragments of interest from the transcription of non-lesion control (A) and N^2-alkyl-dG lesions (B-E) in HEK293T cells. 'WT' represents the 5′-phosphorylated opposite strand of non-mutagenic product, i.e., d(pCCGAATAGCCCGC), whereas 'TT' and 'T' designate the lesion-containing strand that carry GG→TT and G→T mutation, respectively, i.e. , d(CATGGCTTGCTAT) and d(CATGGCGTGCTAT).

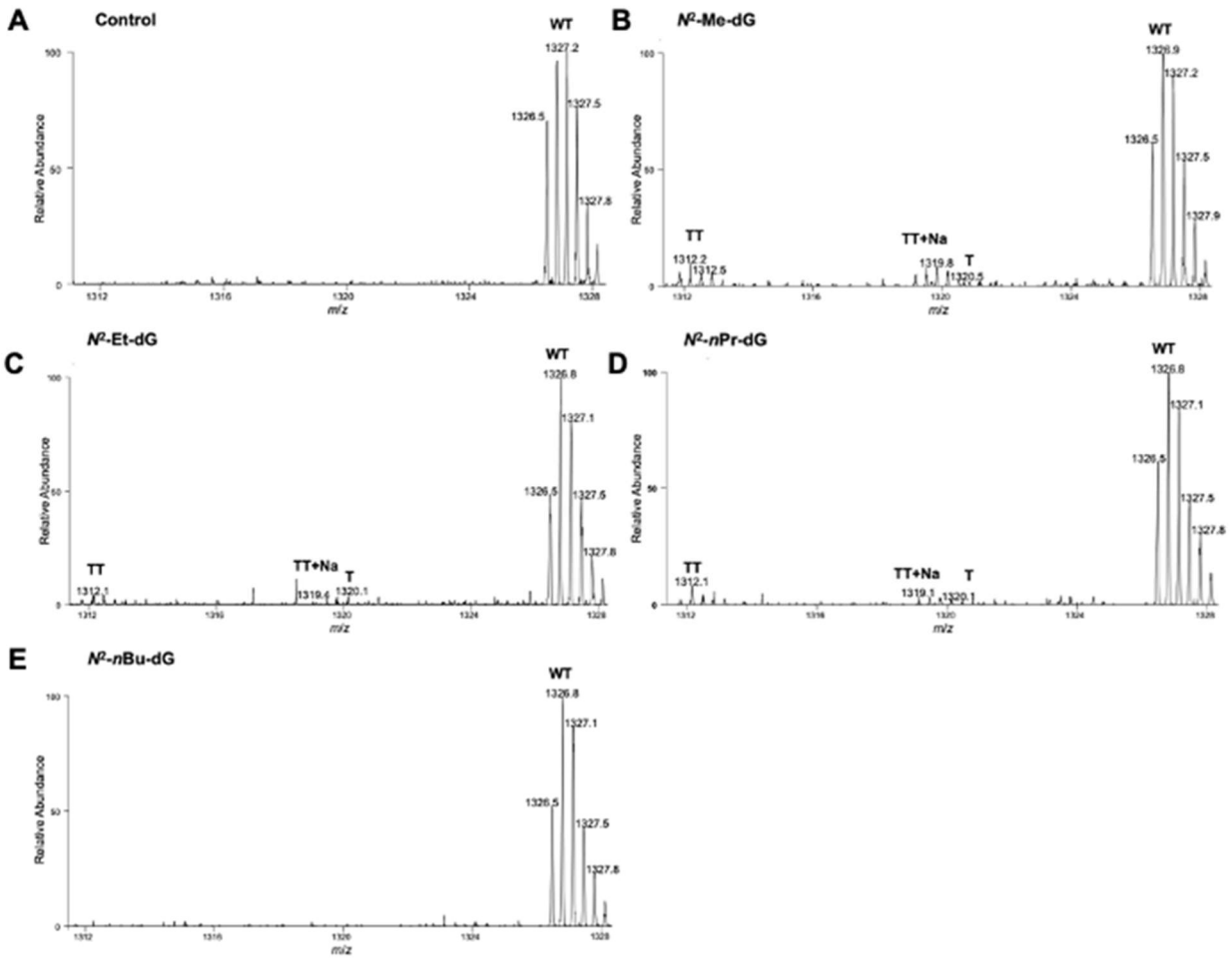
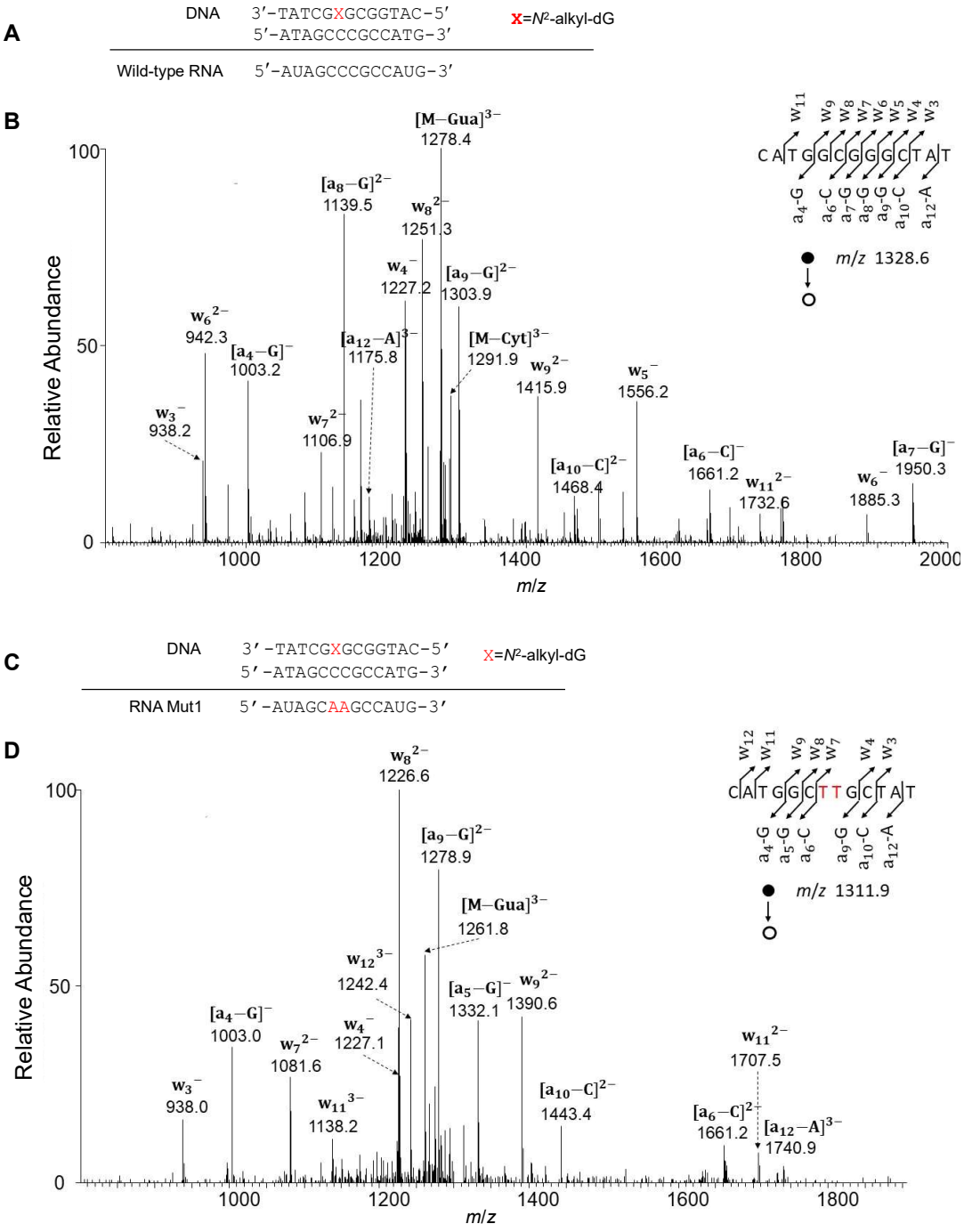


Figure S3. LC-MS/MS for monitoring the restriction digestion products of interest corresponding to wild-type or mutated transcripts arising from N^2-alkyl-dG containing substrates in HEK293T cells. The sequence information of wild-type (A) or mutated (C,E) transcripts are indicated below the double-stranded DNA construct. Shown in (B), (D) and (F) are the product-ion spectra of the ESI-produced $[M-3H]^{3-}$ ions (*m/z* 1328.6, 1311.9 and 1320.2) of the 13-mer fragments of the wild-type sequence [i.e., d(CATGGCGGGCTAT)], 5′-AA mutant sequence [i.e., d(CATGGCTTGCTAT)] and 5′A mutant sequence [i.e., d(CATGGCTGGCTAT)], respectively.



S

E

DNA 3′-TATCGXGCGGTAC-5′ X=N^2-alkyl-dG
5′-ATAGCCCGCCATG-3′

RNA Mut2 5′-AUAGCCAGCCAUG-3′

F

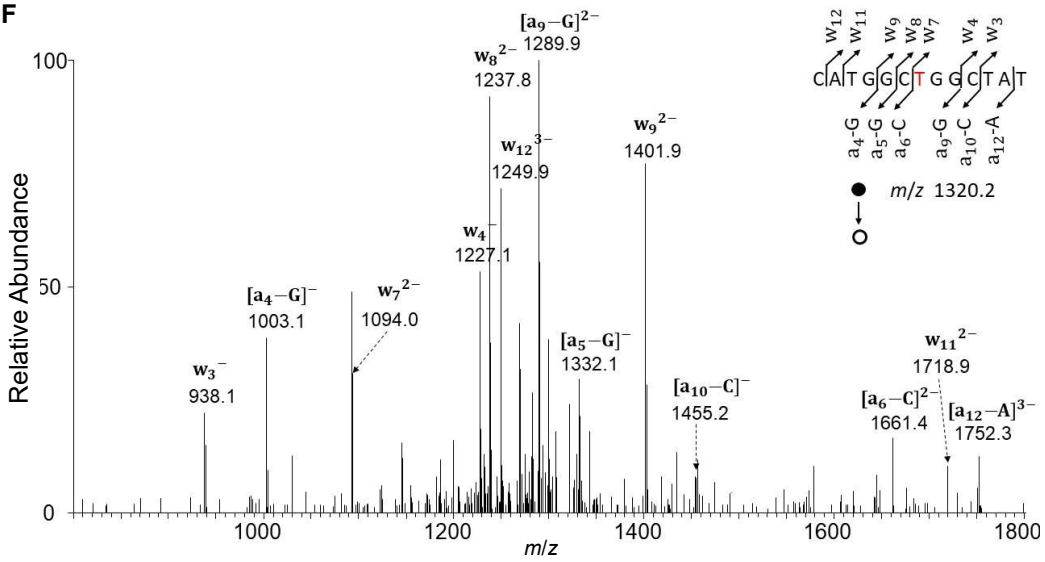


Figure S4. The time-dependent changes in transcriptional alterations of N^2-*n*Pr-dG in parental and TLS polymerase-depleted HEK 293T cells. The N^2-*n*Pr-dG-containing plasmid or control dG-carrying plasmid were mixed individually with the competitor plasmid at defined molar ratios and transfected to HEK293T or the isogenic polymerase knockout cells. The cells were then harvested after 3 h, 8 h or 24 h for RNA extraction and CTAB assay. The lesion-containing plasmids were individually premixed with the competitor plasmid at a molar ratio of 2:1 (lesion/competitor) for transfection into HEK293T cells, 3:1 for *POLH*$^{-/-}$ cells, 5:1 for *POLH*$^{-/-}$/*REV1*$^{-/-}$ cells. (A-B) Representative gel images of restriction fragments arising from RT-PCR products at different time points. (C) Relative bypass efficiencies (RBE) of N^2-*n*Pr-dG at 3 h, 8 h and 24 h. (D) Mutation frequencies (MF) of N^2-*n*Pr-dG at 8 h. The data represent the mean ± S.D. of results obtained from three independent experiments. *, $0.01 < p < 0.05$; **, $0.001 < p < 0.01$; ***, $0.0001 < p < 0.001$; ****, $p < 0.0001$; ns, $p > 0.05$. The multiplicity adjusted *p* values were calculated by using one-way ANOVA and Tukey's multiple comparisons test for the comparisons between different time points within the same cell type, and between different cell types at the same time points.

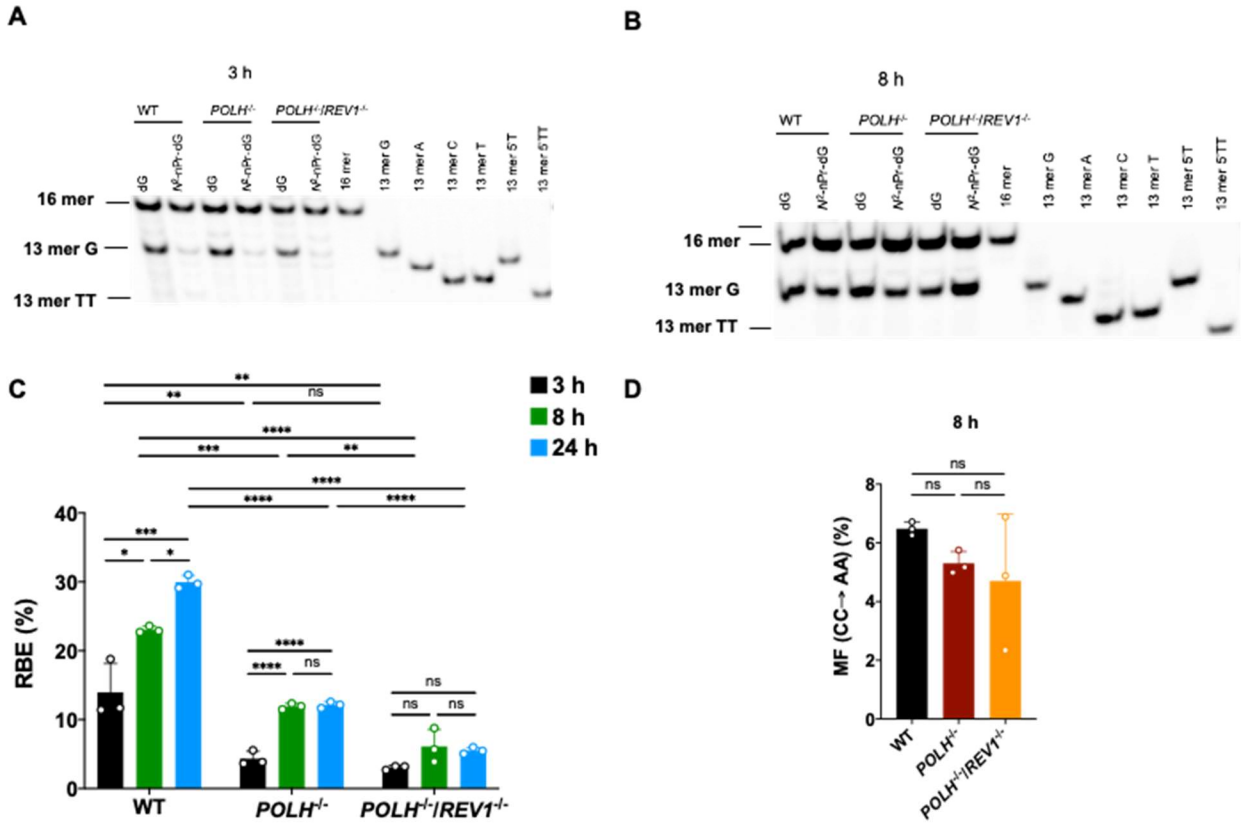


Figure S5. Genotyping of *REV1*/*POLH* double-knockout cells generated with CRISPR/Cas9-based genome editing of *REV1* gene in Pol η-depleted HEK293T cells. (A) Insertion/deletion mutations of the genes encoding translesion synthesis DNA polymerases (i.e., *POLH* and *REV1*, which encode DNA polymerases η and REV1, respectively) were identified by Sanger sequencing. (B) The successful deletion of REV1 in Pol η-deficient background was confirmed by Western blot analysis. The guide sequence for human REV1 gene with the PAM motif being underlined was: ATCAGATGCTGCTATGCAGAAGG.

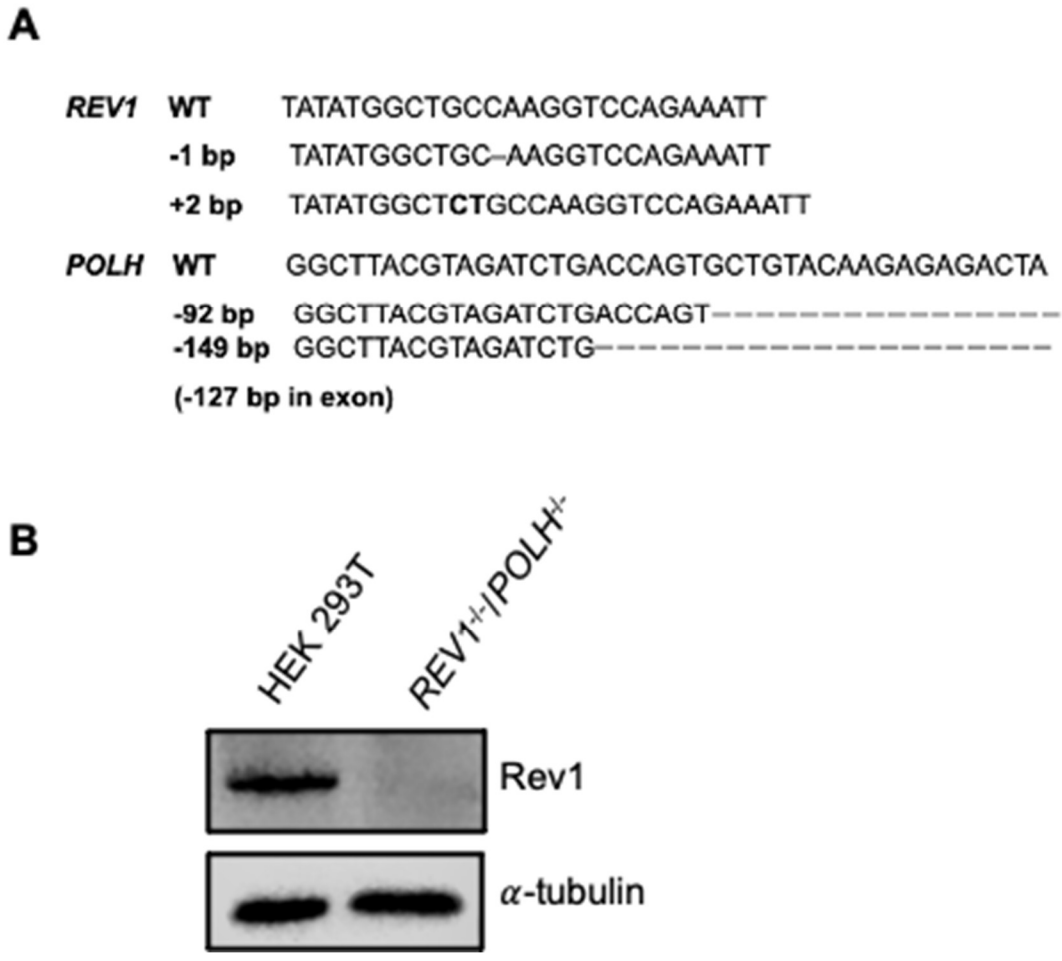


Figure S6. Agarose (1%) gel electrophoresis for monitoring the qualities of purified gapped vector (A) and the fully ligated supercoiled N^2-*n*Pr-dG-containing vector (B). Lane 1 & 5, 1 kb plus DNA ladder; Lane 2, control vector without lesion; Lane 3, gapped vector generated after Nt.BstNBI restriction digestion of the control vector and removal of a 25-mer restriction fragment (details shown in Figure 2b). Lane 4, self-ligation test to ensure complete removal of the 25-mer by incubating gapped vector with T4 DNA ligase and ATP in the absence of insert. Lane 6, ligation reaction mixture where the gapped vector was annealed with the 12-mer N^2-*n*Pr-dG-containing insert and a 13-mer ODN and then ligated with T4 DNA ligase and ATP.

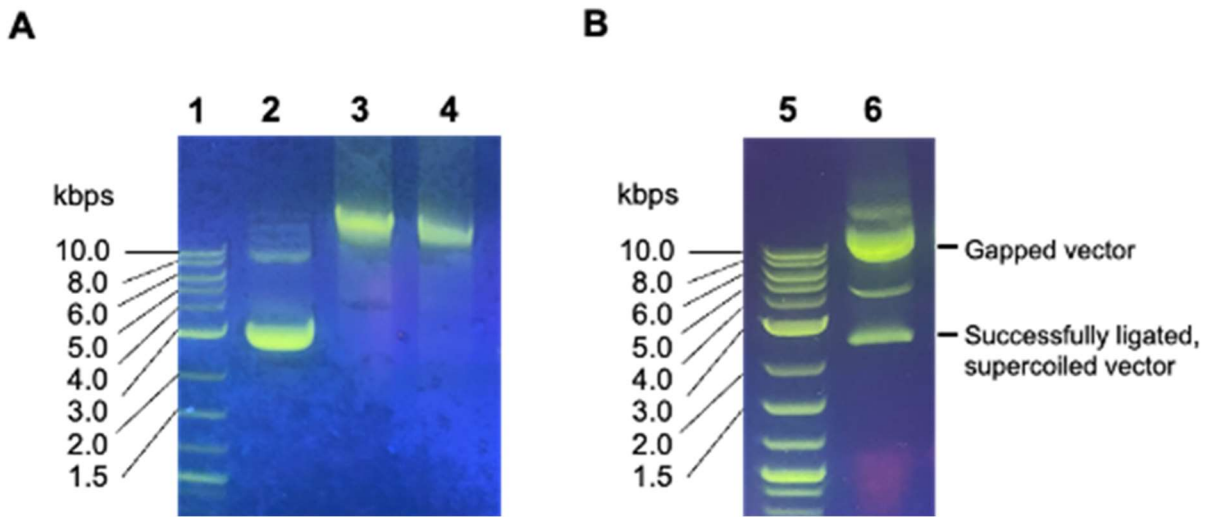


Figure S7. Representative agarose gel image of RT-PCR negative controls (lane 1-15) and normal RT-PCR products (lane 16) of transcripts isolated from *POLH*$^{-/-}$ cells. The negative controls are done by direct PCR amplification from RNA samples without reverse-transcription to ensure that the RNA samples are free of DNA contamination. Lane 1 & 11, 1 kb plus DNA ladder. Shown also are negative controls for 3 biological replicates RNA products of: the control dG-containing substrate (Lanes 2-4); N^2-Me-dG substrate (Lanes 5-7); N^2-Et-dG substrate (Lanes 8-10); N^2-*n*Pr-dG substrate (Lanes 12-14); and N^2-*n*Bu-dG substrate (Lanes 15-17). Lane 18, PCR amplification products of RNA samples of control dG-containing substrate isolated from in *POLH*$^{-/-}$ cells, where reverse transcription reaction was conducted prior to PCR amplification; the expected 500 bp PCR amplicon was only detected for samples with reverse transcription.

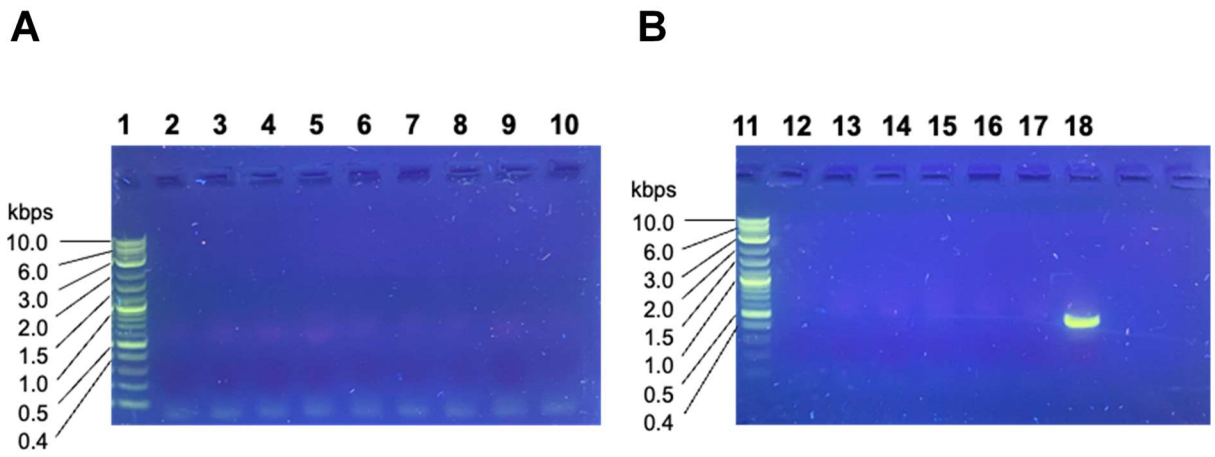


Figure S8. Positive-ion ESI-MS (insets) and MS^2 of N^2-*n*Bu-dG (A) and [D_9]-N^2-*n*Bu-dG (B).

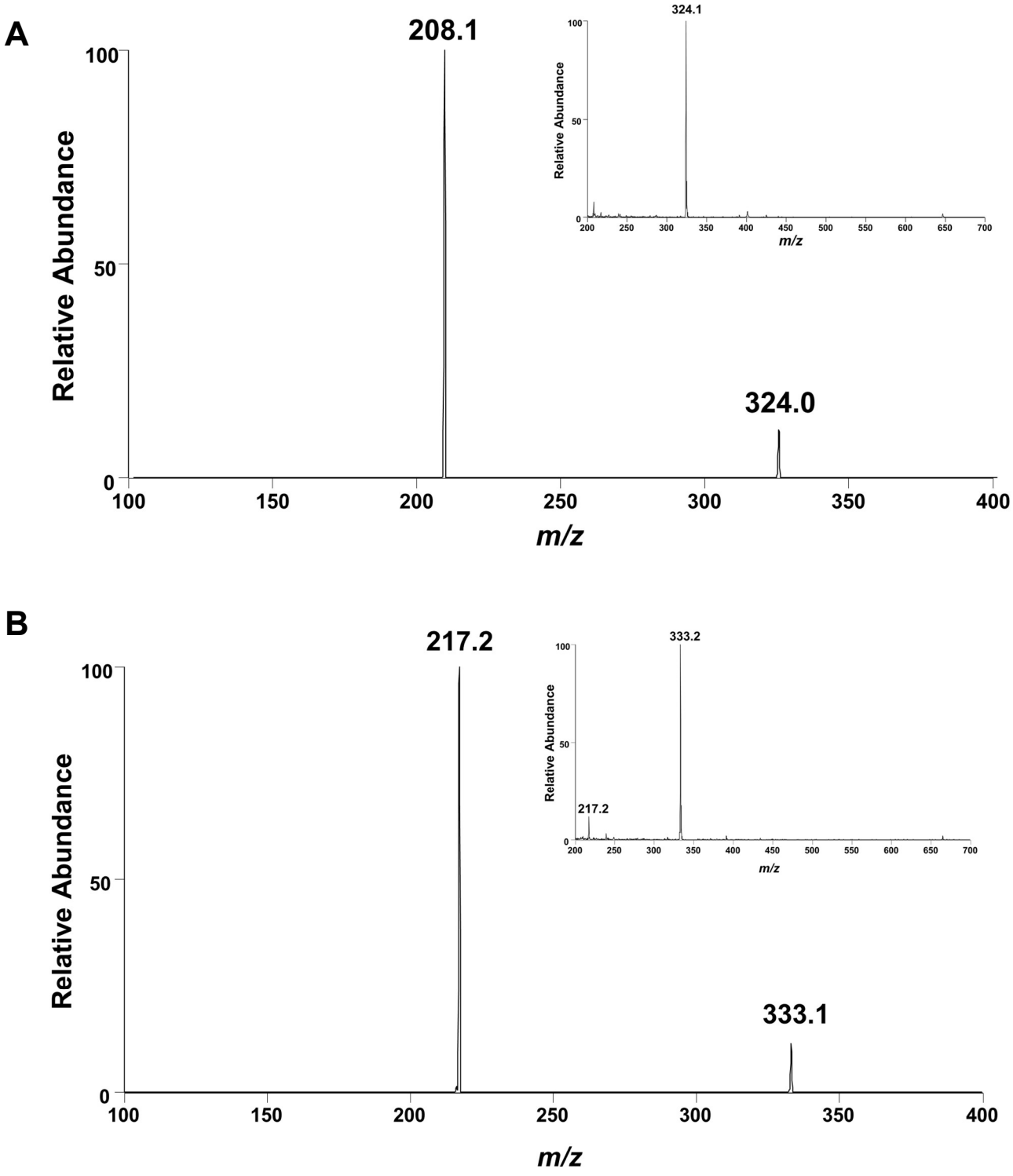


Figure S9. Representative selected-ion chromatograms (SICs) for monitoring *m/z* 324 →208 (top panel) and 333 → 217 (bottom panel) transitions for the protonated ions of N^2-*n*BudG and its corresponding isotope-labeled counterpart.

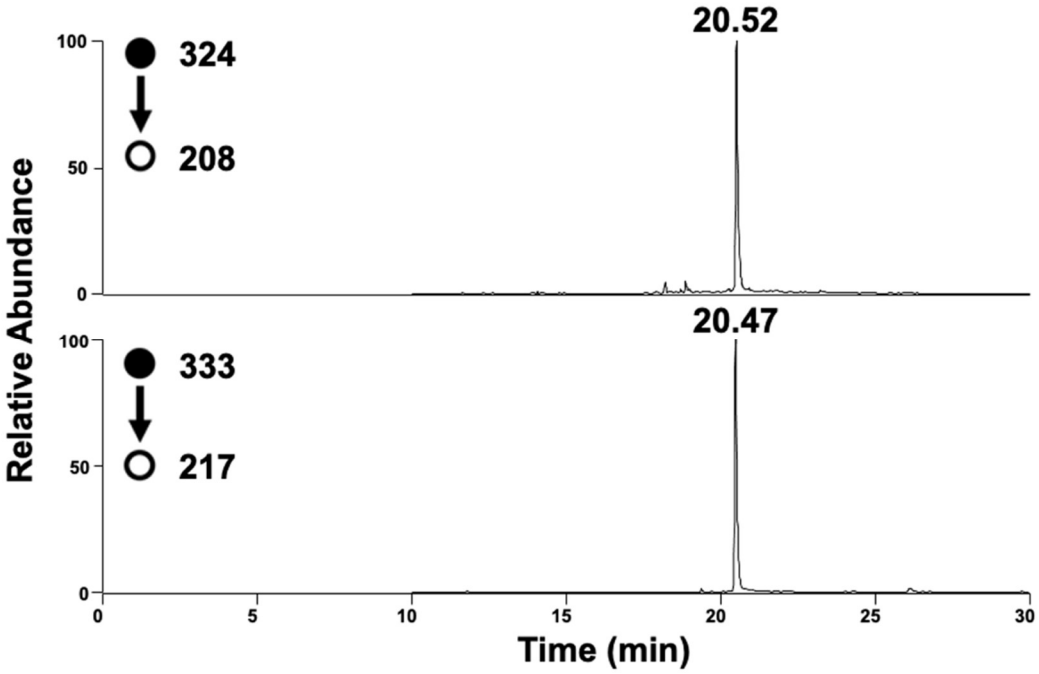


Figure S10. The calibration curve for the quantification of N^2-*n*Bu-dG. The samples used for the construction of the calibration curve compromised 1 μg calf thymus DNA, different amounts (20, 40, 80, 200, 400, 800 and 2000 fmol) of an N^2-*n*Bu-dG containing oligodeoxyribonucleotide (ODN, 5′-ATGGCGXGCTAT-3′, where 'X' represents N^2-*n*Bu-dG) and a fixed amount of [D_9]-N^2-*n*BudG (400 fmol). The calibration curve was constructed by plotting the peak area ratios found in the selected-ion chromatograms for the unlabeled/labeled N^2-*n*Bu-dG vs. the molar ratios of the unlabeled/labeled N^2-*n*Bu-dG.