SUPPLEMENTARY INFORMATION

A quantitative assay for assessing the effects of DNA lesions on transcription

Changjun You[1], Xiaoxia Dai[1], Bifeng Yuan[1], Jin Wang[1], Jianshuang Wang[1], Philip J. Brooks[2], Laura J. Niedernhofer[3], Yinsheng Wang[1,*]

[1]Department of Chemistry, University of California, Riverside, CA 92521-0403

[2]Laboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism, Rockville, MD 20852

[3]Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, 523 Bridgeside Point II, 450 Technology Drive, Pittsburgh, PA 15219.

*To whom correspondence should be addressed: Tel. (951)827-2700; E-mail: Yinsheng.Wang@ucr.edu

SUPPLEMENTARY METHODS

Materials and cell culture conditions

Unmodified oligodeoxynucleotides (ODNs) used in this study were purchased from Integrated DNA Technologies (Coralville, IA). [γ-^{32}P]ATP and [α-^{32}P]-CTP were obtained from Perkin Elmer, and chemicals unless otherwise noted were from Sigma-Aldrich. Shrimp alkaline phosphatase was obtained from USB Corporation, and all other enzymes unless otherwise specified were from New England Biolabs.

The 293T human embryonic kidney epithelial cells were obtained from ATCC. The SV40-transformed GM00637 (repair-proficient) and GM04429 (XPA-deficient) human fibroblast cell lines were kindly provided by Prof. Gerd P. Pfeifer. These cells were cultured at 37 ℃ in 5% CO_2 atmosphere and in Dulbecco's Modified Eagle's medium supplemented with 10% fetal bovine serum (Invitrogen), 100 U/mL penicillin, and 100 μg/mL streptomycin (ATCC). The AA8 (wild-type) and CHO-7-27 (ERCC1-deficient) [1] Chinese hamster ovary cells were provided by Prof. Michael Seidman and grown in Eagle's minimal medium containing 10% fetal bovine serum.

Transcription template construction

By employing the PCR-based site-directed mutagenesis, we first prepared the lesion-free control and competitor vectors using the non-replicating pTGFP-Hha10 plasmid as the original template [2]. The PCR was performed with Phusion high-fidelity DNA polymerase (New England Biolabs), and the PCR amplification started at 98 ℃ for 30 s, followed with 18 cycles at 98 ℃ for 15 s, 60 ℃ for 30 s, and 72 ℃ for 2.5 min, and a 10 min final extension at 72 ℃. A pair of primers (PTGFP-T7-F/-R) were used for the PCR insertion of the T7 promoter sequence into the modified vector to render pTGFP-T7-Hha10, which was used as the control plasmid for *R*- and *S*-N^2-CEdG. Likewise, we prepared a competitor vector for *R*- and *S*-N^2-CEdG by employing the control plasmid as the PCR template and another pair of primers (pTGFP-Comp-1F/-1R). With the use of pTGFP-T7-Hha10 as the template, two pairs of primers (pTGFP-GA-F/-R and pTGFP-GG-F/-R) were separately used for the second round of PCR and a 12-mer sequence, d(ATGGCGXGCTAT) (where 'X' is A or G), was therefore placed between two unique Nt.BstNBI sites of the original vector. These newly constructed vectors were referred to as

pTGFP-T7-GA and pTGFP-T7-GG, and were used as control plasmids for *S*-cdA and *S*-cdG, respectively. The competitor vector for *S*-cdA and *S*-cdG was also constructed by using the pTGFP-T7-Hha10 as the PCR template and one pair of primers (pTGFP-Comp-2F/-2R). The primers used for site-directed mutagenesis are listed in **Supplementary Table 1**, and all the constructed control and competitor plasmids were confirmed by DNA sequencing.

Constructs containing *S*-cdA, *S*-cdG, *R*-N^2-CEdG or *S*-N^2-CEdG were prepared following previously described procedures [2,3]. For subsequent experiments, the supercoiled control and competitor plasmids were isolated from the mixture by agarose gel electrophoresis. Purified lesion-containing or the corresponding lesion-free (control) genome was normalized against the competitor genome as previously described[4-6], and they were subsequently mixed with the competitor genome at given molar ratios as the template for *in vitro* and *in vivo* transcription experiments.

T7 RNAP-mediated transcription with radiolabeling

The lesion-bearing or undamaged control plasmids were digested with EcoRI while the competitor genome was treated with SalI to give linear DNA templates. The EcoRI-treated lesion-bearing or undamaged control genomes were then mixed with the SalI-treated competitor genome at indicated molar ratios and used as DNA template for T7 RNAP-mediated transcription with radiolabeling. The reactions contained 50 ng of DNA template, 0.5 mM each of the three non-radioactive ribonucleotides (ATP, GTP and UTP), 20 μM non-radioactive CTP, 10 μCi [α-^{32}P]CTP (3,000 Ci/mmol) and 20 U T7 RNAP (Promega) in a final volume of 20 μL. The reaction mixture was incubated at 37 ℃ for 60 min, following the manufacturer's recommended procedures. The resulting transcripts were directly separated on a 16% polyacrylamide gel (acrylamide:bis-acrylamide=19:1) with 8 M urea and quantified by phosphorimager analysis with a Typhoon 9410 Variable Mode Imager and ImageQuant software (GE Healthcare).

LC-MS/MS

The ODN mixture was subjected to LC-MS/MS analysis on an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific). A 0.5×250 mm Zorbax SB-C18 column (particle size, 5 μm, Agilent) was used for the separation of the above-enriched lesion fractions, and the flow rate was 8.0 μL/min, which was delivered by using an Agilent 1100 capillary HPLC pump (Agilent

Technologies). A 5-min gradient of 5-35% methanol followed by a 15 min of 35-95% methanol in 400 mM 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP) buffer (pH was adjusted to 7.0 by the addition of triethylamine) was employed.

Real-time quantitative RT-PCR and Western blot analysis

Relative quantification of gene expression was performed on a Bio-Rad iCycler system (Bio-Rad). The primers for *XPC, CSB* and *GAPDH* genes are listed in **Supplementary Table 2**. Real-time PCR was performed in an optical 96-well plate, including iQ SYBR Green Supermix kit (Bio-Rad), 1 μL of the cDNA sample and 0.2 μM of each gene-specific primers, in a final volume of 25 μL. The reactions followed the temperature profile of 95 ℃ for 3 min; 45 cycles of 95 ℃ for 15 sec, 55 ℃ for 30 sec and 72 ℃ for 45 sec. The comparative cycle threshold (Ct) method was used for gene expression data analysis[7].

Western analysis was preformed with a total of 40 μg of whole cell lysate. Antibodies that specifically recognized human CSB (Bethyl Laboratories), XPC (Sigma), and β-actin (Abcam) were used at 1:10,000, 1:50,000 and 1:10,000 dilutions, respectively. Horseradish peroxidase-conjugated secondary goat anti-rabbit antibody (Abcam) was used at a 1:10,000 dilution.

Statistical analysis

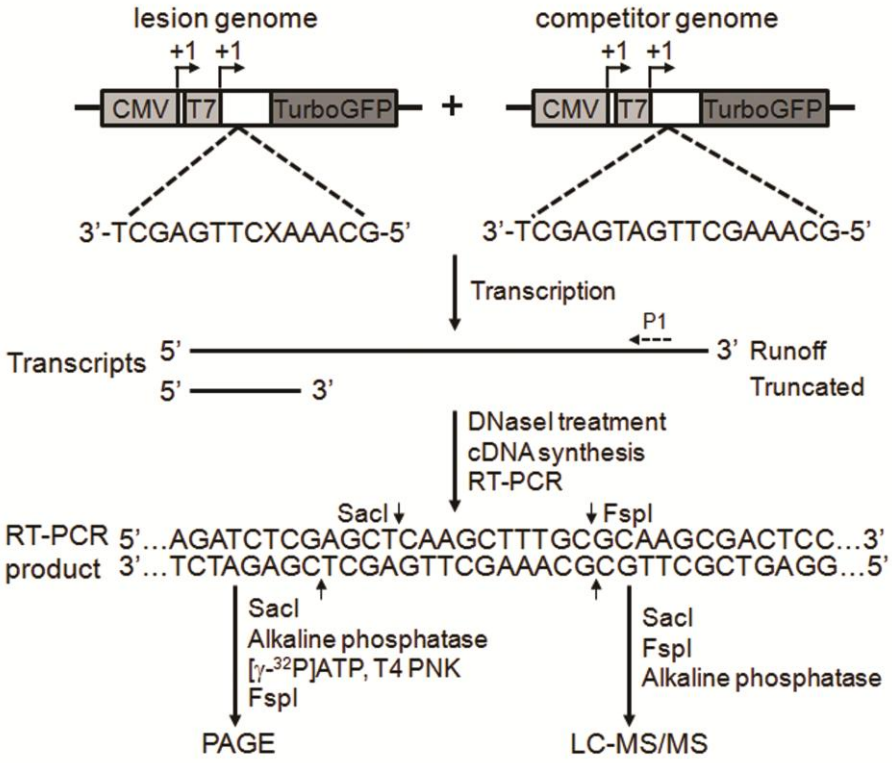
The two-tailed Student's *t*-test (unpaired) was applied for statistical analysis, and the *P* values were calculated using Microsoft Excel.

References:

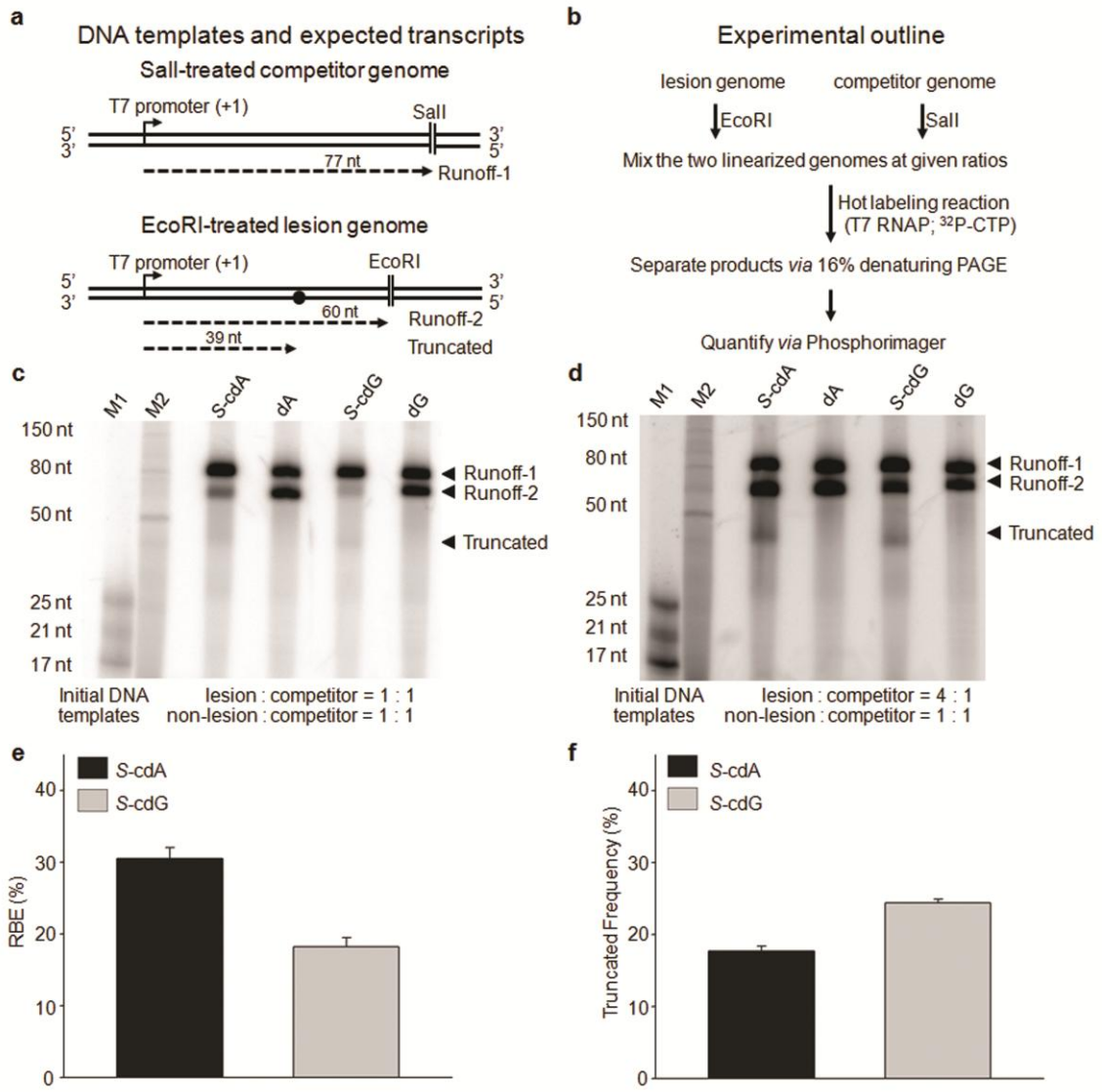
1. Rolig, R.L. et al. Survival, mutagenesis, and host cell reactivation in a Chinese hamster ovary cell ERCC1 knock-out mutant. *Mutagenesis* **12**, 277-83 (1997).
2. Baker, D.J. et al. Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian cells. *J. Biol. Chem.* **282**, 22592-604 (2007).
3. Yuan, B. et al. The roles of DNA polymerases kappa and iota in the error-free bypass of N^2-carboxyalkyl-2'-deoxyguanosine lesions in mammalian cells. *J. Biol. Chem.* **286**, 17503-11 (2011).
4. Delaney, J.C. & Essigmann, J.M. Assays for determining lesion bypass efficiency and mutagenicity of site-specific DNA lesions in vivo. *Methods Enzymol.* **408**, 1-15 (2006).
5. Yuan, B., Cao, H., Jiang, Y., Hong, H. & Wang, Y. Efficient and accurate bypass of N^2-(1-carboxyethyl)-2'-deoxyguanosine by DinB DNA polymerase *in vitro* and *in vivo*. *Proc. Natl. Acad. Sci. USA* **105**, 8679-84 (2008).

6. Hong, H., Cao, H. & Wang, Y. Formation and genotoxicity of a guanine-cytosine intrastrand cross-link lesion in vivo. *Nucleic Acids Res.* **35**, 7118-27 (2007).
7. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ Method. *Methods* **25**, 402-8 (2001).

SUPPLEMENTARY RESULTS

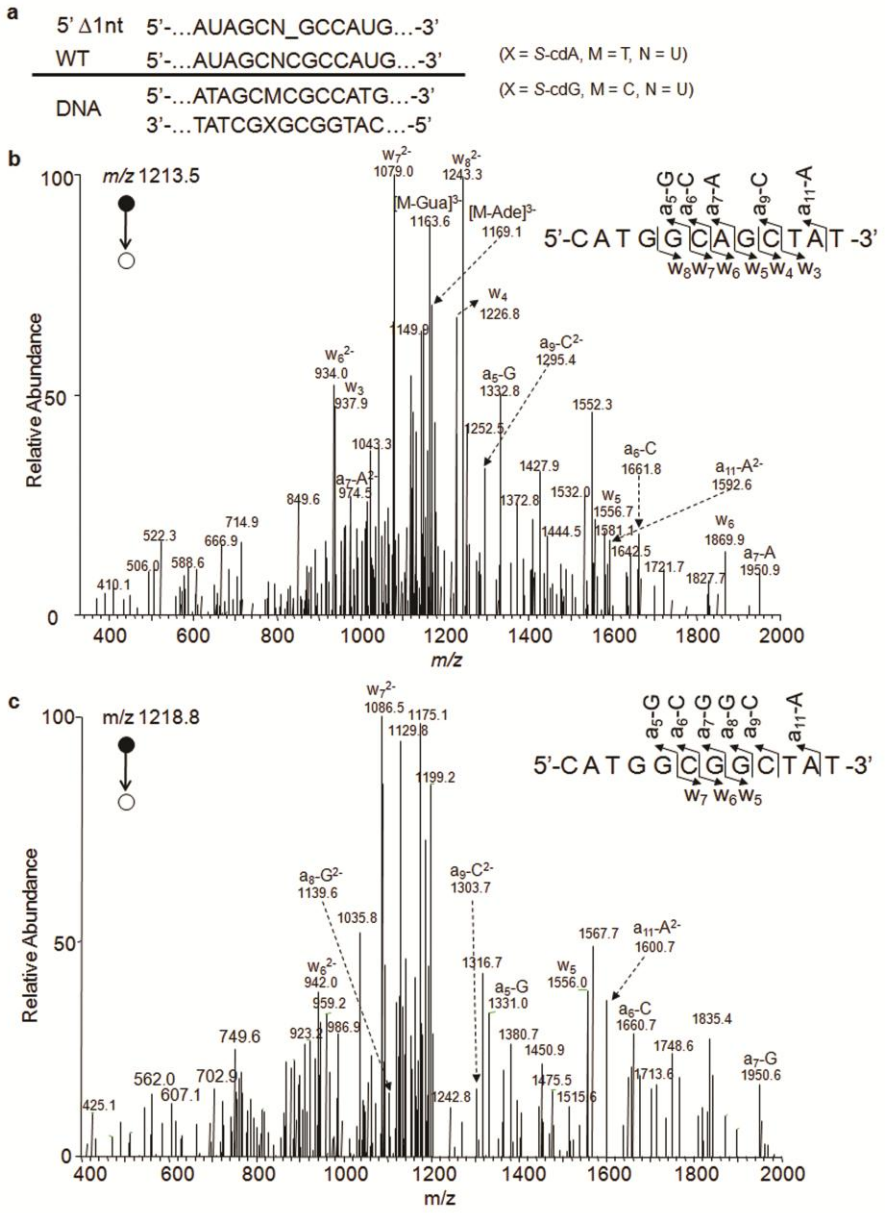


Supplementary Figure 1. A schematic diagram depicts the CTAB assay system for N^2-CEdG lesions. "X" indicates an *R*-/*S*-N^2-CEdG lesion or unmodified dG, which was located on the transcribed strand of *TurboGFP* gene downstream of the CMV and T7 promoters. The arrowheads indicate the +1 transcription start sites. Runoff RNA and truncated RNA resulting from transcription arrest at a lesion site are indicated. P1 represents a gene-specific primer used for reverse transcribing the runoff transcripts. Only the wild-type sequence of the RT-PCR products for the lesion-containing genome is shown, and the RT-PCR products arising from the competitor genome are not depicted. The RT-PCR products were digested with two restriction enzymes (SacI and FspI) and then subjected to PAGE or LC-MS/MS analyses. The cleavage sites of FspI and SacI are designated with arrows.

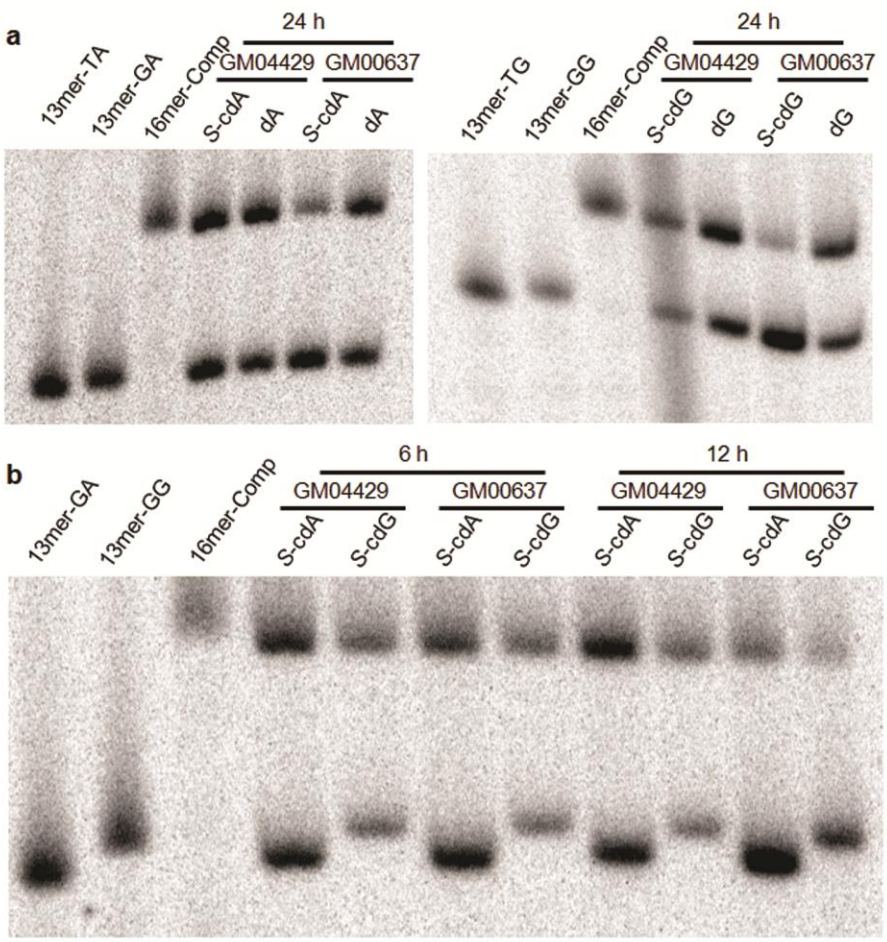


Supplementary Figure 2. RNA gel analysis for determining the effect of a single *S*-cdA or *S*-cdG on transcription by T7 RNAP. (**a**) DNA templates used for transcription reactions by including [α-^{32}P]-CTP in the reaction mixture ("hot-labeling"). (**b**) Experimental outline depicts the T7 RNAP-mediated transcription systems with radiolabeling. (**c**,**d**) RNA gel image showing the runoff and truncated transcripts generated under radiolabeling conditions. The initial control/competitor genome ratio was 1:1 for all experiments while the lesion/competitor ratios were 1:1 and 4:1 for **c** and **d**, respectively. M1 and M2 represent microRNA marker and Low Range ssRNA Ladder (New England Biolabs), respectively. (**e**) Relative bypass efficiencies (RBE) of *S*-cdA and *S*-cdG in hot-labeling transcription by T7 RNAP. The runoff RNA signal ("Runoff-2") arising from the transcription of the lesion-carrying plasmid are normalized to the corresponding runoff RNA signal ("Runoff-1") from the competitor genome taking into consideration the molar ratios of the two genomes used for transcription reactions. The RBE value was then calculated by dividing this ratio with that obtained from the control experiment employing an undamaged DNA substrate. (**f**) Quantification of the truncated transcripts arising

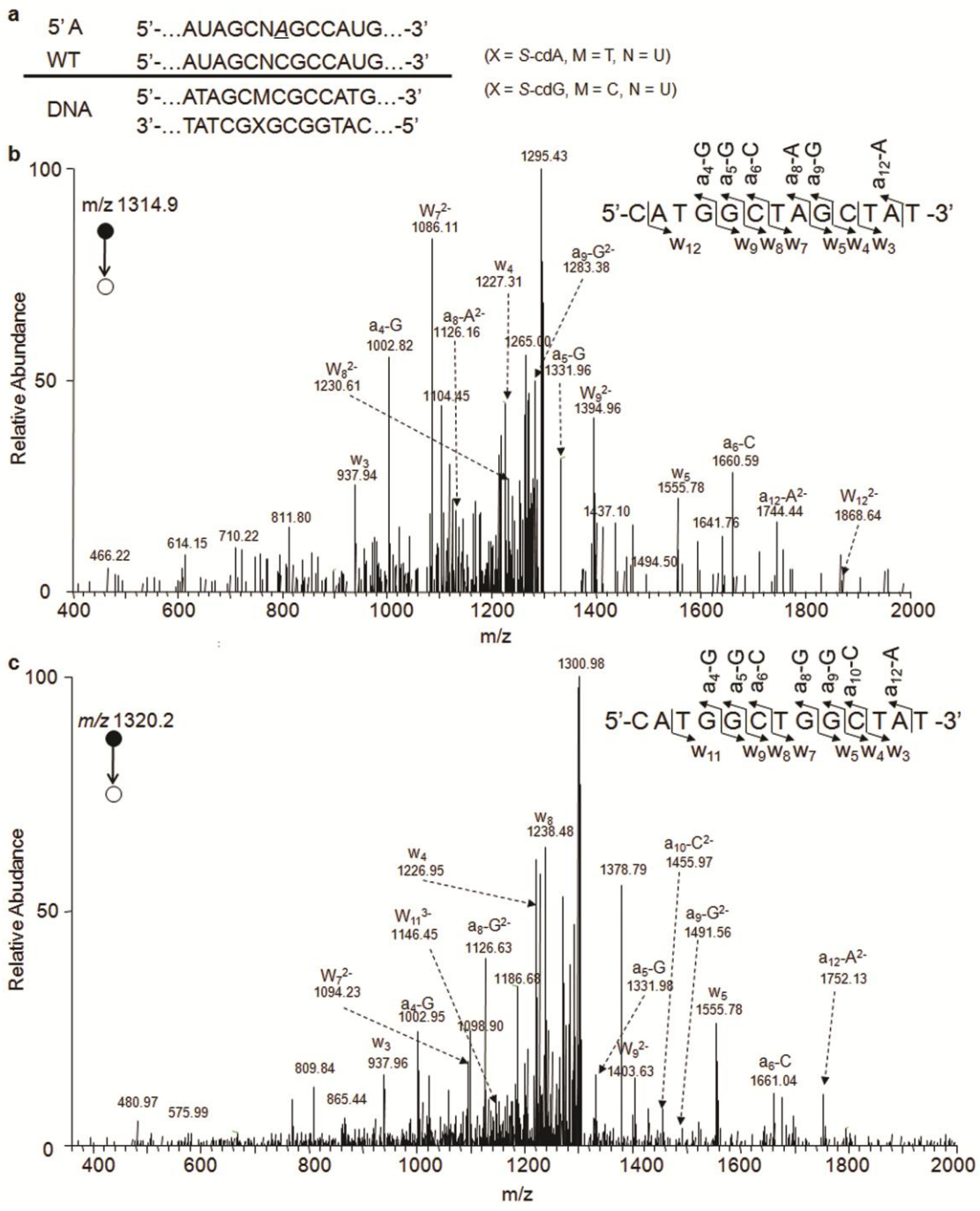
from the transcription of the lesion-carrying plasmid. The truncated frequency was calculated by dividing the truncated RNA signal by the total RNA signal arising from the transcription of the lesion-carrying plasmid taking into consideration of their relative length and cytosine content. The runoff and truncated transcripts contain 18 and 11 cytosines, respectively. The data represent the mean and standard error of results from three independent experiments.



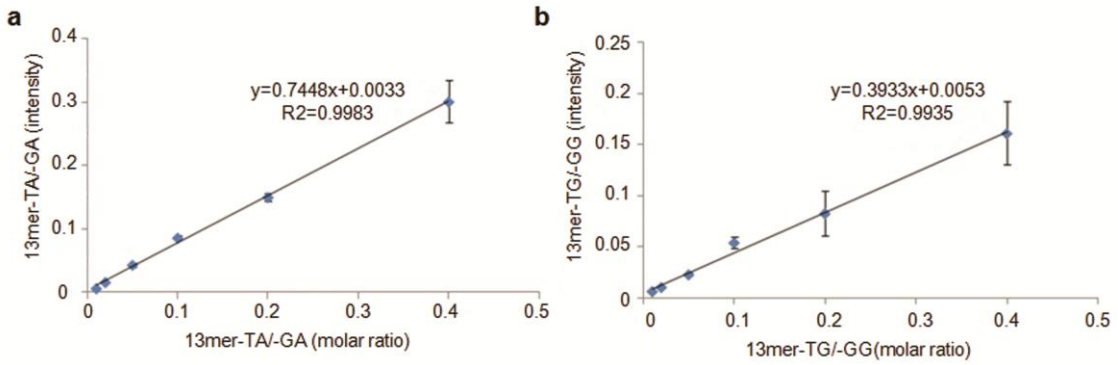
Supplementary Figure 3. Transcriptional mutagenesis induced by *S*-cdA or *S*-cdG in *in vitro* transcription systems using T7 RNAP**.** (**a**) The sequences of wild-type (WT) and mutant (5'Δ1nt) sequences were indicated above the double-stranded DNA construct. (**b**,**c**) LC-MS/MS for monitoring the restriction fragments of interest with 5'Δ1nt mutation. Shown in (**b**) and (**c**) are the MS/MS of the $[M–3H]^{3-}$ ions (*m/z* 1213.5 and 1218.8) of these two ODNs [i.e., d(CATGGCAGCTAT) (12mer-AG) and d(CATGGCGGCTAT) (12-mer GG)]. Shown above the spectrum is a scheme summarizing the observed [a_n – Base] and w_n fragment ions [nomenclature follows that described previously. *J. Am. Soc. Mass Spectrom.* 3, 60-70 (1992)] .



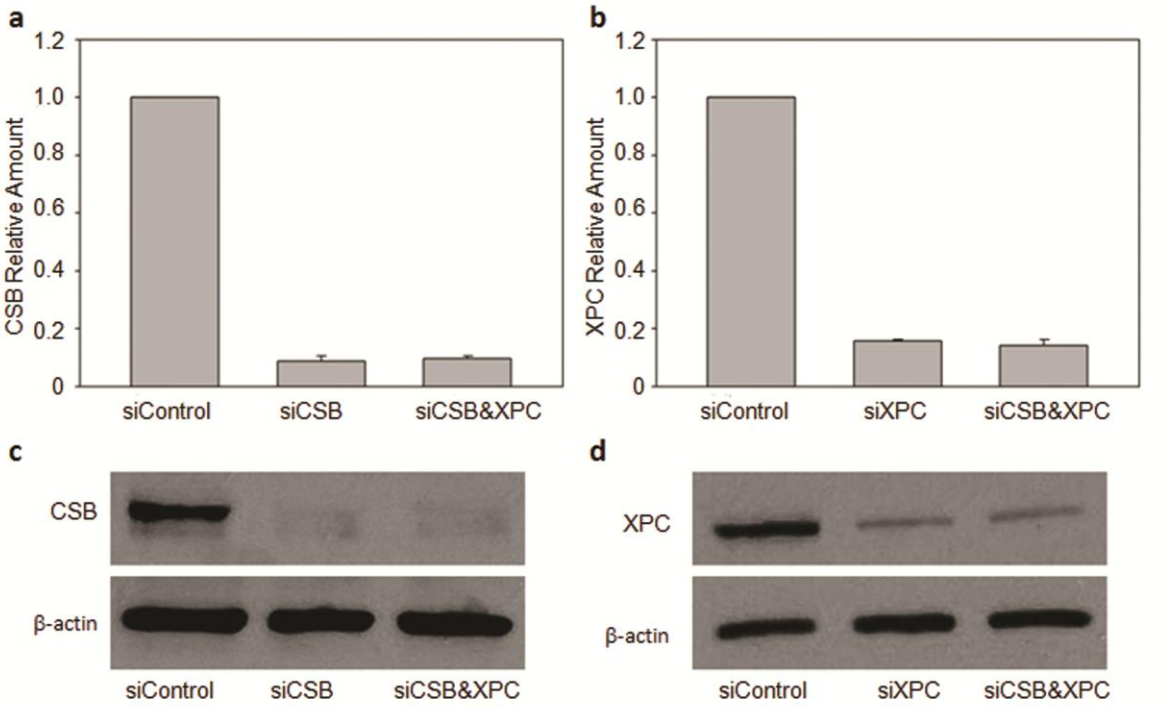
Supplementary Figure 4. PAGE analysis for assessing the RBE values of *S*-cdA and *S*-cdG in NER-proficient (GM00637) and NER-deficient XPA (GM04429) cells at 24 h (**a**), 6 and 12 h (**b**) following transfection. The control/competitor genome and lesion/competitor ratios used for these transcription templates were 1:1 and 19:1, respectively. The restriction fragment arising from the competitor genome, i.e., d(CATGGCGATATGCTAT), is designated with '16mer-Comp'; '13mer-GA', '13mer-TA', '13mer-GG', and '13mer-TG' represent standard ODNs d(CATGGCGAGCTAT), d(CATGGCTAGCTAT), d(CATGGCGGGCTAT), and d(CATGGCTGGCTAT), respectively.



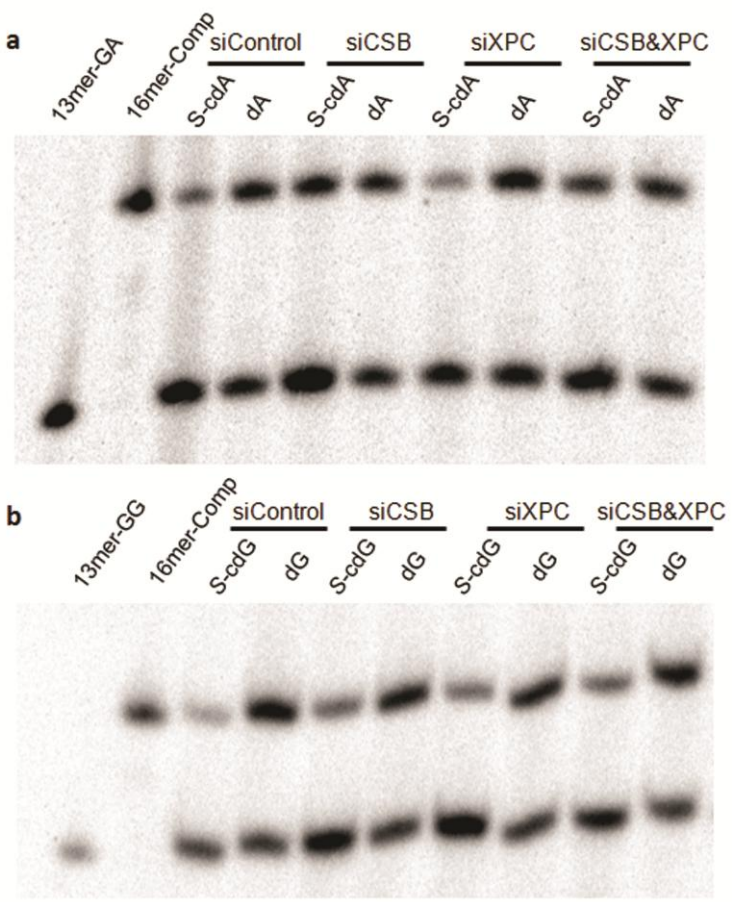
Supplementary Figure 5. Transcriptional mutagenesis induced by *S*-cdA or *S*-cdG in *in vitro* transcription systems with NER-deficient XPA (GM04429) cells. (**a**) The sequences of wild-type (WT) and mutant (5' A) sequences were indicated above the double-stranded DNA construct. (**b**,**c**) LC-MS/MS for monitoring the restriction fragments of interest with 5'A mutation. Shown in (**a**) and (**b**) are the MS/MS of the $[M–3H]^{3-}$ ions (*m/z* 1314.9 and 1320.2) of these two ODNs [i.e., d(CATGGCTAGCTAT) (13mer-TA) and d(CATGGCTGGCTAT) (13-mer TG)]. Shown above the spectrum is a scheme summarizing the observed [a_n – Base] and w_n fragment ions



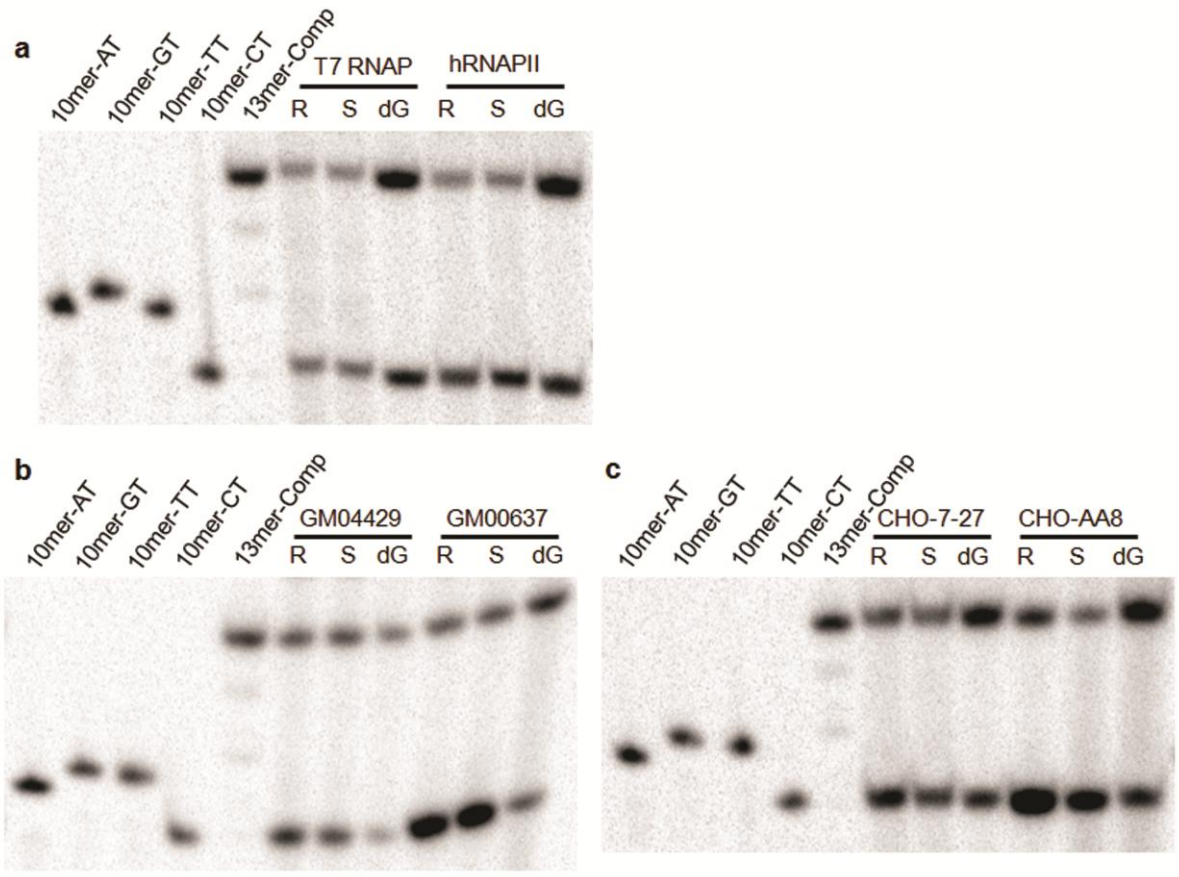
Supplementary Figure 6. Calibration curves for quantification of 13mer-TA [i.e., d(CATGGCTAGCTAT)] (**a**) and 13mer-TG [i.e., d(CATGGCTAGCTAT)] (**b**). The ratios of peak areas 13mer-TA and 13mer-TG over their corresponding wild-type sequences (13mer-GA and 13mer-GG) found in the selected-ion chromatograms (SICs) were plotted against the molar ratio for these ODNs to give calibration curves. The w_4^{2-}, $[a_4\text{-Base}]^{2-}$ and $[a_5\text{-Base}]^{2-}$ ions were selected for plotting the SICs of these 13-mer ODNs. Identical LC-MS/MS conditions were used for analyzing the CTAB samples and the standards for calibration. The data represent the mean and standard error of results from three independent measurements.



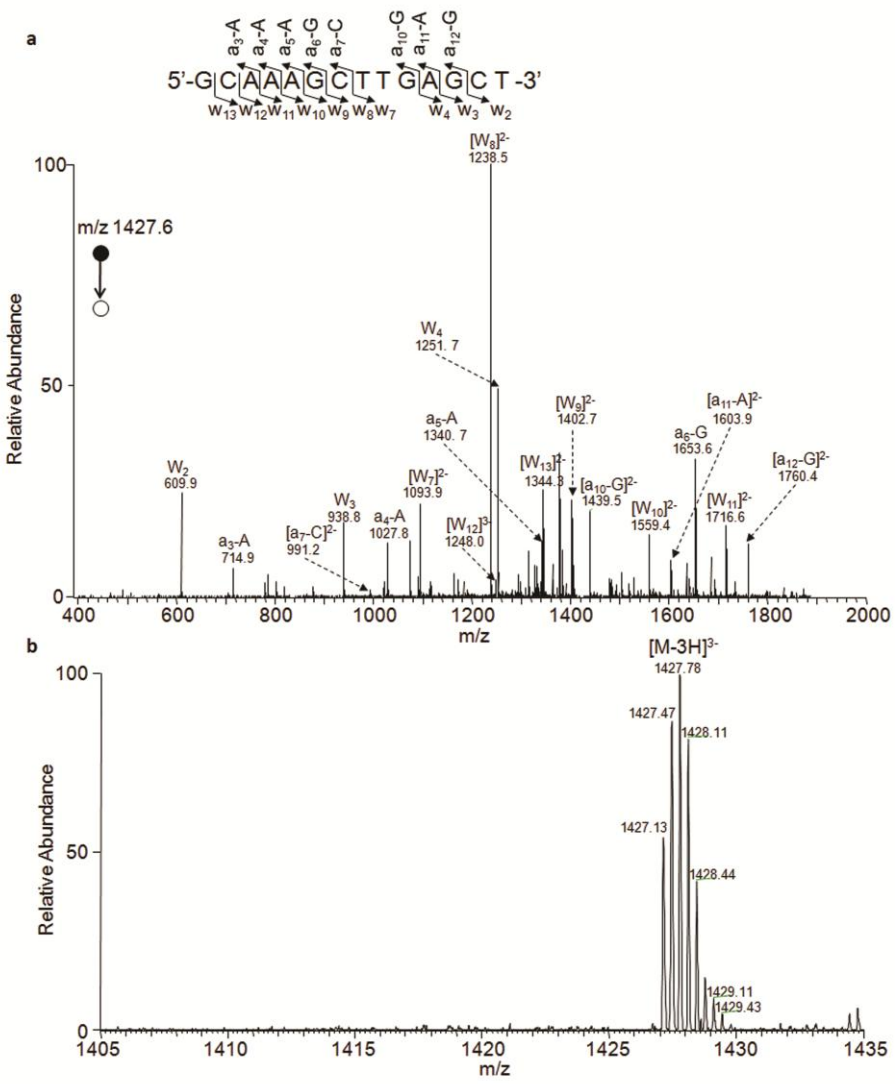
Supplementary Figure 7. Real-time RT-PCR (**a**,**b**) and Western blot (**c**,**d**) analysis for monitoring the siRNA-induced knockdown of *CSB* and/or *XPC* in 293T cells. GAPDH was used as control for real-time RT-PCR analysis, and β-actin was used as the loading control for Western analysis. The data shown in **a** and **b** represent the mean and standard error of results from three separate experiments.



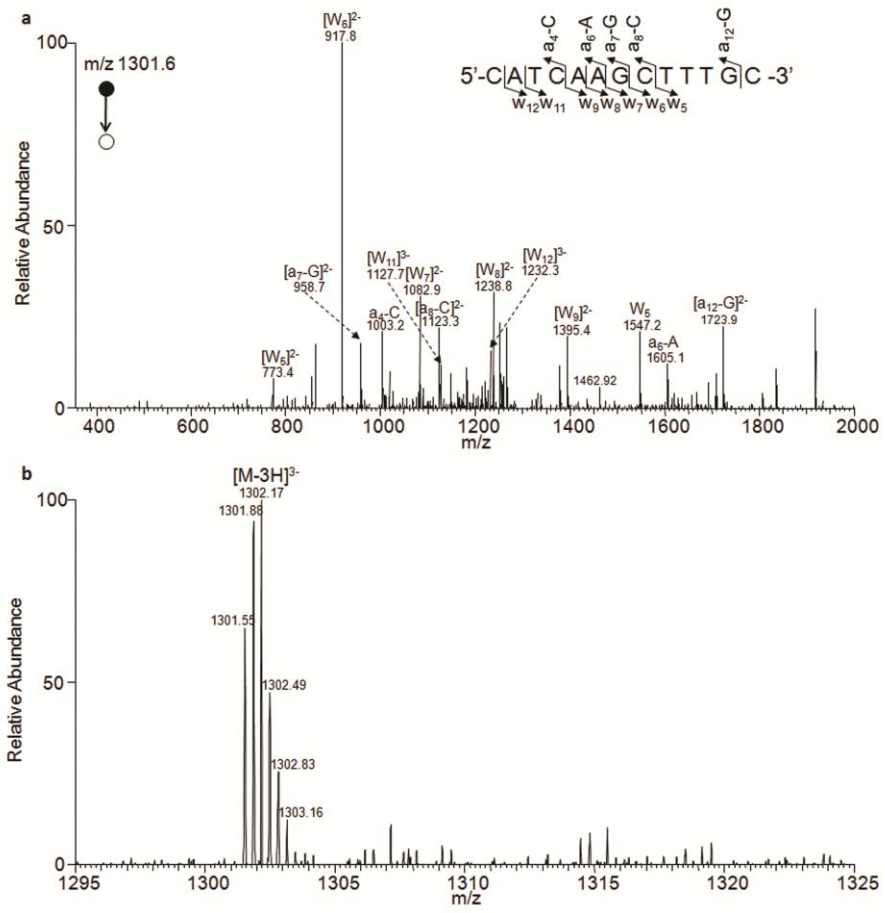
Supplementary Figure 8. PAGE analysis for assessing the RBE values of *S*-cdA (**a**) and *S*-cdG (**b**) in 293T cells treated with either or both the *CSB* and *XPC* siRNAs. The control/competitor genome and lesion/competitor ratios used for these transcription templates were 1:1 and 19:1, respectively. The restriction fragment arising from the competitor genome, i.e., d(CATGGCGATATGCTAT), is designated with '16mer-Comp'; '13mer-GA', '13mer-TA', '13mer-GG', and '13mer-TG'represent standard ODNs d(CATGGCGAGCTAT), d(CATGGCTAGCTAT), d(CATGGCGGGCTAT) and d(CATGGCTGGCTAT), respectively.



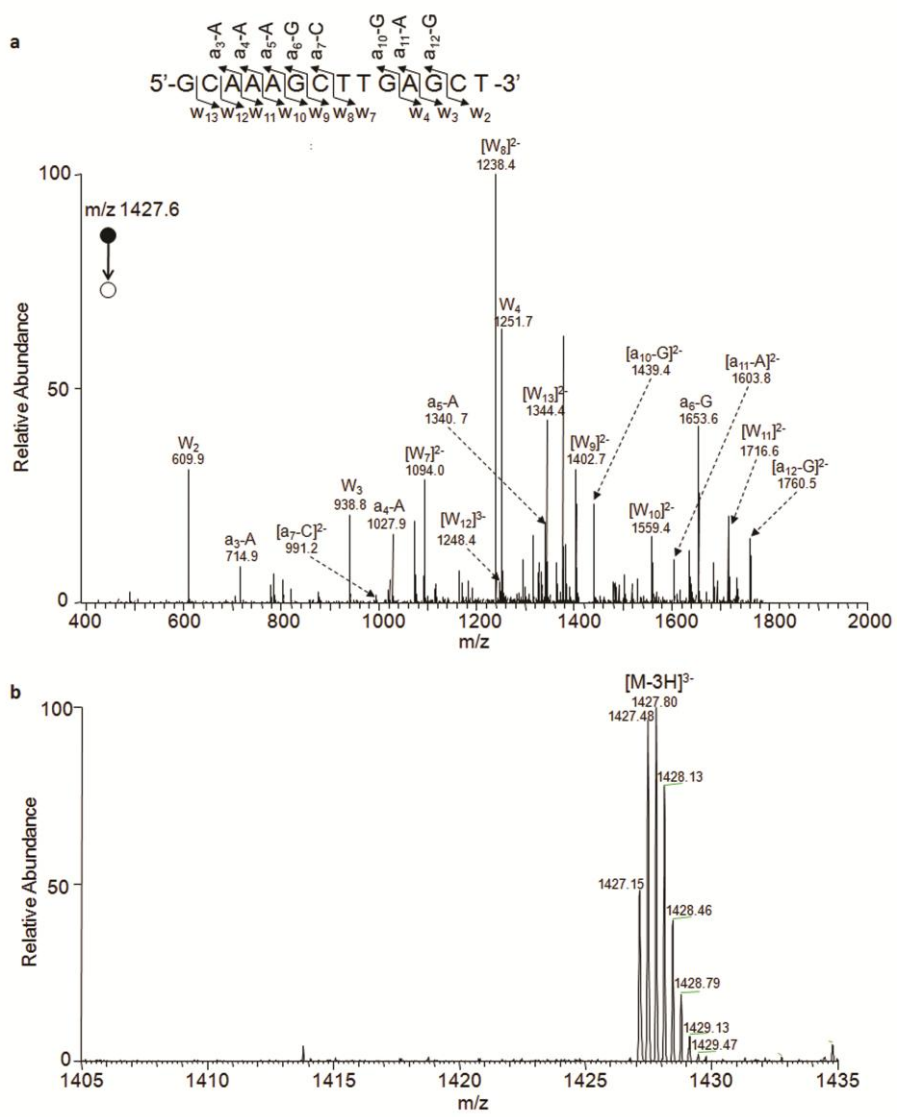
Supplementary Figure 9. PAGE analysis for the determination of the effect of *R-/S-N*^{2}-CEdG on transcription *in vitro* and in mammalian cells. Gel image showing the restriction fragments released from the RT-PCR products corresponding to the competitor genome and the control or lesion-bearing genome in *in vitro* transcription systems using T7 RNAP or Hela nuclear extract (hRNAPII) (**a**) and *in vivo* transcription systems with NER-deficient and repair-proficient cells at 24 h (**b**,**c**). These cells include XPA-deficient (GM04429) and repair-proficient (GM00637) human fibroblast cells and ERCC1-deficient (CHO-7-27) and corresponding repair-proficient (CHO-AA8) Chinese hamster ovary cells. The control/competitor genome and lesion/competitor ratios used for these transcription templates were 1:1 and 19:1, respectively. The restriction fragment arising from the competitor genome, i.e., d(CATCAAGCTTTGC), is designated with '13mer-Comp'; '10mer-AT', '10mer-GT', '10mer-TT' and '10mer-CT' represent standard ODNs d(CAAGMNTTGC), where 'MN' represent 'AT', 'GT', 'TT' and 'CT', respectively. 'R', 'S', and 'dG' represent *R-N*^{2}-CEdG, *S-N*^{2}-CEdG and the control substrate, respectively.



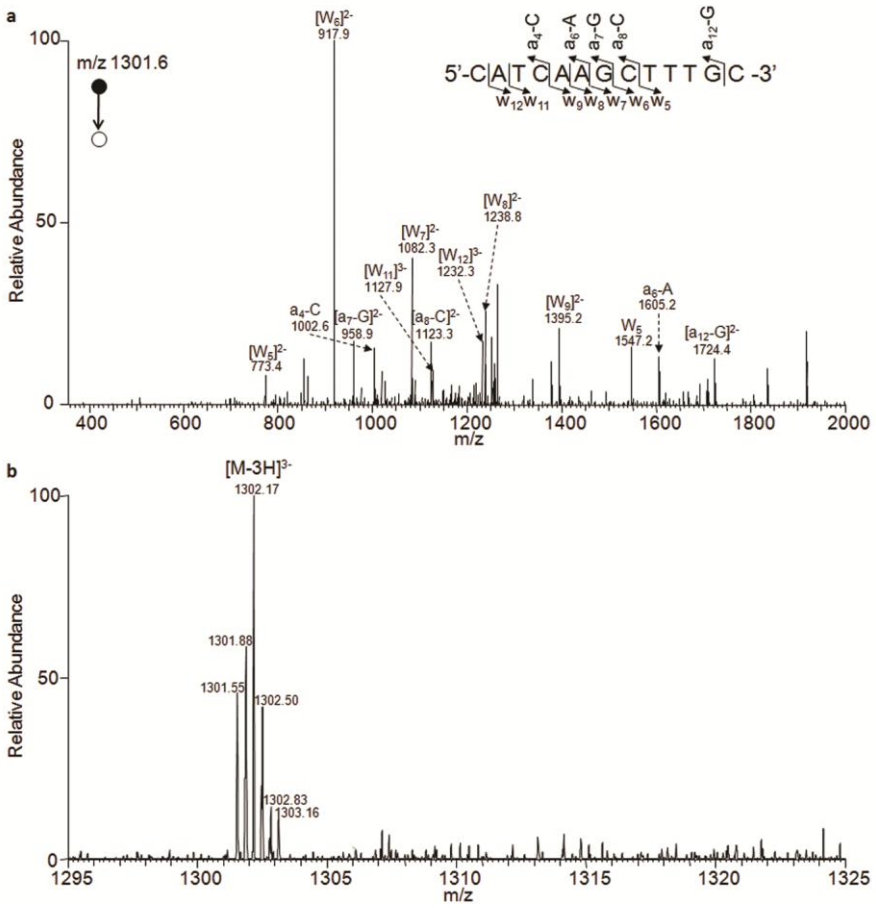
Supplementary Figure 10. LC-MS and MS/MS for monitoring the 14-mer restriction fragments of *R*-N^2-CEdG-bearing substrate from the *in-vitro* transcription with T7 RNAP. (**a**) Product-ion spectrum (MS/MS) of the $[M–3H]^{3–}$ ion (*m/z* 1427.6) of the complementary 14-mer fragment of the wild-type sequence d(GCAAAGCTTGAGCT). Shown above the spectrum is a scheme summarizing the observed [a_n – Base] and w_n fragment ions. (**b**) High-resolution "ultra zoom-scan" ESI-MS revealed the presence of the $[M–3H]^{3–}$ ion of d(GCAAAGCTTGAGCT), but the absence of the $[M–3H]^{3–}$ ions of single-base substitution products of d(GCAAAPCTTGAGCT) (P = A, C, or T, with the calculated *m/z* being 1422.2, 1414.2 and 1419.2, respectively) or the $[M–3H]^{3–}$ ion of d(GCAATGCTTGAGCT) (calculated *m/z* 1424.6), which contains a misincorporation of an adenosine (A) opposite the next nucleotide downstream of the lesion (5'A mutation).



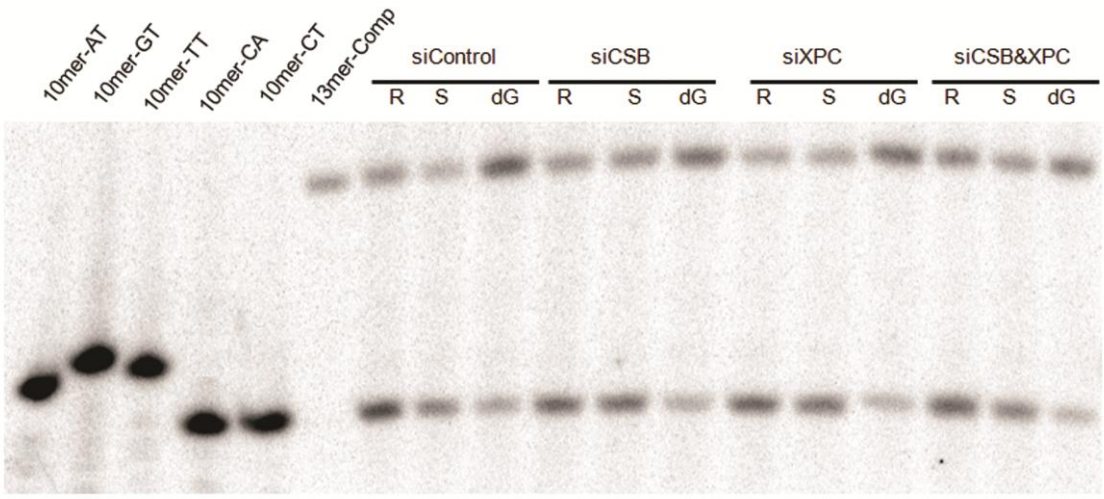
Supplementary Figure 11. LC-MS and MS/MS for monitoring the 13-mer restriction fragments of *R*-N^2-CEdG-bearing substrate from the *in-vitro* transcription with T7 RNAP. (**a**) MS/MS of the $[M-3H]^{3-}$ ion (*m/z* 1301.6) of the 13-mer fragment arising from the competitor genome, i.e., d(CATCAAGCTTTGC). Shown above the spectrum is a scheme summarizing the observed [a_n – Base] and w_n ions. (**b**) High-resolution "ultra zoom-scan" ESI-MS revealed the presence of the $[M-3H]^{3-}$ ion of the 13-mer fragment arising from the competitor genome, but the absence of the 13-mer fragment from lesion-containing genome with a single-nucleotide deletion opposite the lesion, i.e., d(GCAAACTTGAGCT) (calculated *m/z* 1317.9 for the $[M-3H]^{3-}$ ion).



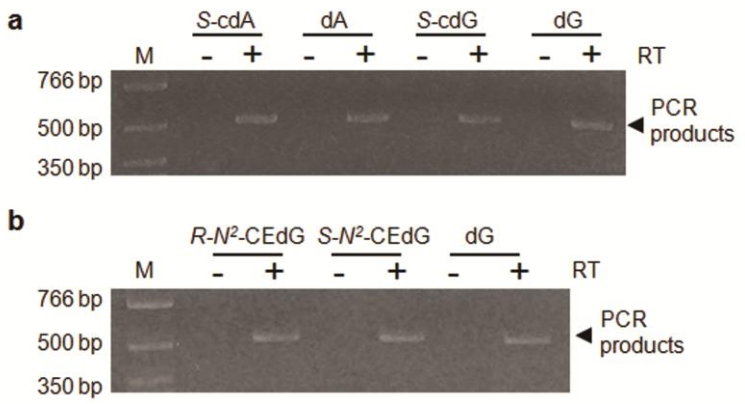
Supplementary Figure 12. LC-MS and MS/MS for monitoring the 14-mer restriction fragments resulting from the *in vivo* transcription of *R*-N^2-CEdG-bearing substrate in GM04429 cells. (**a**) MS/MS of the $[M\text{-}3H]^{3-}$ ion (*m/z* 1427.6) of the complementary 14-mer fragment of the wild-type sequence d(GCAAAGCTTGAGCT). Shown above the spectrum is a scheme summarizing the observed [a_n – Base] and w_n fragment ions. (**b**) High-resolution "ultra zoom-scan" ESI-MS revealed the presence of the $[M–3H]^{3–}$ ion of d(GCAAAGCTTGAGCT), but the absence of the $[M–3H]^{3–}$ ion of single-base substitution products of d(GCAAAPCTTGAGCT) (P = A, C, or T, with the calculated *m/z* being 1422.2, 1414.2 and 1419.2, respectively), or the $[M–3H]^{3–}$ ion of d(GCAATGCTTGAGCT) (calculated *m/z* 1424.6), which contains a misincorporation of an adenosine (A) opposite the next nucleotide downstream of the lesion (5'A mutation).



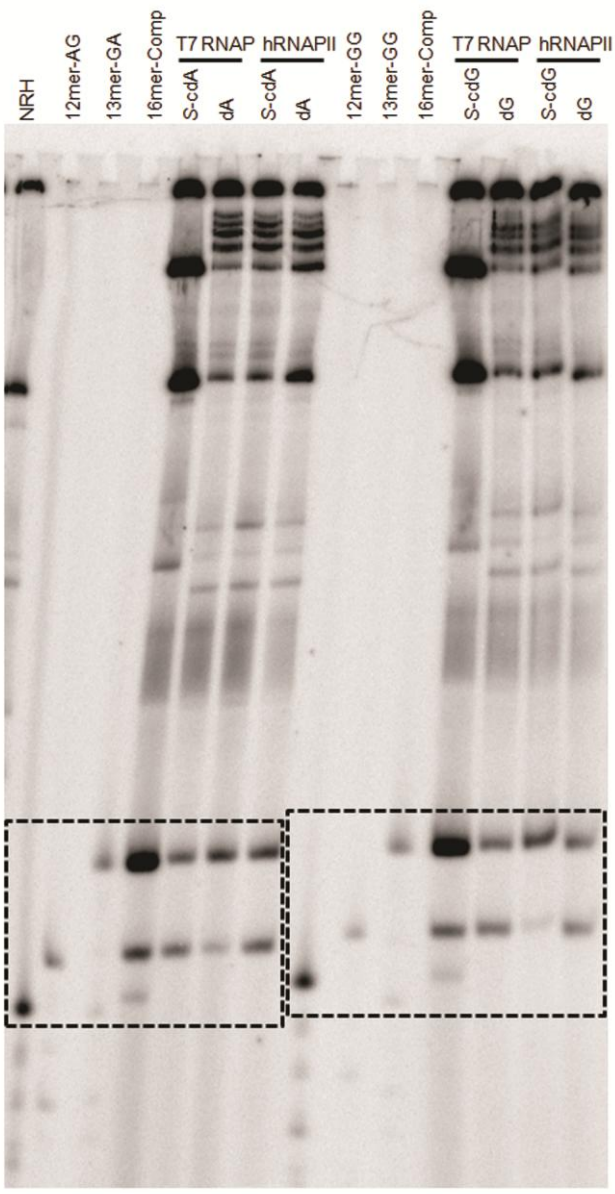
Supplementary Figure 13. LC-MS and MS/MS for monitoring the 13-mer restriction fragments from the *in vivo* transcription of *R*-N^2-CEdG-bearing substrate in GM04429 cells. (**a**) MS/MS of the $[M-3H]^{3-}$ ion (*m/z* 1301.6) of the 13-mer fragment arising from the competitor genome, i.e., d(CATCAAGCTTTGC). Shown above the MS/MS is a scheme summarizing the observed [a_n – Base] and w_n fragment ions. (**b**) High-resolution "ultra zoom-scan" ESI-MS revealed the presence of the $[M-3H]^{3-}$ ion of the 13-mer fragment arising from the competitor genome, but the absence of the 13-mer fragment from lesion-containing genome with a single-nucleotide deletion opposite the lesion, i.e., d(GCAAACTTGAGCT) (calculated *m/z* 1317.9 for the $[M-3H]^{3-}$ ion).



Supplementary Figure 14. PAGE analysis for assessing the RBE values of *R-/S-*N^2-CEdG in 293T cells treated with either or both of the *CSB* and XPC siRNAs. The control/competitor genome and lesion/competitor ratios used for these transcription templates were 1:1 and 19:1, respectively. '13mer-Comp' represents the restriction fragment arising from the competitor genome, i.e., d(CATCAAGCTTTGC). '10mer-CT' represents the wild-type sequence arising from the lesion-containing or control genome, i.e., d(CAAGCTTTGC). 'R', 'S', and 'dG' represent *R-*N^2-CEdG, *S-*N^2-CEdG and the control substrate, respectively.



Supplementary Figure 15. Representative agarose gel images (**a**,**b**) for analyzing the RT-PCR products of DNaseI-treated RNA arising from *in vitro* transcription systems using T7 RNAP. The RNA was treated with (+) or without (-) reverse transcription (RT) and then used as the template for PCR reaction. M represents Low Molecular Weight DNA Ladder (New England Biolabs).



Supplementary Figure 16. Full gel corresponding to Figure 3a. NRH, not reported here.

Primer	Sequence (5'→3')
pTGFP-T7-F	AGTGAACCGTCTAATACGACTCACTATAGGGAGATCCGCTAGCGC
pTGFP-T7-R	GCGCTAGCGGATCTCCCTATAGTGAGTCGTATTAGACGGTTCACT
pTGFP-comp-1F	CTAGCGCTACCGGACTCAGATCTCGAGCTCATCAAGCTTTGCGCAAGCGACTCCG
pTGFP-comp-1R	AATTCGGAGTCGCTTGCGCAAAGCTTGATGAGCTCGAGATCTGAGTCCGGTAGCG
pTGFP-GA-F	CAGATCTCGAGCTCCATGGCGAGCTATAATTGCGCAAGCGACTCC
pTGFP-GA-R	GGAGTCGCTTGCGCAATTATAGCTCGCCATGGAGCTCGAGATCTG
pTGFP-GG-F	CAGATCTCGAGCTCCATGGCGGGCTATAATTGCGCAAGCGACTCC
pTGFP-GG-R	GGAGTCGCTTGCGCAATTATAGCCCGCCATGGAGCTCGAGATCTG
pTGFP-comp-2F	CAGATCTCGAGCTCCATGGCGACATGCTATAATTGCGCAAGCGACTCC
pTGFP-comp-2R	GGAGTCGCTTGCGCAATTATAGCTTGTCGCCATGGAGCTCGAGATCTG

Supplementary Table 1. Primers for site-directed mutagenesis in producing pTGFP vectors for *in-vitro* and *in-vivo* transcription assays.

Gene	Forward primer (5'→3')	Reverse primer (5'→3')
CSB	CTTGACCCGGCTGGTATTGT	TGGCTGGAAGTTTCTGAGGAG
XPC	GTGCAGCATCCTTCATCAAC	GCATCTTCTGCTCTTCCAC
GAPDH	TTTGTCAAGCTCATTTCCTGGTATG	TCTCTTCCTCTTGTGCTCTTGCTG

Supplementary Table 2. Primers for quantitative RT-PCR analysis.