HHS Public Access

Author manuscript

J Am Chem Soc. Author manuscript; available in PMC 2023 January 12.

Published in final edited form as:

J Am Chem Soc. 2022 January 12; 144(1): 454–462. doi:10.1021/jacs.1c10490.

DNA–Protein Cross-Linking Sequencing for Genome-Wide Mapping of Thymidine Glycol

Feng Tang,
Department of Chemistry, University of California, Riverside, Riverside, California 92521-0403, United States

Sauvage Center for Molecular Sciences, Department of Chemistry, Wuhan University, Wuhan 430072, China

Jun Yuan,
Department of Chemistry, University of California, Riverside, Riverside, California 92521-0403, United States

Bi-Feng Yuan,
Sauvage Center for Molecular Sciences, Department of Chemistry, Wuhan University, Wuhan 430072, China

Yinsheng Wang
Department of Chemistry, University of California, Riverside, Riverside, California 92521-0403, United States

Abstract

Thymidine glycol (Tg) is the most prevalent form of oxidatively induced pyrimidine lesions in DNA. Tg can arise from direct oxidation of thymidine in DNA. In addition, 5-methyl-2′-deoxycytidine (5-mdC) can be oxidized to 5-mdC glycol, and its subsequent deamination also yields Tg. However, Tg's distribution in the human genome remains unknown. Here, we presented a DNA–protein cross-linking sequencing (DPC-Seq) method for genome-wide mapping of Tg in human cells. Our approach capitalizes on the specificity of a bifunctional DNA glycosylase, i.e., NTHL1, for the covalent labeling, as well as DPC pulldown, SDS-PAGE fractionation, and membrane transfer for highly efficient and selective enrichment of Tg-bearing DNA. By employing DPC-Seq, we detected thousands of Tg sites in the human genome, where dual ablation of NTHL1 and NEIL1, the major DNA glycosylases responsible for Tg repair, led to pronounced increases in the number of Tg peaks. In addition, Tg is depleted in genomic regions associated with active transcription but enriched at nucleosome-binding sites, especially at heterochromatin

Corresponding Authors: Yinsheng Wang – Department of Chemistry, University of California, Riverside, Riverside, California 92521-0403, United States; Yinsheng.Wang@ucr.edu; **Bi-Feng Yuan** – Sauvage Center for Molecular Sciences, Department of Chemistry, Wuhan University, Wuhan 430072, China; bfyuan@whu.edu.cn.

Supporting Information

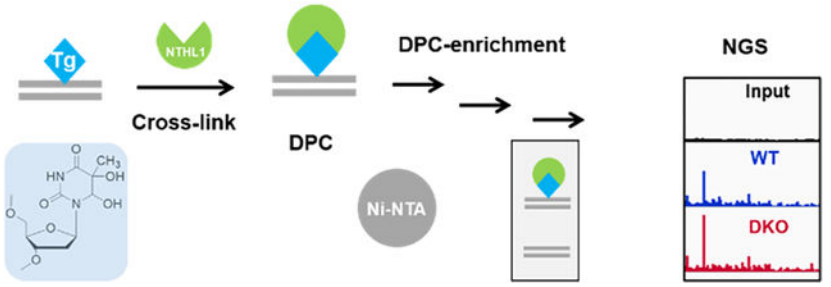
The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c10490.
Model DNA sequences, LC-MS/MS characterizations of the Tg-ODN, purification of recombinant NTHL1, NTHL1 labeling and enrichment optimization, CRISPR KO verification and NGS data analysis, including Tables S1 and Figures S1–S21 (PDF)

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c10490

The authors declare no competing financial interest.

sites marked with H3K9me2. Collectively, we developed a DPC-Seq method for highly efficient enrichment of Tg-containing DNA and for genome-wide mapping of Tg in human cells. Our work offers a robust tool for future functional studies of Tg in DNA, and we envision that the method can also be adapted for mapping other modified nucleosides in genomic DNA in the future.

Graphical Abstract



INTRODUCTION

Endogenous metabolism and exposure to environmental agents and ionizing radiation can result in augmented formation of reactive oxygen species (ROS) in human cells.[1] ROS can attack genomic DNA to yield a plethora of DNA lesions, such as 8-oxo-7,8-dihydro-2′-deoxyguanosine (8-oxodG)[2] and 5,6-dihydroxy-5,6-dihydrothymidine (thymidine glycol, Tg).[3] Tg, the most prevalent form of ROS-induced pyrimidine lesions in DNA, can arise from direct oxidation of thymidine in DNA[3] or from oxidation of 5-methyl-2′-deoxycytidine (5-mdC) to 5-mdC glycol and its subsequent deamination.[4] Tg may be a more reliable indicator of oxidative stress than 8-oxodG, partly because 8-oxodG is more readily generated than Tg during sample preparation.[5] Moreover, Tg is a predominant type of DNA damage induced by ionizing radiation, rendering it directly relevant to cancer radiation therapy.[6]

The knowledge about the distributions of DNA lesions in the human genome is important for understanding mutations observed in cancer genomes.[7,8] Such knowledge will also provide new information for exploring how site-specific induction and repair of DNA lesions are modulated by genomic context,[9] e.g., nucleosome and transcription factor binding, epigenetic state of chromatin, transcription status, DNA repair capacity, etc. In this respect, a growing number of studies have been conducted to locate DNA lesions, such as 8-oxodG, DNA photoproducts, cisplatin-induced intrastrand cross-links, and the benzo[*a*]pyrene diol epoxide (BPDE) adduct of dG in genomic DNA.[10–16] However, no method has been reported for mapping Tg at the genome-wide scale.

By taking advantage of their high substrate specificities, Burrows and co-workers[17,18] employed DNA glycosylases to selectively remove 8-oxodG and its further oxidation product, spiroiminodihydantoin-2′-deoxyribonucleoside, and marked the damage sites by sealing the resultant single-nucleotide gaps with a DNA ligase or by incorporating a non-natural nucleotide, which enabled the identification of initial damage sites with Sanger sequencing. This strategy of combining glycosylase cleavage with non-natural nucleotide

incorporation was later employed for developing sequencing methods for genome-wide mapping of 8-oxodG and 2′-deoxyuridine.[14,19] We reasoned that DNA glycosylases might also be harnessed for mapping Tg in the human genome. In this vein, Tg in human cells can be repaired primarily by two DNA glycosylases, endonuclease III-like protein 1 (NTHL1) and Nei-like DNA glycosylase 1 (NEIL1).[20–23]

Here, we present a DNA–protein cross-linking sequencing (DPC-Seq) method to map Tg in the human genome (Figure 1A). The method capitalizes on NTHL1-mediated removal of thymine glycol, covalent binding of the ensuing abasic site with the Lys residue at its active site, reduction to stabilize the resulting Schiff base (Figure 1B), and a multistep procedure for the highly selective and efficient enrichment of DNA-NTHL1 cross-link for subsequent library construction and next-generation sequencing analysis. By conducting DPC-Seq with two different cell lines, we identified thousands of Tg peaks, which provided the first insight about the genomic distribution of Tg in the context of DNA repair competence, nucleosome occupancy, histone epigenetic modifications, and transcription status.

RESULTS

NTHL1-Mediated Covalent Labeling of Tg-Containing DNA.

Previous studies revealed that, after removing its substrate nucleobases, bifunctional DNA glycosylases are able to cross-link with the ensuing abasic site in DNA, and the resulting Schiff base intermediate can be reduced by $NaBH_4$ to yield stable DPCs.[24] This borohydride trapping strategy has been used to isolate and characterize many bifunctional DNA glycosylases.[25–27] NTHL1 is one of the bifunctional glycosylases that can selectively excise thymine glycol from genomic DNA. Therefore, we set out to explore the feasibility of using this glycosylase to label and enrich Tg-containing DNA.

We first prepared an oligodeoxyribonucleotide (ODN) harboring a site-specifically inserted Tg (Figure S1) and purified recombinant NTHL1 with a C-terminal 6xHis tag (Figure S2). After the Tg-ODN was incubated with NTHL1 in the presence of $NaBH_4$, we observed a stable product with a substantially diminished mobility on SDS-PAGE (Figure 2A). Following proteinase K treatment, the product migrates similarly to double-stranded DNA, confirming the DPC formation. No DPC was detectable for the corresponding reactions with the use of damage-free DNA or in the absence of $NaBH_4$, manifesting the high specificity of the labeling reaction (Figure 2A). We also optimized a number of experimental conditions, including types and amounts of reducing reagents, solution pH, and amounts of NTHL1 (Figures S3–S6). We found that a better labeling efficiency can be attained with the use of $NaBH_3CN$ than $NaBH_4$ (Figure S3); hence, $NaBH_3CN$ was used for all the subsequent experiments. Under our optimized conditions, we were able to achieve an 80% labeling efficiency of the Tg-containing duplex DNA with NTHL1 (Figure S6).

DPC-Seq Enables Enrichment of Tg-Containing DNA with High Efficiency and Selectivity.

With the availability of this highly efficient and specific labeling method, we examined the enrichment efficiency of the Tg-containing DNA in the presence of an excess amount of genomic DNA. After covalent labeling of the Tg-containing DNA with NTHL1-His_6,

the resulting DPCs were first enriched from the reaction mixture using Ni-NTA beads. Under our optimized enrichment conditions, a model Tg-containing DNA can be enriched by several hundred-fold compared to the corresponding damage-free DNA (Figures 2B, S7, and S8), which is similar to what was previously reported for dU- or 8-oxodG-containing DNA.[13,19] Considering the relatively low frequency of Tg in genomic DNA (at a level of 0.8–10 per 10^6 nucleosides in different experimental systems),[28,29] we decided to further enrich the DPC by virtue of its chemical stability. In this vein, we found that the reduced NTHL1-abasic site DNA cross-link is very stable even after denaturing the protein with boiling.

We asked if we could employ SDS-PAGE to further remove the residual DPC-free DNA in the eluent obtained from affinity purification with Ni-NTA beads. We found that, at a similar molecular weight, free DNA migrates much faster than protein in SDS-PAGE gel (Figure S9). For instance, a 500 bp dsDNA (~300 kDa) displays a faster mobility than a 50 kDa marker protein. However, after being cross-linked with a protein, the ensuing DPC displays a migration property resembling that of a protein on SDS-PAGE, indicating that SDS-PAGE can efficiently separate DPC from free DNA. In addition, SDS-PAGE separation can disrupt nonspecific interactions between DNA and beads. Inspired by the previously reported protocols for CLIP-Seq, which harnesses membrane transfer to enrich RNA-protein cross-link,[30] we employed membrane transfer to further remove free DNA and to achieve a better recovery of NTHL1 DPC. Our results showed that almost all free DNA was removed after the membrane transfer, even in the experiment conducted in the presence of an excessive amount of genomic DNA (Figure S10). By using these optimized enrichment conditions, we found that, after sequential purification with Ni-NTA beads, SDS-PAGE fractionation, and membrane transfer, a model Tg-containing DNA can be enriched by 2780-fold (Figure 2B), which is approximately 10-fold higher than what was reported previously in other DNA damage sequencing methods. Therefore, our DPC-Seq method integrates specific Tg labeling (with NTHL1), affinity pulldown of DPC using Ni-NTA beads, SDS-PAGE separation, and membrane transfer to achieve highly efficient enrichment of Tg-containing DNA.

DPC-Seq Reveals the Genome-Wide Distribution of Tg in Cultured Human Cells.

We next utilized DPC-Seq to map Tg in the human genome. Apart from Tg, NTHL1 can also remove other DNA lesions from genomic DNA, e.g., 2,6-diamino-4-hydroxy-5-formamidopyrimidine (FapyG), 5,6-di-hydro-2′-deoxyuridine (5,6-DHdU), and 5-hydroxy-2′-deoxyuridine (5-OHdU), and label abasic sites induced endogenously and/or formed from BER processes. Hence, we pretreated genomic DNA with Fpg, SMUG1, and EndoIV, which allowed for removal of these lesions and free abasic sites without affecting appreciably the labeling of Tg (Figures 2C and S11A). In this vein, we found that NTHL1 is also able to label an abasic site-containing duplex DNA, albeit at a lower efficiency than that for the corresponding Tg-containing duplex DNA, which might be due to poorer binding of the abasic site- than Tg-containing DNA with NTHL1. After pretreatment with Fpg, SMUG1, and EndoIV, the abasic site-containing DNA, however, can no longer be subjected to NTHL1 labeling (Figure S11B), suggesting the effectiveness of the pretreatment on removing other lesions.

To assess the feasibility of DPC-Seq, we spiked 30 fmol of Tg-containing DNA into 1 μg of fragmented genomic DNA isolated from HEK293T cells, which represents a spike-in level of approximately 1 Tg modification per 10^5 nucleosides. Our results obtained from DPC enrichment (Figure S12) and TA-clone sequencing showed that, among the ten clones that we sequenced, six had insertions of the initial Tg-containing DNA, two had insertions that can be aligned to the human genome, one could be aligned with blocking DNA (i.e., salmon sperm DNA), and one was attributed to self-ligation. These results suggest that Tg-containing DNA can be efficiently enriched with our method and correctly sequenced (Figure S13). In this vein, it is worth noting that we did not detect any mutation at the initial Tg site, which might be attributed to the preferential amplification of the strand complementary to the DPC-containing strand, where DPC is known to block polymerase bypass.[31] Indeed our primary extension assay results revealed that the DNA-peptide cross-link remnant arising from proteinase K cleavage of the NTHL1-abasic site cross-link completely stalls primer extension mediated by the Q5 DNA polymerase used in our library construction (Figure S14).

In mammalian cells, NTHL1 and NEIL1 are the main BER glycosylases for Tg repair.[21–23] To further understand the distribution of Tg damage, we used CRISPR-Cas9 to knock out *NTHL1* and *NEIL1* genes to examine the Tg landscape in genomic DNA of human cells without its major BER glycosylases. Western blot and Sanger sequencing analyses confirmed the successful knockout of these two genes (Figures S15 and S16). Next, we applied the DPC-Seq method to map Tg in fragmented genomic DNA samples extracted from parental HEK293T (WT) cells and the isogenic NTHL1/NEIL1 double-knockout (DKO) cells. After peak calling with the control and overlap between two biological replicates (Figure S17), we identified 3688 and 9373 high-confidence Tg peaks for parental HEK293T and the isogenic DKO cells, respectively (Figure 3A). Approximately 76% (2812 out of 3688) of peaks detected for parental HEK293T cells overlapped with those detected in DKO cells (Figure 3B), suggesting the facile formation and/or poor repair of Tg in some genomic regions. This result also underscored that, even with active BER, Tg at some sites persists long enough in the genome to be captured. In addition, our results showed that there is a nearly 2.5-fold increase in peak numbers observed in the DKO cells compared to WT cells. With a 4- or 5-fold cutoff in enrichment, there are still markedly higher numbers of peaks in the DKO than WT background (Figure 3A). Moreover, the common peaks detected for the two cell lines exhibit higher signal intensity in DKO cells than parental cells (Figures 3C and 3D). These results are in accordance with the fact that NTHL1 and NEIL1 constitute the major DNA glycosylases for Tg repair in human cells.[32]

Correlation of Tg Sites with Genomic Features.

We next analyzed the correlation of Tg peaks with genomic features (Figure S18A). For parental HEK293T cells, 60.2%, 37.9%, and 1.7% of the Tg peaks are located in intergenic regions, introns, and other regions [e.g., 3′-untranslated region (3′-UTR), transcriptional termination sites (TTS), and transcriptional start sites (TSS)], respectively. In addition, Tg exhibits a slight enrichment in intron regions and depletion in other regions (Figure S18B). The distributions of Tg peaks are similar in the DKO cells, where most Tg sites are located within intergenic and intron regions. Tg sites' signal intensity meta profiles

are clearly depleted at the TSS and TTS regions (Figure S18C), which are consistent with the previous observations made for DNA photoproducts, benzo[*a*]pyrene-dG adducts, and cisplatin adducts.[12,16,33] These results suggest a strong influence of transcription on the genome-wide distribution of Tg. In this vein, transcription-coupled repair is an NER pathway that removes DNA lesions from the genome.[34] While transcription-coupled BER is still under debate, some studies suggest such a possibility.[34–37] Although Tg is mainly repaired by NTHL1- and NEIL1-mediated BER pathways in mammalian cells, some studies show that oxidatively induced single-nucleobase lesions may be repaired by the NER pathway.[38,39] Thus, depletion of Tg at TSS and TTS may arise from more active BER or NER at these sites. Interestingly, we also observed an apparent depletion of the Tg signal at TSS and TTS in the DKO cells (Figure S18C), indicating that, in BER-deficient cells, the TC-NER pathway may participate in the removal of Tg in actively transcribed regions. Nevertheless, we could not formally exclude the possibility that DNA glycosylase(s) other than NTHL1 and NEIL1 may be involved in Tg repair, e.g., NEIL3.[40] Future experiments are needed to examine this possibility.

We also investigated the influence of nucleosome binding and epigenetic state of chromatin on Tg distribution. We found that Tg peaks are enriched at histones H3 and H4 binding sites (Figure 4A) but depleted at DNase I hypersensitive (DHS) sites (Figure S19), which is consistent with the notion that the packaging of DNA into nucleosomes impedes DNA repair.[41,42]

We next explored Tg distribution around the center of four histone epigenetic marks, i.e., H3K9me2, H3K4me1, H3K4me3, and H3K27ac (Figure 4A). Interestingly, a higher enrichment of Tg peaks was observed at H3K9me2, a mark of heterochromatin that is strongly associated with transcriptional repression.[43] An Integrative Genomics Viewer (IGV) plot also shows a high degree of colocalization between Tg and H3K9me2 peaks (Figure 4B). In addition, motif analysis revealed an enrichment of highly repetitive sequences (Figure S20), which are characteristic of heterochromatin.[44] However, we failed to observe similar enrichment patterns for histone marks that are characteristic of active transcription, i.e., H3K4me1, H3K4me3, and H3K27ac (Figure 4B). Additionally, when we overlapped Tg peaks with those of the histone epigenetic marks, a much higher overlap rate was observed in the heterochromatin region than those with active histone marks (Figure 4C). These data are in agreement with the notions that DNA lesions in active and open chromatin regions are more readily repaired than those in repressed and heterochromatin regions[45] and that transcription-coupled NER/BER may be involved in Tg removal.[46] Moreover, mitochondrial DNA, which is devoid of nucleosome binding or NER, displays a higher level of Tg compared to the nuclear DNA (Figure S21), which is in keeping with previous findings for other DNA lesions.[47,48] While Tg is known to be induced from the oxidation of 5-mdC to 5-mdC glycol and its subsequent deamination, we were not able to detect an enrichment of Tg in CpG islands (Figure S18B). This is likely attributed to the lack of adequate resolution of our mapping method. In this vein, the resolution of our method is approximately 100–300 bp, which is imposed by the size of DNA fragments employed for NTHL1 labeling and sequencing library construction.

DISCUSSION

In this study, we developed a genome-wide mapping method for Tg detection in human cells. Our approach harnesses the high specificity of NTHL1, a natural DNA repair enzyme for Tg. The $NaBH_3CN$-mediated reductive cross-linking enables efficient removal of non-cross-linked free DNA through rigorous washing. In addition, SDS-PAGE separation and membrane transfer allow for further removal of free DNA, which provides an extremely high enrichment efficiency (i.e., by ~2780-fold) of a Tg-containing DNA over the corresponding unmodified DNA. By employing this highly efficient method, we generated the first genome-wide distribution maps of Tg in HEK293T and the isogenic NTHL1/NEIL1-DKO cells. A high degree of overlap (76%) was observed between the peaks in the two cell lines, suggesting the susceptible induction and/or poor repair of Tg in some genomic regions. Even with active BER, some Tg sites persist long enough in the genome to be captured. An additional 6000 peaks were identified in the DKO cells, suggesting the highly efficient repair of Tg in repair-competent human cells.

Our results revealed a depletion of Tg peaks at TSS and TTS in both parental and DKO HEK293T cells. This observation underscores that Tg may be repaired by transcription-coupled repair pathway(s) that is(are) independent of NEIL1 and NTHL1, which is in agreement with the observation that Tg can be recognized by the human NER factors XPA and XPC.[38] Future research is needed to verify this. In addition, our results suggest that Tg persists within nucleosomes, which is in keeping with the *in vitro* results showing that the packaging of DNA into nucleosomes hinders the repair of oxidatively induced DNA lesions.[49,50] Moreover, Tg peaks are highly enriched in heterochromatin relative to euchromatin regions, which is attributed to differential accessibility of DNA repair enzymes to these chromatin regions.

In summary, we developed a novel DPC-based strategy for highly efficient and selective enrichment of Tg-containing DNA over unmodified DNA. The method, in conjunction with second-generation sequencing analysis, facilitated the first genome-wide mapping of Tg in human cells. Our results revealed an enrichment of Tg in nucleosome-binding regions, especially in heterochromatin regions, and unveiled the modulations of the genome-wide occupancy of this modified nucleoside by DNA repair competence and transcription activity. It can be envisaged that the strategy can be employed to assess the genome-wide distributions of Tg in tissue samples and other organisms, and the method can also be adapted for assessing, in the future, the occurrence of other DNA lesions and modifications in the human genome. In addition, this method is amenable for monitoring the dynamics of Tg repair in the future.

EXPERIMENTAL SECTION

Cell Lines.

HEK293T cells were purchased from ATCC. NTHL1/NEIL1 DKO cells were generated in the HEK293T background using CRISPR-Cas9, following previously reported procedures.[51] The guide RNAs were designed according to a previously described method,[52] and their sequences are listed in Table S1. The guide sequences for NEIL1 and NTHL1 sites were

inserted into the pX330 plasmid (Addgene). At 24 h following plasmid transfection, the cells were diluted and seeded into a 96-well plate at a concentration of approximately 0.5 cell per well. After 2 weeks, single colonies were transferred to a 24-well plate. Western blot was performed to screen for the knockout clones by using an anti-NEIL1 antibody (Proteintech). Sanger sequencing was performed to confirm the mutation of the sequence. The NEIL1-KO cell line was used for the further knockout of the *NTHL1* gene by following previously described procedures.[53] The deletion of NTHL1 was screened and confirmed by PCR and Sanger sequencing.

The cell lines used in this study were confirmed to be free of mycoplasma contamination with the use of the e-Myco Mycoplasma PCR Detection Kit (Bulldog Bio, Inc.). The cells were maintained in DMEM (Life Technologies) supplemented with 10% FBS (Invitrogen) and 1% penicillin/streptomycin at 37 °C in a humidified incubator with 5% CO_2. The cells were harvested at 80% confluence, and genomic DNA was extracted by using a Monarch Genomic DNA Purification Kit (NEB, T3010S). Butylated hydroxytoluene (0.1 mM) (Sigma) was added during DNA extraction to minimize artificial generation of Tg during sample preparation.[14]

Preparation of Model Tg-Containing DNA Sequences.

We synthesized an 18-mer ODN harboring a site-specifically inserted Tg by employing automated solid-phase DNA synthesis on a Beckman Oligo 1000 M DNA synthesizer (Fullerton, CA, USA) with a commercially available Tg phosphoramidite (Glen Research) following the vendor's recommended protocols. Synthesized ODNs were cleaved from controlled pore glass and deprotected by treating with 30% ammonium hydroxide at room temperature for 2 h. After solvent removal using a SpeedVac, the solid residues were dissolved in 0.5 mL of TEA·3HF and kept at 40 °C overnight to remove the TBDMS protecting groups. The ODN was subsequently purified by denaturing PAGE and characterized by ESI-MS and MS/MS analyses (Figure S1). The ESI-MS and MS/MS results support the site-specific insertion of Tg. The 18-mer Tg-containing ODN was ligated with a 30-mer ODN in the presence of a 24-mer scaffold ODN to generate a 48-mer Tg-containing ODN. The ligation product was purified by denaturing PAGE and used as model Tg-ODN for the following labeling and enrichment experiments. Detailed sequences are listed in Table S1.

Generation of Recombinant NTHL1 Protein.

The coding sequence of the human *NTHL1* gene was amplified by PCR and inserted into the pET30a, followed by a 6xHis tag at the C terminus. For purification of NTHL1, Rosetta (DE3) *E. coli* cells were transformed with the plasmid and grown on LB agar with 100 μg/mL ampicillin at 37 °C overnight. A single colony was amplified in a 5 mL LB medium containing 100 μg/mL ampicillin at 37 °C overnight and then diluted to a 500 mL LB medium. After the optical density (OD600) reached approximately 0.6, the culture was cooled to 16 °C, and the protein expression was induced with 1 mM IPTG (Sigma) at 16 °C for approximately 20 h. The cells were then collected by centrifugation and lysed at 4 °C by sonication in a 15 mL 1× PBS buffer (Fisher BioReagents) containing 1 mM phenylmethylsulfonyl fluoride. The suspension was centrifuged at 14,000 g for 15 min, and

the supernatant was collected and incubated with 0.5 mL of HisPur Ni-NTA Resin (Thermo Scientific, 88221) at 4 °C for 30 min. The resin was then washed 6 times with a 1 mL washing buffer (PBS with 25 mM imidazole, pH 7.4). The His_6-tagged NTHL1 was eluted with a 1 mL PBS buffer containing 250 mM imidazole (pH 7.4) 4 times. The purity of NTHL1 was assessed by SDS-PAGE analysis. The pure glycosylase-containing fractions were pooled and dialyzed against PBS with 1 mM DTT and concentrated using Slide-A-Lyzer Dialysis Cassettes, 10 kDa MWCO (Thermo Scientific, 66380). Protein concentration was determined by using a Bradford Protein Assay Kit (Bio-Rad). Aliquots of purified proteins were stored at −80 °C until use.

Labeling of Tg-ODN with NTHL1.

For the labeling experiments, 250 fmol of 5′−^{32}P-labeled T/Tg-ODN was annealed with its complementary strand. The resulting double-stranded DNA was incubated with 2 pmol of NTHL1 in 10 μL of 50 mM PBS buffer (pH 7.0) containing 1 mM EDTA, 1 mM DTT, and 50 mM $NaBH_4$ at 37 °C for 30 min, followed by digestion with 8 U proteinase K (NEB, P8107S) at 37 °C for 20 min. The reaction was quenched by a 4× Laemmli Sample Buffer (Biorad, #1610747) and heated at 90 °C for 5 min. The samples were separated using 10% SDS-PAGE at 90 V for 60 min and imaged on a Typhoon 9410 imager. For pH optimization, 250 fmol of T/Tg-ODN was incubated with 2 pmol of NTHL1 in 10 μL of 50 mM PBS buffer (pH 5–8) containing 1 mM EDTA, 1 mM DTT, and 50 mM $NaBH_3CN$ or $NaBH_4$ at 37 °C for 30 min. The results showed that $NaBH_3CN$ gave better labeling efficiency (Figure S3); hence, we subsequently optimized the concentrations of $NaBH_3CN$ and NTHL1 in the labeling reactions.

Other oxidatively induced DNA lesions that may interfere with NTHL1 labeling were removed by incubating the isolated DNA with an enzyme cocktail containing 1 μL of endonuclease IV (NEB, M0304), 1 μL of Fpg (NEB, M0240S), 1 μL of hSMUG1 (NEB, M0336S) in a buffer containing 25 mM Tris (pH 7.5), 50 mM NaCl, 5 mM $MgCl_2$, and 0.5 mM DTT at 37 °C for 60 min and then at 85 °C for 20 min to deactivate the enzymes. The mixture was then labeled with 2 pmol of NTHL1 in a buffer containing 50 mM Tris-HCl (pH 7.0), 1 mM EDTA, 1 mM DTT, and 10 mM $NaBH_3CN$ at 37 °C for 60 min.

Enrichment of Tg-Induced NTHL1 DPC with Bead Purification.

To optimize the washing times in bead purification, we first spiked 250 fmol of ^{32}P-labeled T/Tg-ODN into 10 μg of fragmented genomic DNA isolated from HEK293T cells. To the mixture was then added 8 pmol of NTHL1 in a 10 μL buffer containing 50 mM Tris (pH 7.0), 1 mM EDTA, 1 mM DTT, and 10 mM $NaBH_3CN$, and the mixture was incubated at 37 °C for 60 min. DPC was enriched from the mixture by incubating with 15 μL of HisPur Ni-NTA resin, which was washed twice with PBS prior to use, in 800 μL of PBS at 4 °C for 2 h. The sample was centrifuged, and the supernatant was discarded. After washing with 1 mL of PBS 5 times, DNA was eluted from the beads by incubating with a 15 μL PBS buffer containing 500 mM imidazole in a Thermomixer at 37 °C for 15 min with interval mixing at 1,200 rpm. The eluents were resolved by 10% SDS-PAGE at 90 V for 60 min and imaged by using a Typhoon 9410 imager.

To minimize nonspecific adsorption of DNA to Ni-NTA beads, we optimized the amount of blocking DNA. Toward this end, 3 fmol of T/Tg-ODN was spiked into 10 μg of genomic DNA isolated from HEK293T cells and labeled with 8 pmol of NTHL1 following the aforementioned procedures. Different amounts of salmon sperm DNA (0–50 μg) were subsequently added to the binding buffer to block the beads. After elution, the eluents were diluted 2 times and subjected to qPCR quantification by using Luna Universal qPCR Master Mix (NEB, M3003X) on the CFX96 RT-qPCR detection system (Bio-Rad). Primers used for qPCR are listed in Supplementary Table S1. The "enrichment fold", quantified by the $\Delta\Delta C_t$ method,[54] represents the fold change of Tg-ODN/T-ODN after enrichment compared to the input sample. First, the difference in C_t values (ΔC_t) between input and enriched samples was calculated, which was used to calculate the recovery rate. The difference in ΔC_t ($\Delta\Delta C_t$) between Tg-ODN and T-ODN was then calculated and transformed into the enrichment fold.

Enrichment of Tg-ODN with Bead Purification, SDS-PAGE, and Membrane Transfer.

For the two-step enrichment procedure, 3 fmol of T/Tg-ODNs was spiked into 10 μg of HEK293T genomic DNA and labeled with 8 pmol of NTHL1, followed by purification using Ni-NTA beads. The eluents were separated by 10% SDS-PAGE at 90 V for 60 min and transferred to the nitrocellulose blotting membrane (Biorad, 162–0112). The membrane was rinsed with PBST and then placed on a clean glass surface. The DNA–protein cross-link band in the molecular weight range of 50–250 kDa was cut from the membrane. The membrane was then cut into 1–2 mm slices, transferred to a new 1.5 mL tube, and washed with 500 μL of Tris-HCl buffer (30 mM, pH 8.0). To the mixture were added proteinase K (NEB, P8107S) (20 μL) and 130 μL of Tris-HCl buffer (30 mM, pH 8.0), and the mixture was incubated in a Thermomixer at 55 °C overnight with interval mixing at 1,200 rpm. The Thermomixer temperature was subsequently raised to 70 °C, and the incubation was continued at 70 °C for 20 min with interval mixing at 1,200 rpm. The liquid (150 μL) was transferred into a new DNA LoBind tube and purified with QIAquick Nucleotide Removal Kit (Qiagen, 28304) by following the manufacturer's guidelines. After purification, the eluents were employed for qPCR quantification, with the primers being listed in Table S1.

DPC-Seq.

Genomic DNA was digested to 100–300 bp with NEBNext dsDNA Fragmentase (NEB, M0348) following the manufacturer's guidelines. Sonication was avoided during the fragmentation process since it may introduce artificial generation of oxidatively induced DNA lesions. Purification and size selection were performed using NEBNext Sample Purification Beads (NEB).

Genomic DNA (10 μg) was pretreated with SMUG1, Fpg, and Endo IV in a buffer containing 25 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5 mM $MgCl_2$, and 0.5 mM DTT at 37 °C for 60 min. DNA was purified with NEBNext Sample Purification Beads. The genomic DNA was labeled with 0.8 μM NTHL1 in 40 μL of Tris-HCl buffer (50 mM, pH 7.0) with 1 mM EDTA, 1 mM DTT, and 10 mM $NaBH_3CN$ at 37 °C for 60 min. DPC was enriched from the reaction mixture by incubating with 20 μL of HisPur Ni-NTA Resin in 800 μL of PBS containing 50 μg of single-stranded salmon sperm DNA (Sigma, D7656). The mixture was incubated at 4 °C overnight. After centrifugation at 6000*g*, the supernatant

was discarded, and the beads were washed twice with 1 mL of PBS and then twice with 1 mL of washing buffer (20 mM Tris-HCl, pH 7.5, 0.2% Tween). The NEB adaptor was ligated (NEB, E7103S) to double-stranded DNA on Ni-NTA resin. The ligation assay was performed by following the manufacturer's guidelines with minor modifications. To the mixture were subsequently added 50 μL of ddH_2O, 3 μL of End Prep Enzyme Mix, and 7 μL of End Prep Reaction Buffer (NEB, E7103S). The end repair mixture was incubated in a Thermomixer at 20 °C for 1 h and then at 65 °C for 30 min with interval mixing at 1,200 rpm. Ligation Master Mix (30 μL), 2.5 μL of NEBNext Adaptor, and 1 μL of Ligation Enhancer (NEB, E7103S) were then added to the mixture. The mixture was incubated at 20 °C for 1 h, and the tube was flipped in every 10 min for better mixing. Three microliters of USER Enzyme (NEB, E7500S) was added to the ligation mixture, and the mixture was incubated at 37 °C for 30 min. Input samples' ligation was performed in parallel by following the manufacturer's guidelines and purified with NEBNext Sample Purification Beads.

After completing the adaptor ligation, the beads were washed twice with PBS and then twice with a washing buffer. After washing, DNA was eluted from the beads using 20 μL of PBS with 500 mM imidazole at 37 °C for 30 min with interval mixing at 1,200 rpm. The subsequent enrichment procedures (SDS-PAGE and membrane transfer) were the same as described above. Purified DNA was quantified and verified on an Agilent 2100 Bioanalyzer. The library was then constructed using NEBNext Ultra DNA Library Prep Kit (NEB, E7103S) following the manufacturer's instructions. Subsequently, the purified DNA libraries were assessed using an Agilent 2100 Bioanalyzer and multiplexed for sequencing on a HiSeq4000 system (Illumina).

DPC-Seq Data Processing and Analysis.

The sequencing reads of DPC-seq were checked with FastQC and aligned to the human hg38 reference genome using Bowtie2 with the default configuration.[55] For genome-wide Tg site identification, peak calling was performed using the model-based analysis of ChIP-seq (MACS2)[56] with the following configuration: macs2 callpeak -t treat.bam -c input.bam -f BAM -g 2.7e+9 -n --ourdir macs2. The Integrated Genome Browser[57] was used to visualize the mapping results. Overlap of two biological replicates was defined as high-confidence peaks, which were employed in further analysis. Genomic annotations were performed using PAVIS.[58] The intersection between bed files was performed using BEDTools.[59] Plotprofile and heatmap were performed using deepTools.[60] The H3, H4, H3K9me2, H3K4me1, H3K4me3, H3K27ac, TSS, and DHS sequencing data of HEK293T cell line were downloaded from GEO databases (GSM1482823; GSM1482824; GSM1644653; GSM1249888; GSM1249885; GSM1249889) and ENCODE (ENCSR294YNI; ENCF-F285OXK).[61] Motif analysis was conducted using MEME-ChIP.[62] The DPC-Seq data generated in this study have been deposited into the NCBI Gene Expression Omnibus (GEO) under accession number GSE184204.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (R01 CA210072 to Y.W.)

REFERENCES

(1). Finkel T; Holbrook NJ Oxidants, oxidative stress and the biology of ageing. Nature 2000, 408, 239–47. [PubMed: 11089981]

(2). Cadet J; Loft S; Olinski R; Evans MD; Bialkowski K; Richard Wagner J; Dedon PC; Møller P; Greenberg MM; Cooke MS Biologically relevant oxidants and terminology, classification and nomenclature of oxidatively generated damage to nucleobases and 2-deoxyribose in nucleic acids. Free Radic Res 2012, 46, 367–81. [PubMed: 22263561]

(3). Yu Y; Cui Y; Niedernhofer LJ; Wang Y Occurrence, biological consequences, and human health relevance of oxidative stress-induced DNA damage. Chem. Res. Toxicol 2016, 29, 2008–2039. [PubMed: 27989142]

(4). Cao H; Jiang Y; Wang Y Kinetics of deamination and Cu(II)/H_2O_2/ascorbate-induced formation of 5-methylcytosine glycol at CpG sites in duplex DNA. Nucleic Acids Res 2009, 37, 6635–43. [PubMed: 19706732]

(5). Collins AR; Cadet J; Möller L; Poulsen HE; Vina J Are we sure we know how to measure 8-oxo-7,8-dihydroguanine in DNA from human cells? Arch. Biochem. Biophys 2004, 423, 57–65. [PubMed: 14989265]

(6). Breimer LH; Lindahl T Thymine lesions produced by ionizing radiation in double-stranded DNA. Biochemistry 1985, 24, 4018–22. [PubMed: 3902079]

(7). Sabarinathan R; Mularoni L; Deu-Pons J; Gonzalez-Perez A; López-Bigas N Nucleotide excision repair is impaired by binding of transcription factors to DNA. Nature 2016, 532, 264–267. [PubMed: 27075101]

(8). Perera D; Poulos RC; Shah A; Beck D; Pimanda JE; Wong JW Differential DNA repair underlies mutation hotspots at active promoters in cancer genomes. Nature 2016, 532, 259–263. [PubMed: 27075100]

(9). Amente S; Scala G; Majello B; Azmoun S; Tempest HG; Premi S; Cooke MS Genome-wide mapping of genomic DNA damage: methods and implications. Cell. Mol. Life Sci 2021, 78, 6745–6762. [PubMed: 34463773]

(10). Ding Y; Fleming AM; Burrows CJ Sequencing the mouse genome for the oxidatively modified base 8-oxo-7,8-dihydroguanine by OG-Seq. J. Am. Chem. Soc 2017, 139, 2569–2572. [PubMed: 28150947]

(11). Tang F; Liu S; Li QY; Yuan J; Li L; Wang Y; Yuan BF; Feng YQ Location analysis of 8-oxo-7,8-dihydroguanine in DNA by polymerase-mediated differential coding. Chem. Sci 2019, 10, 4272–4281. [PubMed: 31015952]

(12). Li W; Hu J; Adebali O; Adar S; Yang Y; Chiou YY; Sancar A Human genome-wide repair map of DNA damage caused by the cigarette smoke carcinogen benzo[*a*]pyrene. Proc. Natl. Acad. Sci. U.S.A 2017, 114, 6752–6757. [PubMed: 28607059]

(13). Fang YX; Zou P Genome-wide mapping of oxidative DNA damage via engineering of 8-oxoguanine DNA glycosylase. Biochemistry 2020, 59, 85–89. [PubMed: 31618020]

(14). Wu J; McKeague M; Sturla SJ Nucleotide-resolution genome-wide mapping of oxidative DNA damage by click-code-Seq. J. Am. Chem. Soc 2018, 140, 9783–9787. [PubMed: 29944356]

(15). Amente S; Di Palo G; Scala G; Castrignano T; Gorini F; Cocozza S; Moresano A; Pucci P; Ma B; Stepanov I; Lania L; Pelicci PG; Dellino GI; Majello B Genome-wide mapping of 8-oxo-7,8-dihydro-2'-deoxyguanosine reveals accumulation of oxidatively-generated damage at DNA replication origins within transcribed long genes of mammalian cells. Nucleic Acids Res 2019, 47, 221–236. [PubMed: 30462294]

(16). Hu JC; Lieb JD; Sancar A; Adar S Cisplatin DNA damage and repair maps of the human genome at single-nucleotide resolution. Proc. Natl. Acad. Sci. U.S.A 2016, 113, 11507–11512. [PubMed: 27688757]

(17). Riedl J; Ding Y; Fleming AM; Burrows CJ Identification of DNA lesions using a third base pair for amplification and nanopore sequencing. Nat. Commun 2015, 6, 8807. [PubMed: 26542210]

(18). Riedl J; Fleming AM; Burrows CJ Sequencing of DNA lesions facilitated by site-specific excision via base excision repair DNA glycosylases yielding ligatable gaps. J. Am. Chem. Soc 2016, 138, 491–494. [PubMed: 26741640]

(19). Shu XT; Liu MH; Lu ZK; Zhu CX; Meng HW; Huang SH; Zhang XX; Yi CQ Genome-wide mapping reveals that deoxyuridine is enriched in the human centromeric DNA. Nat. Chem. Biol 2018, 14, 680. [PubMed: 29785056]

(20). Evans MD; Dizdaroglu M; Cooke MS Oxidative DNA damage and disease: induction, repair and significance. Mutat. Res 2004, 567, 1–61. [PubMed: 15341901]

(21). Ikeda S; Biswas T; Roy R; Izumi T; Boldogh I; Kurosky A; Sarker AH; Seki S; Mitra S Purification and characterization of human NTH1, a homolog of Escherichia coli endonuclease III. Direct identification of Lys-212 as the active nucleophilic residue. J. Biol. Chem 1998, 273, 21585–93. [PubMed: 9705289]

(22). Hazra TK; Izumi T; Boldogh I; Imhoff B; Kow YW; Jaruga P; Dizdaroglu M; Mitra S Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. Proc. Natl. Acad. Sci. U.S.A 2002, 99, 3523–8. [PubMed: 11904416]

(23). Rosenquist TA; Zaika E; Fernandes AS; Zharkov DO; Miller H; Grollman AP The novel DNA glycosylase, NEIL1, protects mammalian cells from radiation-mediated cell death. DNA Repair 2003, 2, 581–91. [PubMed: 12713815]

(24). Fromme JC; Verdine GL Structure of a trapped endonuclease III-DNA covalent intermediate. EMBO J 2003, 22, 3461–71. [PubMed: 12840008]

(25). Schrock RD; Lloyd RS Reductive Methylation of the Amino Terminus of Endonuclease-V Eradicates Catalytic Activities - Evidence for an Essential Role of the Amino Terminus in the Chemical Mechanisms of Catalysis. J. Biol. Chem 1991, 266, 17631–17639. [PubMed: 1894643]

(26). Sun B; Latham KA; Dodson ML; Lloyd RS Studies on the catalytic mechanism of five DNA glycosylases. Probing for enzyme-DNA imino intermediates. J. Biol. Chem 1995, 270, 19501–8. [PubMed: 7642635]

(27). Girard PM; Guibourt N; Boiteux S The Ogg1 protein of *Saccharomyces cerevisiae*: a 7,8-dihydro-8-oxoguanine DNA glycosylase/AP lyase whose lysine 241 is a critical residue for catalytic activity. Nucleic Acids Res 1997, 25, 3204–11. [PubMed: 9241232]

(28). Frelon S; Douki T; Ravanat JL; Pouget JP; Tornabene C; Cadet J High-performance liquid chromatography-tandem mass spectrometry measurement of radiation-induced base damage to isolated and cellular DNA. Chem. Res. Toxicol 2000, 13, 1002–10. [PubMed: 11080049]

(29). Bailey DT; DeFedericis HC; Greene KF; Iijima H; Budzinski EE; Patrzyc HB; Dawidzik JB; Box HC A novel approach to DNA damage assessments: measurement of the thymine glycol lesion. Radiat. Res 2006, 165, 438–44. [PubMed: 16579656]

(30). He X; Yuan J; Wang Y G3BP1 binds to guanine quadruplexes in mRNAs to modulate their stabilities. Nucleic Acids Res 2021, 49, 11323–11336. [PubMed: 34614161]

(31). Yeo JE; Wickramaratne S; Khatwani S; Wang YC; Vervacke J; Distefano MD; Tretyakova NY Synthesis of Site-Specific DNA-Protein Conjugates and Their Effects on DNA Replication. ACS Chem. Biol 2014, 9, 1860–1868. [PubMed: 24918113]

(32). Hazra TK; Das A; Das S; Choudhury S; Kow YW; Roy R Oxidative DNA damage repair in mammalian cells: a new perspective. DNA Repair (Amst) 2007, 6, 470–80. [PubMed: 17116430]

(33). Hu J; Adar S; Selby CP; Lieb JD; Sancar A Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. Genes Dev 2015, 29, 948–60. [PubMed: 25934506]

(34). Spivak G Transcription-coupled repair: an update. Arch. Toxicol 2016, 90, 2583–2594. [PubMed: 27549370]

(35). Chakraborty A; Tapryal N; Islam A; Mitra S; Hazra T Transcription coupled base excision repair in mammalian cells: So little is known and so much to uncover. DNA Repair (Amst) 2021, 107, 103204. [PubMed: 34390916]

(36). Menoni H; Wienholz F; Theil AF; Janssens RC; Lens H; Campalans A; Radicella JP; Marteijn JA; Vermeulen W The transcription-coupled DNA repair-initiating protein CSB promotes

XRCC1 recruitment to oxidative DNA damage. Nucleic Acids Res 2018, 46, 7747–7756. [PubMed: 29955842]

(37). Fleming AM; Ding Y; Burrows CJ Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. Proc. Natl. Acad. Sci. U.S.A 2017, 114, 2604–2609. [PubMed: 28143930]

(38). Brown KL; Roginskaya M; Zou Y; Altamirano A; Basu AK; Stone MP Binding of the human nucleotide excision repair proteins XPA and XPC/HR23B to the 5R-thymine glycol lesion and structure of the cis-(5R,6S) thymine glycol epimer in the 5′-GTgG-3′ sequence: destabilization of two base pairs at the lesion site. Nucleic Acids Res 2010, 38, 428–40. [PubMed: 19892827]

(39). Kumar N; Raja S; Van Houten B The involvement of nucleotide excision repair proteins in the removal of oxidative DNA damage. Nucleic Acids Res 2020, 48, 11227–11243. [PubMed: 33010169]

(40). Liu MM; Bandaru V; Bond JP; Jaruga P; Zhao XB; Christov PP; Burrows CJ; Rizzo CJ; Dizdaroglu M; Wallace SS The mouse ortholog of NEIL3 is a functional DNA glycosylase in vitro and in vivo. P Natl. Acad. Sci. USA 2010, 107, 4925–4930.

(41). Hinz JM; Rodriguez Y; Smerdon MJ Rotational dynamics of DNA on the nucleosome surface markedly impact accessibility to a DNA repair enzyme. Proc. Natl. Acad. Sci. U.S.A 2010, 107, 4646–51. [PubMed: 20176960]

(42). Olmon ED; Delaney S Differential ability of five DNA glycosylases to recognize and repair damage on nucleosomal DNA. ACS Chem. Biol 2017, 12, 692–701. [PubMed: 28085251]

(43). Wang Z; Zang C; Rosenfeld JA; Schones DE; Barski A; Cuddapah S; Cui K; Roh TY; Peng W; Zhang MQ; Zhao K Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet 2008, 40, 897–903. [PubMed: 18552846]

(44). Lohe AR; Hilliker AJ; Roberts PA Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. Genetics 1993, 134, 1149–74. [PubMed: 8375654]

(45). Adar S; Hu JC; Lieb JD; Sancar A Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. Proc. Natl. Acad. Sci. U.S.A 2016, 113, E2124–E2133. [PubMed: 27036006]

(46). Charlet-Berguerand N; Feuerhahn S; Kong SE; Ziserman H; Conaway JW; Conaway R; Egly JM RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. EMBO J 2006, 25, 5481–5491. [PubMed: 17110932]

(47). Alhegaili AS; Ji Y; Sylvius N; Blades MJ; Karbaschi M; Tempest HG; Jones GDD; Cooke MS Genome-Wide Adductomics Analysis Reveals Heterogeneity in the Induction and Loss of Cyclobutane Thymine Dimers across Both the Nuclear and Mitochondrial Genomes. Int. J. Mol. Sci 2019, 20, 5112.

(48). Wauchope OR; Mitchener MM; Beavers WN; Galligan JJ; Camarillo JM; Sanders WD; Kingsley PJ; Shim HN; Blackwell T; Luong T; deCaestecker M; Fessel JP; Marnett LJ Oxidative stress increases M1dG, a major peroxidation-derived DNA adduct, in mitochondrial DNA. Nucleic Acids Res 2018, 46, 3458–3467. [PubMed: 29438559]

(49). Maher RL; Prasad A; Rizvanova O; Wallace SS; Pederson DS Contribution of DNA unwrapping from histone octamers to the repair of oxidatively damaged DNA in nucleosomes. DNA Repair 2013, 12, 964–71. [PubMed: 24051050]

(50). Kennedy EE; Caffrey PJ; Delaney S Initiating base excision repair in chromatin. DNA Repair 2018, 71, 87–92. [PubMed: 30170831]

(51). Sander JD; Joung JK CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol 2014, 32, 347–55. [PubMed: 24584096]

(52). Li L; Williams P; Ren W; Wang MY; Gao Z; Miao W; Huang M; Song J; Wang Y YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. Nat. Chem. Biol 2021, 17, 161–168. [PubMed: 33199912]

(53). Bauer DE; Canver MC; Orkin SH Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. J. Vis. Exp 2015, e52118. [PubMed: 25549070]

(54). Chen CF; Tan RY; Wong L; Fekete R; Halsey J Quantitation of MicroRNAs by Real-Time RT-qPCR. Methods Mol. Biol 2011, 687, 113–134. [PubMed: 20967604]

(55). Langmead B; Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat. Methods 2012, 9, 357–359. [PubMed: 22388286]

(56). Feng JX; Liu T; Qin B; Zhang Y; Liu XS Identifying ChIP-seq enrichment using MACS. Nat. Protoc 2012, 7, 1728–1740. [PubMed: 22936215]

(57). Nicol JW; Helt GA; Blanchard SG; Raja A; Loraine AE The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 2009, 25, 2730–2731. [PubMed: 19654113]

(58). Huang WC; Loganantharaj R; Schroeder B; Fargo D; Li LP PAVIS: a tool for Peak Annotation and Visualization. Bioinformatics 2013, 29, 3097–3099. [PubMed: 24008416]

(59). Quinlan AR; Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010, 26, 841–2. [PubMed: 20110278]

(60). Ramirez F; Dundar F; Diehl S; Gruning BA; Manke T deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 2014, 42, W187–91. [PubMed: 24799436]

(61). Davis CA; Hitz BC; Sloan CA; Chan ET; Davidson JM; Gabdank I; Hilton JA; Jain K; Baymuradov UK; Narayanan AK; Onate KC; Graham K; Miyasato SR; Dreszer TR; Strattan JS; Jolanki O; Tanaka FY; Cherry JM The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res 2018, 46, D794–D801. [PubMed: 29126249]

(62). Machanick P; Bailey TL MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics 2011, 27, 1696–1697. [PubMed: 21486936]

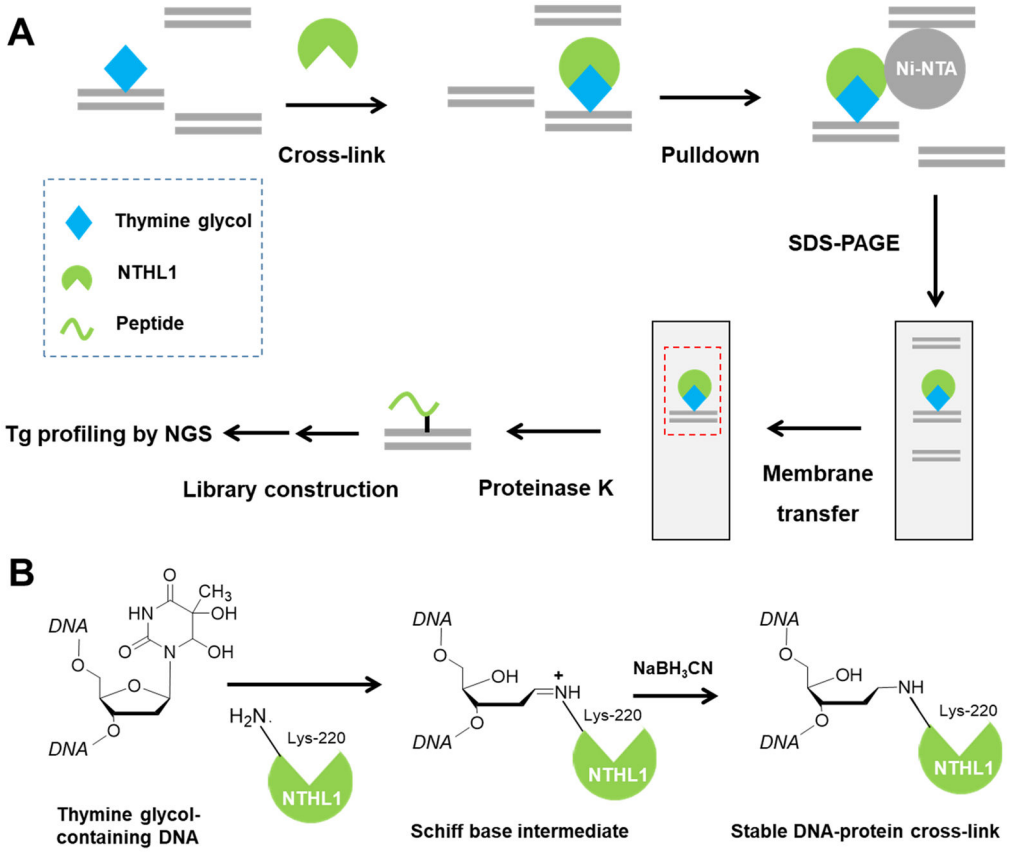


Figure 1.

DNA–protein cross-link sequencing (DPC-Seq) for genome-wide profiling of Tg. (A) A schematic diagram illustrating the workflow of DPC-Seq. Fragmented genomic DNA was first cross-linked with NTHL1-His_6 to form DPC. Subsequently, DPC was stabilized by $NaBH_4$ reduction and enriched with Ni-NTA beads. The eluents were further purified by SDS-PAGE and membrane transfer. After proteinase K digestion, DNA was recovered from the membrane and employed for library construction and next-generation sequencing. (B) NTHL1-mediated cleavage of Tg, DPC formation, and $NaBH_3CN$ reduction of the ensuing Schiff base intermediate.

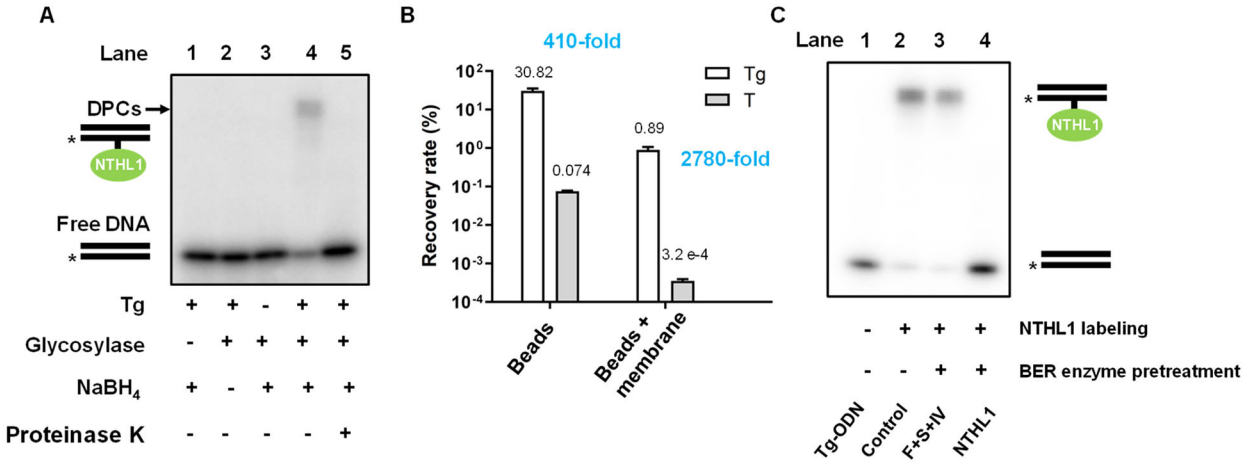


Figure 2.

NTHL1-mediated covalent labeling and enrichment of Tg-containing DNA. (A) SDS-PAGE analysis of the Tg-containing DPC products. (B) DPC-Seq enables efficient enrichment of Tg-containing DNA. The enrichment fold was calculated based on the ratio of recovery rate of Tg-ODN over that of the corresponding unmodified ODN, where the recovery rates were calculated from qPCR results in three replicates. Error bars represent standard deviations. (C) Tg-ODN pretreated with BER enzymes: Lane 1, Tg-ODN control without labeling; Lane 2, labeling of Tg-ODN without pretreatment with BER enzymes; Lane 3, the sample was pretreated with Fpg (F), SMUG1 (S), and EndoIV (IV) prior to cross-linking with NTHL1; Lane 4, absence of NTHL1 labeling of Tg-ODN when the sample was pretreated with EndoIII and EndoIV.

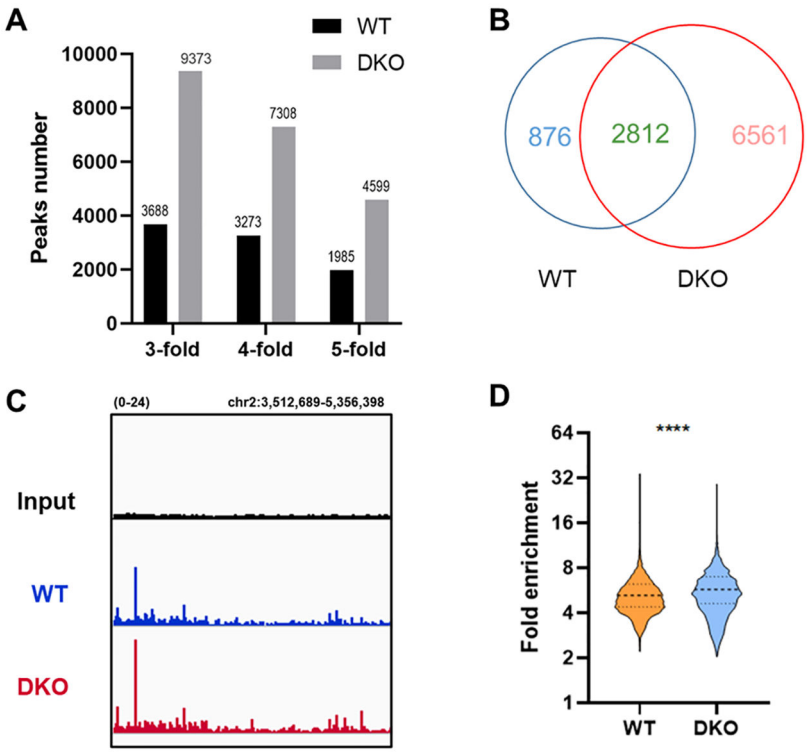


Figure 3.

Analysis of NGS reads obtained from DPC-Seq of genomic DNA isolated from HEK293T and the isogenic DKO cell lines. (A) Peak numbers obtained from the two cell lines with at least 3-, 4-, and 5-fold enrichment relative to input. (B) A Venn diagram illustrating the overlap of high-confidence peaks between the two cell lines. (C) Genome browser view displaying the peak enrichment in parental and DKO HEK293T cells. (D) Fold enrichment of overlapped peaks between the two cell lines.

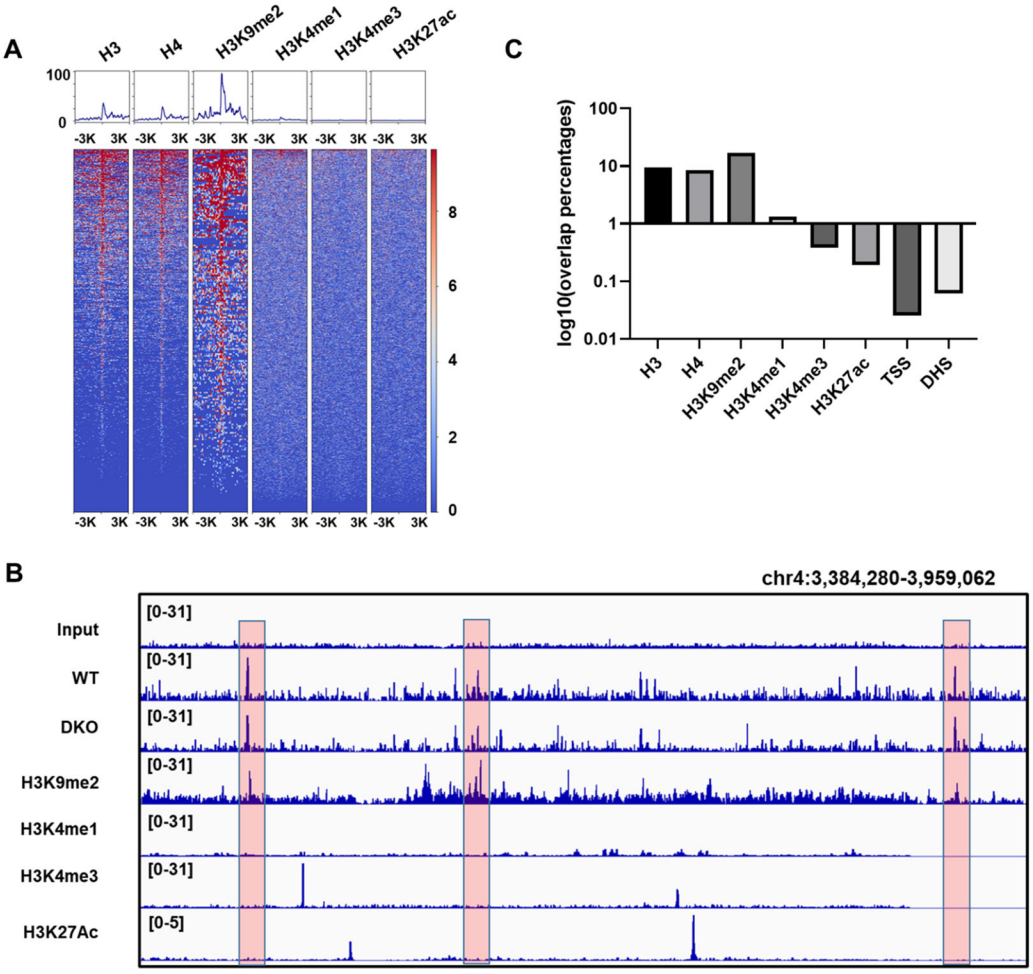


Figure 4.

Tg peaks are enriched in a nucleosome region, especially in a heterochromatin region. (A) The heatmap representation of Tg signals in the regions ±3 kb from the centers of H3, H4, H3K9me2, H3K4me1, H3K4me3, and H3K27ac peaks in HEK293T cells. (B) Genome browser view to illustrate peak colocalization between Tg and the corresponding histone marks. (C) Percentages of the H3, H4, H3K9me2, H3K4me1, H3K4me3, H3K27ac, TSS, and DHS peaks that overlapped with Tg peaks. Overlap percentages = (number of peaks overlapped with Tg peaks/number of peaks in the whole genome) × 100%.