Supporting Information

SI Materials and Methods

UVDE-seq and UVDE-HS-seq library preparation and sequencing

For yeast library preparation, wild-type (BY4741) and *rad16*Δ cells were grown in yeast extract peptone dextrose (YPD) medium to an OD_{600} of approximately 0.8 (between 0.6 and 1.0). Cells were then pelleted and re-suspended in sterile Millipore water. Cells were taken for the "No UV" control group, while the remaining cells were exposed to 600 J/m^2 UVC light. Cells were then pelleted and stored at -80^oC until genomic DNA (gDNA) was isolated using phenol:chloroform:isoamylalcohol (PCI), lysis buffer, and glass beads. Cells used for naked DNA samples were cultured as described, but not exposed to UVC. Following gDNA isolation, samples were spotted onto a microscope cover slide on ice in 10µL aliquots. Samples were then exposed to 400J/m^2 or 500J/m^2 UVC light (approximately 80% of the dose exposed to cellular samples). Samples were recovered from the cover slide and combined in a TPX tube for sonication (described below). At least two replicate samples were prepared and sequenced for each UV-irradiated yeast strain (e.g., WT, *rad16*Δ) and processed using UVDE-seq.

Following isolation of yeast genomic DNA, purified gDNA was sonicated using a Bioruptor 300 Sonicator (Diagenode, UCD-300 TM) for 15 cycles (30 second ON/OFF intervals) to create fragments between 200 and 500bp in length. DNA fragments were then end-repaired (NEB, E6050L) and dA-tailed (NEB, E6053L), and a double stranded trP1 adapter was ligated to both ends of the fragments via a quick ligase reaction (NEB, E6056L). trP1 adapter ligation was PCR confirmed using primers complimentary to the

trP1 adapter. Following confirmation, free 3'-OH groups were blocked with Terminal Transferase (NEB, M0315L) and either dideoxy ATP (ddATP) or dideoxy GTP (ddGTP; Roche Diagnostics, 03732738001). Samples were then treated with *E. coli* CPD photolyase and incubated under UVA light for 2 hours at room temperature. It is important to note that this UVA treatment could convert many of the 6-4PP to Dewar isomers (1), which presumably are still cleaved by UVDE. DNA fragments were purified using a phenol:chloroform:isoamylalcohol (PCI) extraction, followed by ethanol precipitation. Those libraries prepared using the UVDE-HS-seq method were treated with uracil DNA glycosylase inhibitor (UGI) during CPD photolyase treatment, and under the same conditions. Samples were then treated with *T. thermophilus* UVDE for 45 minutes at 55^{o}C. 5' phosphate groups were removed using shrimp alkaline phosphatase (Affymetrix, AF78390500), and DNA was subsequently denatured at 95^{o}C for 5 minutes and snap-cooled on ice. A second double stranded adapter, the A adapter, was then ligated to the 3'-OH created immediately downstream from the cleaved UV lesion (NEB, E6056L) and PCR confirmed using a Cy3-labeled primer complimentary to the A adapter. Each A adapter contains a unique barcode that allows for downstream sample identification following multiplexed DNA sequencing techniques, as well as a biotin-labeled strand. DNA containing the biotin label was purified using Streptavidin beads (Thermo Fisher Scientific, 11205D), while the DNA strand lacking the biotin label was removed using 0.15M NaOH. The remaining single stranded DNA was then used as a template for second strand synthesis, in which the second strand of adapter A was used as the extension primer. Libraries were PCR amplified for 8 cycles using primers complimentary to the trP1 and A adapters. Finally, samples were

combined typically with equal volumes and submitted for Ion Proton sequencing (Life Technologies).

Oligonucleotides used for UVDE-seq

The following oligonucleotides were used for the first ligation:

trP1-top (5′-CCTCTCTATGGGCAGTCGGTGAT-phosphorothioate-T-3′)

trP1-bottom (5′-phosphate-ATCACCGACTGCCCATAGAGAGGC-dideoxy-3′).

The following barcoded oligonucleotides (A1, A2, A3, A4, A5, or A6) were used for the second ligation:

A1-top (5′-phosphate-ATCCTCTTCTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3′), A1-bottom (5′-biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATNNNNNN-C3 phosphoramidite-3′);

A2-top (5′-phosphate-ATCACGAACTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3′), A2-bottom (5′-biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATNNNNNN-C3 phosphoramidite-3′);

A3-top (5′-phosphate-ATCTCAGGCTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3′), A3-bottom (5′-biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGAGATNNNNNN-C3 phosphoramidite-3′).

A4-top (5’-phosphate-ATCGCGATCTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3’); A4-bottom (5’- biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGATCGCGATNNNNNN-C3 phosphoramidite-3’);

A5-top (5′- phosphate-ATCCAGTACTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3′); A5-bottom (5′- biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTGGATNNNNNN-C3 phosphoramidite-3′)

A6-top (5′- phosphate-ATCAGTTCCTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3′); A6-bottom (5′- biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGGAACTGATNNNNNN-C3 phosphoramidite-3′)

Oligonucleotides used for PCR confirmation and library amplification:

Primer A (5′-CCATCTCATCCCTGCGTGTCTCCGAC-3′)

Primer trP1 (5′-CCTCTCTATGGGCAGTCGGTGATT-3′).

Analysis of UV photoproduct formation in nucleosomes

For nucleosome analysis, peak rotational and translational periodicity and their corresponding signal-to-noise ratio (SNR) was determined using the mutperiod software package (2). For rotational data, periods between 5 and 25 bp were analyzed, while periods between 50 and 250 bp were analyzed for translational periodicities. The SNR statistic essentially quantifies the relative strength of the peak periodicity relative to the

median strength of the other periodicities analyzed, as previously described (2, 3). In plotting the rotational data, outward rotational settings (dashed lines in Fig. 2) were indicated every 10.17 bp, centered at the nucleosome dyad (position 0). In plotting the translational data, for yeast nucleosomes a 165 bp nucleosome repeat length was used (i.e., 147 bp of nucleosomal DNA and 18 bp of linker DNA).

Analysis of UV photoproduct formation in TFBS

We analyzed 6-4PP and TA-PP formation at published TF binding sites identified by ChIP-exo (4, 5), using our previously described methods (6). We excluded all TF binding sites within 1 kb of a telomere end. For Reb1 binding sites, we also analyzed low occupancy sites (Reb1 occupancy < 10) identified in a previous study (7), excluded Reb1 sites that overlapped with repetitive genomic regions with aberrantly high sequencing depth, such as the ribosomal DNA locus, etc., as a control. Custom perl scripts were used to determine the density of lesions, typically normalized by the number of lesion-forming dipyrimidine or TA sequences, at positions relative to the TFBS midpoint. Lesion density derived from UVDE-seq libraries of UV-irradiated yeast was normalized using UV-irradiated naked DNA controls, which were scaled so that there were equal numbers of lesions as the cellular samples. This calculation yielded 6-4PP or TA-PP enrichment. Similar analysis was performed on published CPD-seq data for UV-irradiated yeast cells and the naked DNA controls (6). Cluster plot analysis of CPD or 6-4PP induction at individual Hap2/Hap3/Hap5 binding sites was performed using custom Perl scripts that calculated the difference between UV-irradiated cellular

CPDs or 6-4PPs at each location in each binding site relative to the scaled naked DNA control. Cluster plot graphics were generated using Java Treeview (8).

SI Figures

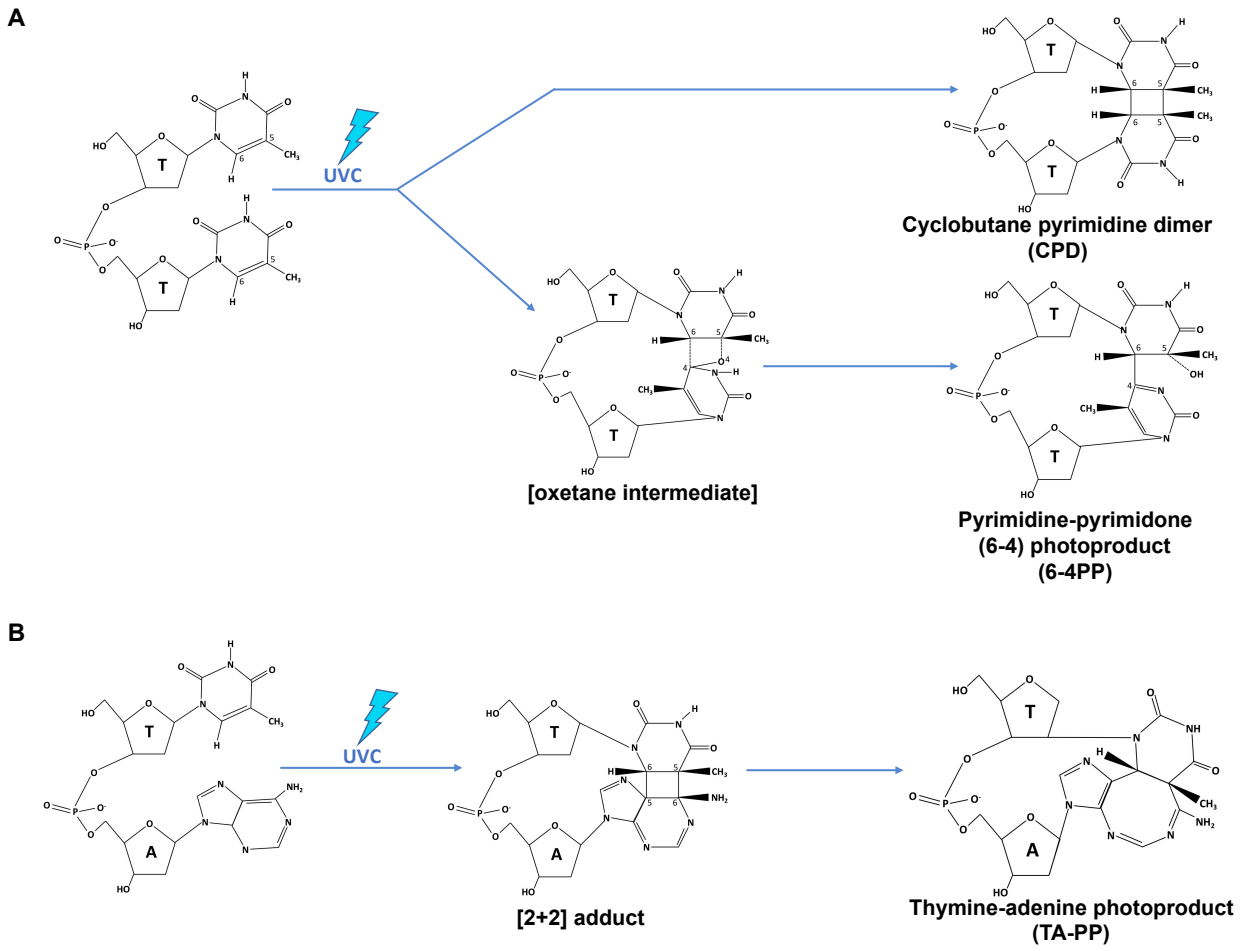

Figure S1. Structure of canonical UV-induced DNA lesions. (**A)** Top panel shows the formation of cyclobutane pyrimidine dimers (CPDs), which are formed rapidly and constitute approximately 75% of the damage induced by UV light. The bottom panel displays the formation of pyrimidine-pyrimidone (6-4) photoproducts (6-4PPs). These lesions are formed more slowly than CPDs and form an oxetane intermediate before forming their final structure. 6-4PPs constitute about 25% of the damage formed by UV light and are thus considered rare when compared to CPDs. (**B)** Formation of thymine-adenine photoproducts (TA-PPs). Upon exposure to UV-C light, neighboring thymine

and adenine bases will bond at their C5-C6 carbons to form an intermediate [2+2] adduct before forming the final TA-PP lesion. Lesion structures adapted from (9).

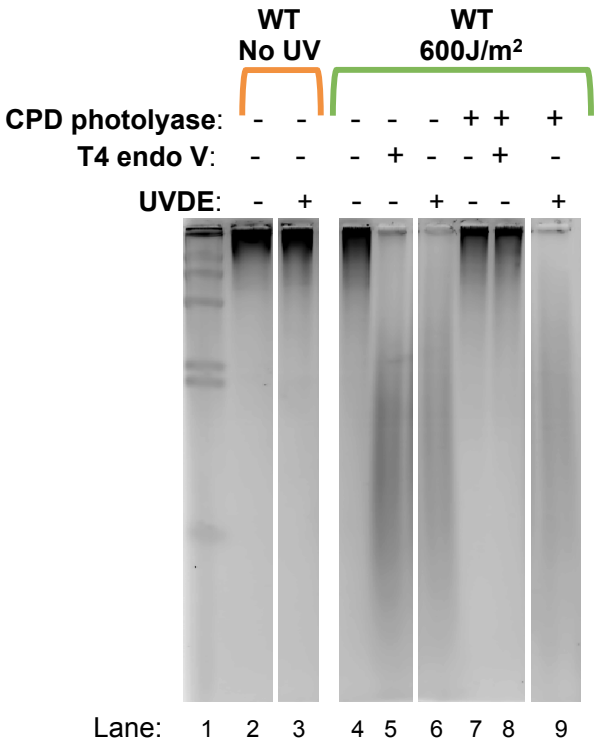

Figure S2. Alkaline gel optimization of UVDE-seq method. Wild-type yeast were either untreated (No UV) or treated with 600J/m^2 UVC light. Genomic DNA (gDNA) was isolated and samples were treated solely or in combination with CPD photolyase, T4 endonuclease V, and UVDE and run on a 1.2% alkaline gel at 30V for 20 hours to resolve DNA fragments. Samples treated with CPD photolyase were incubated at room temperature under UVA light for 2 hours. Samples treated with T4 endonuclease V (T4 endo V) were incubated at 37^{o}C for 1.5 hours. Samples treated with UVDE were incubated at 55^{o}C for 45 minutes. Optimal CPD removal by photolyase was confirmed when similar DNA fragmenting patterns were seen in CPD photolyase treated samples compared to samples treated with both CPD photolyase and T4 endo V (compare lanes 7 and 8). Optimal cleavage of remaining rare (i.e., 6-4PPs) and atypical (i.e., TA-PP)

lesions was obtained using 3μL of UVDE (lane 9). Gel image was cropped to eliminate lanes containing samples not pertinent to this study.

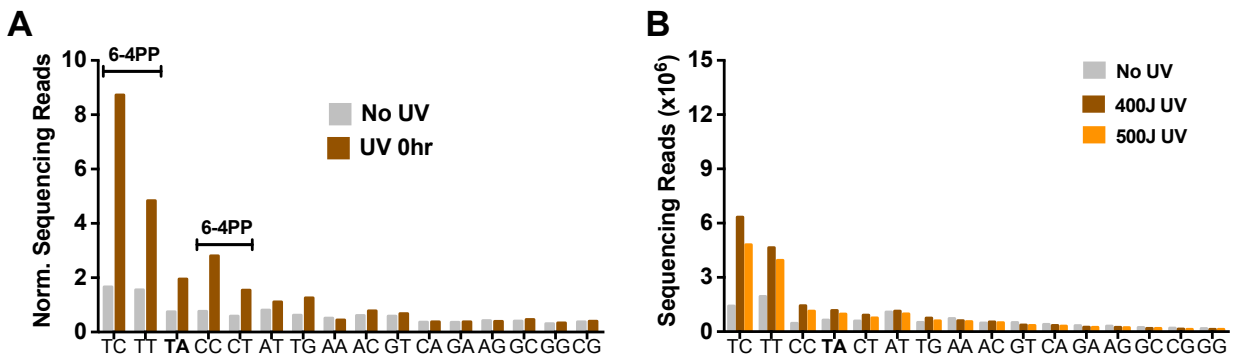

Figure S3. Dinucleotide counts of putative lesions associated with cellular and naked DNA UVDE-seq reads. (**A**) UVDE-seq reads for un-irradiated (No UV) and UV-irradiated cells (UV 0hr) associated with lesions at different dinucleotides, normalized to the frequency of each dinucleotide in the yeast genome. (**B**) Number of UVDE-seq reads associated with different dinucleotide sequences for UV-irradiated naked DNA (either 400J/m^2 or 500J/m^2 of UVC light) relative to the No UV control.

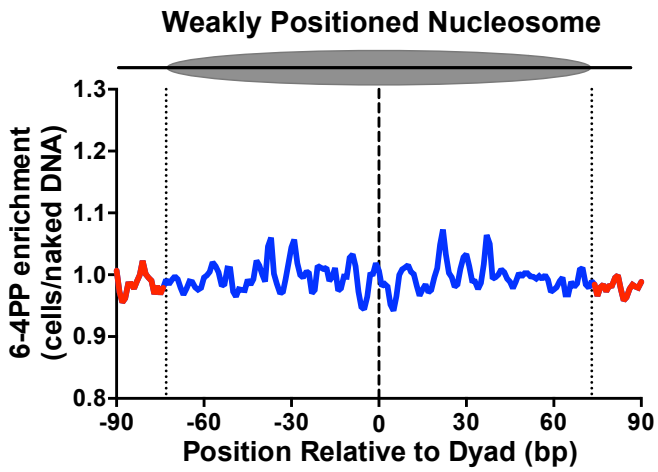

Figure S4. Analysis of translational periodicity of 6-4PP enrichment in weakly positioned nucleosomes. Nucleosomal DNA is in blue and linker DNA is depicted in red. Dashed lines at -73 and +73 represent the border between nucleosomal and linker DNA.

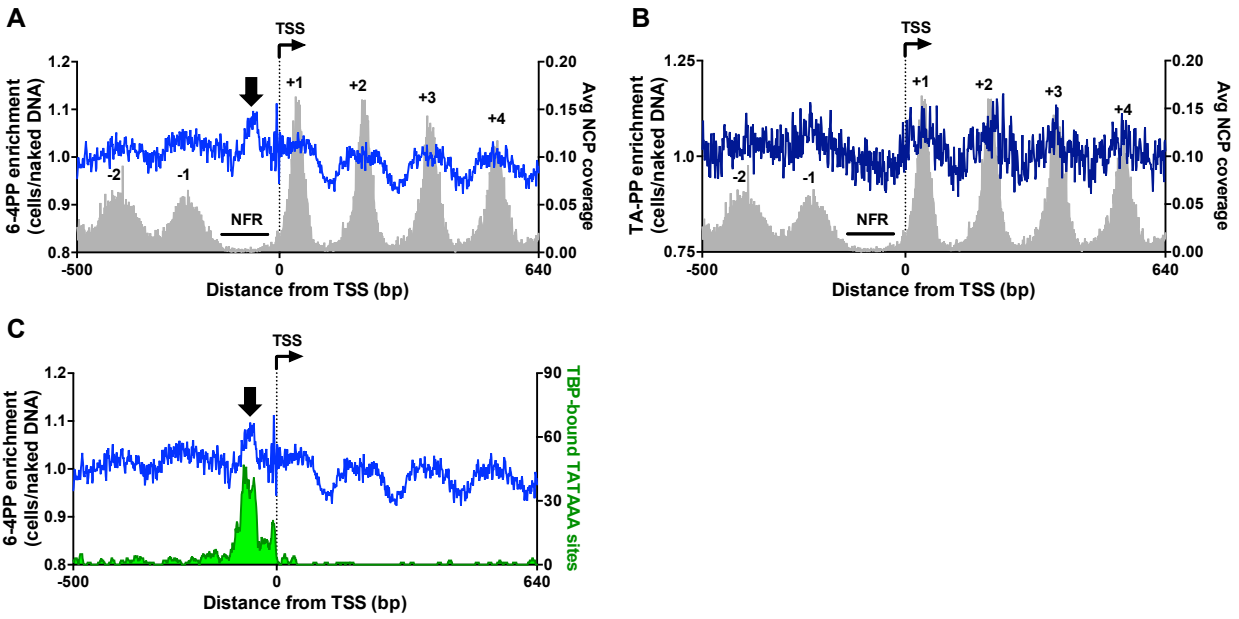

Figure S5. Analysis of 6-4PP and TA-PP enrichment around the transcription start site (TSS) of ~5200 yeast genes. (**A**) 6-4PP and (**B**) TA-PP enrichment was calculated by normalizing 6-4PP or TA-PP counts from UVDE-HS-seq data in UV-irradiated yeast cells to the scaled naked DNA control. TSS coordinates are from (10). Average nucleosome core particle (NCP) coverage is derived from (11), as previously described (12). (**C**) Same as panel **A**, except the number of TBP-bound TATAAA sites is plotted. Each binding site consisted of a 9 bp DNA segment.

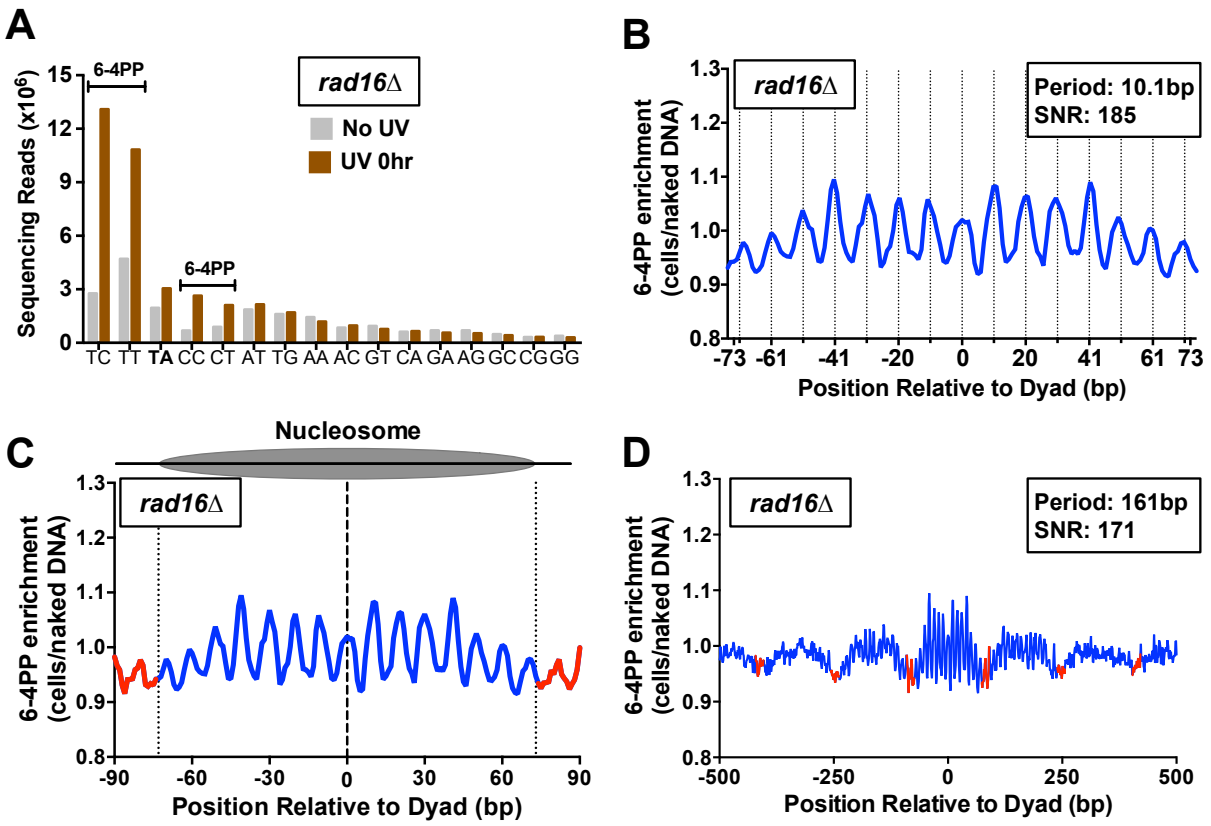

Figure S6. (**A)** UVDE-seq reads mapping putative lesions at different dinucleotide sequences in a GG-NER deficient *rad16*Δ mutant. Enrichment of reads observed at canonical dipyrimidine sequences (TC, TT, CC, CT) as well as atypical TA sequences in comparison to unexposed (No UV) samples. (**B)** Rotational periodicity of 6-4PPs within all (~10,000) strongly positioned yeast nucleosomes. Dashed lines correspond to outward rotational settings, and 0 represents the dyad, or center, of the nucleosome. Nucleosome map from (13). Data plotted as cellular lesion counts normalized to naked DNA control. Peak rotational periodicity and its SNR was calculated using mutperiod (2). (**C)** Translational periodicity examining 6-4PP enrichment in nucleosomal and linker DNA in strongly positioned nucleosomes. Dashed lines at -73 and +73 represent the

border between nucleosomal and linker DNA, with linker DNA shown in red. (**D)** Translational periodicity of 6-4PPs spanning numerous strongly positioned yeast nucleosomes. Nucleosomal DNA is shown in blue, and predicted linker DNA is shown in red. Peak translational periodicity and its SNR was calculated using mutperiod (2).

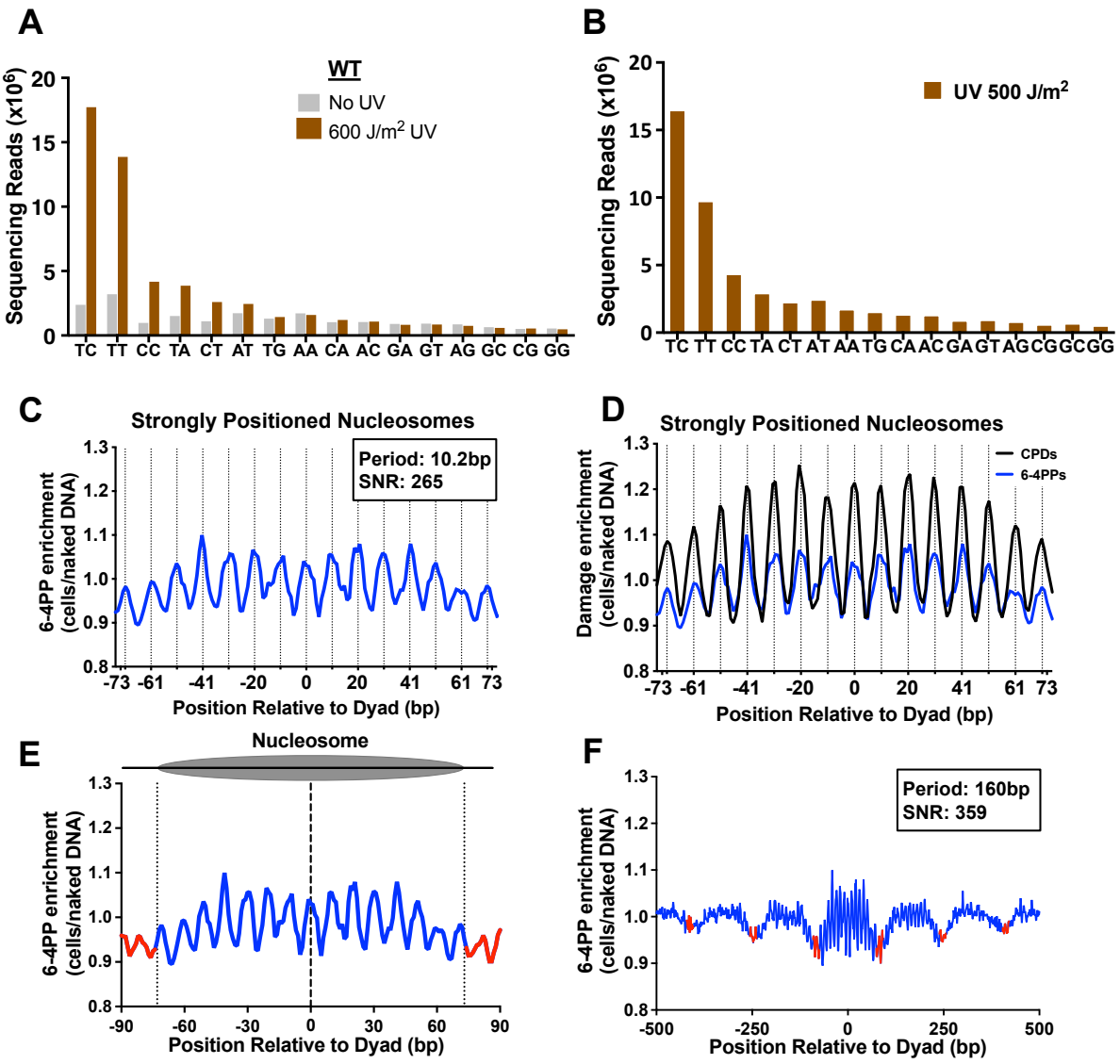

Figure S7. (**A)** UVDE-HS-seq reads mapping putative lesions at different dinucleotide sequences in UV-irradiated (600 J/m^2) WT yeast cells. Enrichment of reads is observed at canonical dipyrimidine sequences (TC, TT, CC, CT), as well as atypical TA sequences, in comparison to unexposed (No UV) samples. (**B**) Same as panel **A**, except for UV-irradiated (500 J/m^2) naked DNA control. (**C**) Rotational periodicity of cellular 6-4PPs, as measured by UVDE-HS-seq, within all (~10,000) strongly positioned

yeast nucleosomes (13) in WT yeast. Data plotted as cellular lesion counts normalized to a naked DNA control. Peak rotational periodicity and its SNR was calculated using mutperiod (2). (**D**) Same as panel **C**, except cellular CPD-seq data (6) is also plotted. (**E**) Translational periodicity of 6-4PP enrichment from UVDE-HS-seq data in nucleosomal and linker DNA in strongly positioned nucleosomes (13). Dashed lines at -73 and +73 represent the border between nucleosomal and linker DNA, with linker DNA shown in red. (**F**) Translational periodicity of 6-4PPs in 1000 bp window centered on strongly positioned yeast nucleosomes. Nucleosomal DNA is shown in blue, predicted linker DNA is shown in red. UVDE-HS-seq data plotted is cellular lesion counts normalized to a naked DNA control. Peak translational periodicity and its SNR was calculated using mutperiod (2).

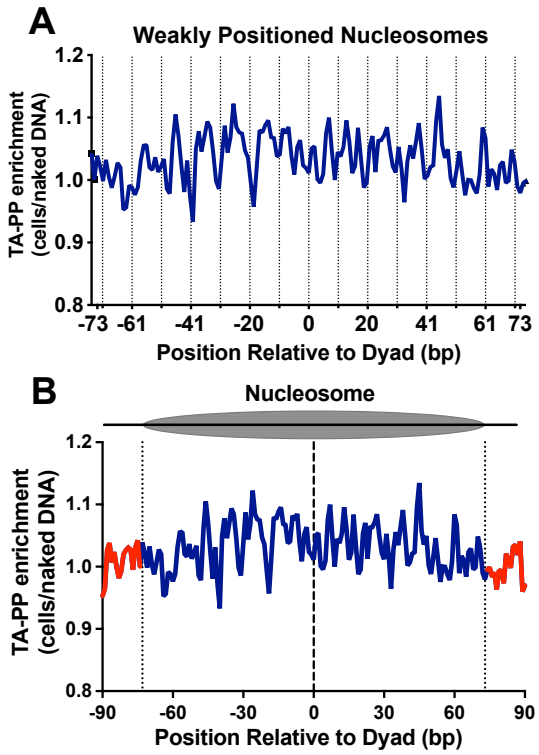

Figure S8. (**A**) Rotational periodicity of cellular TA-PPs, as measured by UVDE-HS-seq, within ~7500 weakly positioned yeast nucleosomes (13) in WT yeast. Data plotted as cellular lesion counts normalized to a naked DNA control. (**B**) Translational periodicity of TA-PP enrichment (from UVDE-HS-seq data) in nucleosomal and linker DNA in weakly positioned nucleosomes (13). Dashed lines at -73 and +73 represent the border between nucleosomal and linker DNA, with linker DNA shown in red.

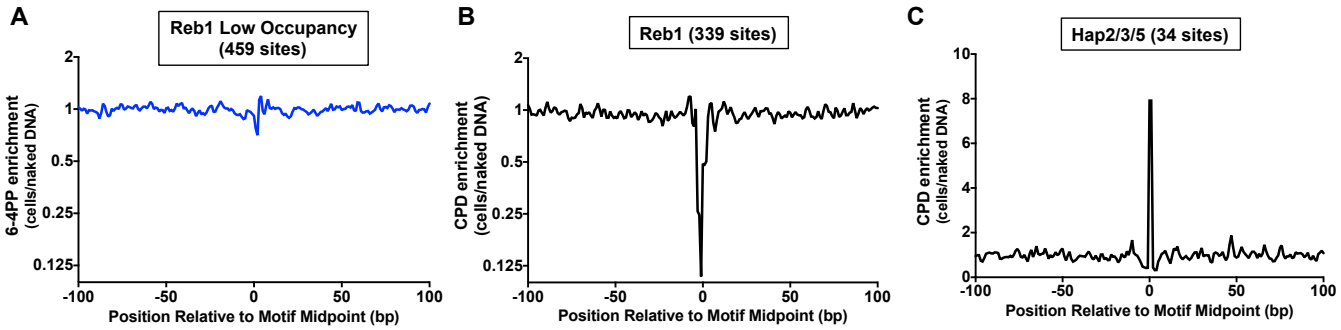

Figure S9. (**A**) Formation of 6-4PPs at control low occupancy Reb1 binding sites (identified by (7)). Data plotted as cellular lesion counts normalized to a naked DNA control. (**B-C**) Formation of CPD lesions at (**B**) Reb1 and (**C**) Hap2/3/5 binding sites. Binding site data from ChIP-exo studies (4). CPD-seq data from (6).

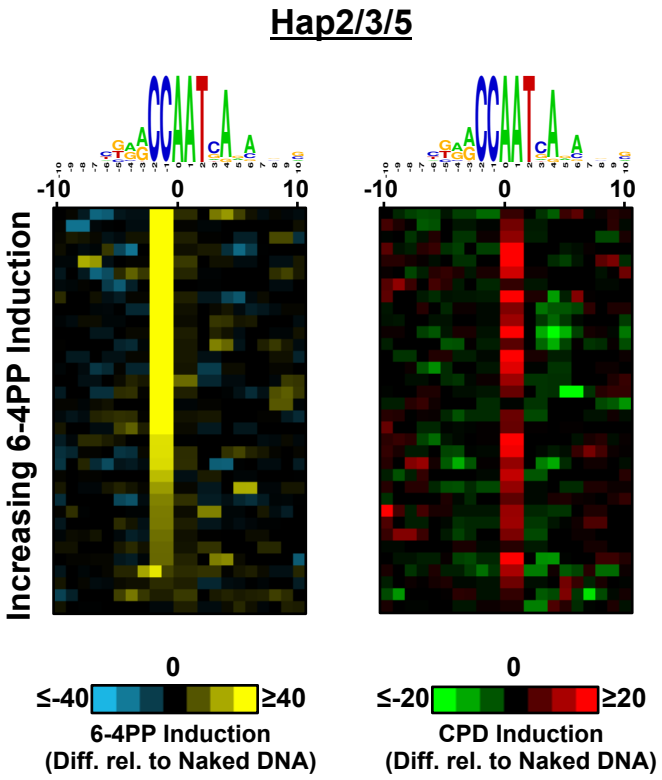

Figure S10. 6-4PP and CPD induction occur at overlapping Hap2/Hap3/Hap5 binding sites. Cluster plot showing the level of 6-4PP (left panel) and CPD (right panel) induction at 34 Hap2/Hap3/Hap5 binding sites. 6-4PP or CPD induction was calculated by determining the difference in the number of UVDE-seq or CPD-seq reads in UV-irradiated cells relative to scaled UV-irradiated naked DNA samples. Each row represents a single binding site, which are ordered based on increasing 6-4PP induction at position -1 in the Hap2/Hap3/Hap5 binding consensus. The left and right panels have the same row order. The color indicates the level of induction for 6-4PPs or CPDs. The consensus sequence of the Hap2/Hap3/Hap5 binding sites is shown above each cluster plot, and was created using weblogo (14).

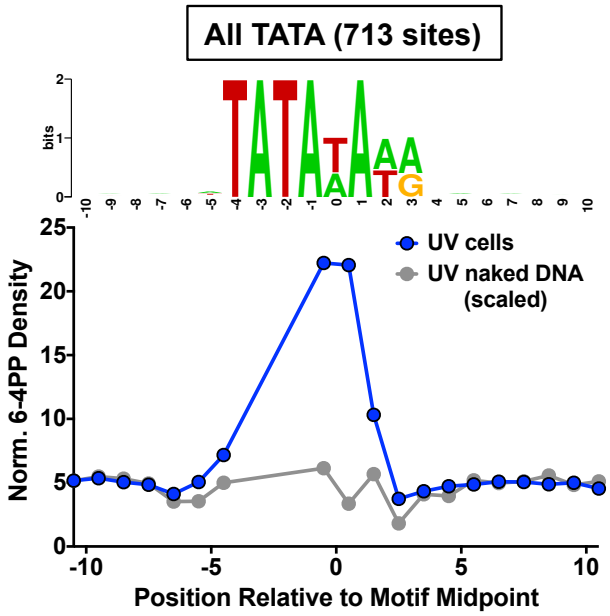

Figure S11. High-resolution analysis of 6-4PP formation within all yeast TBP-bound TATA sites identified by ChIP-exo (5). Top panel is the consensus sequence of yeast TBP binding sites, created using weblogo (14).

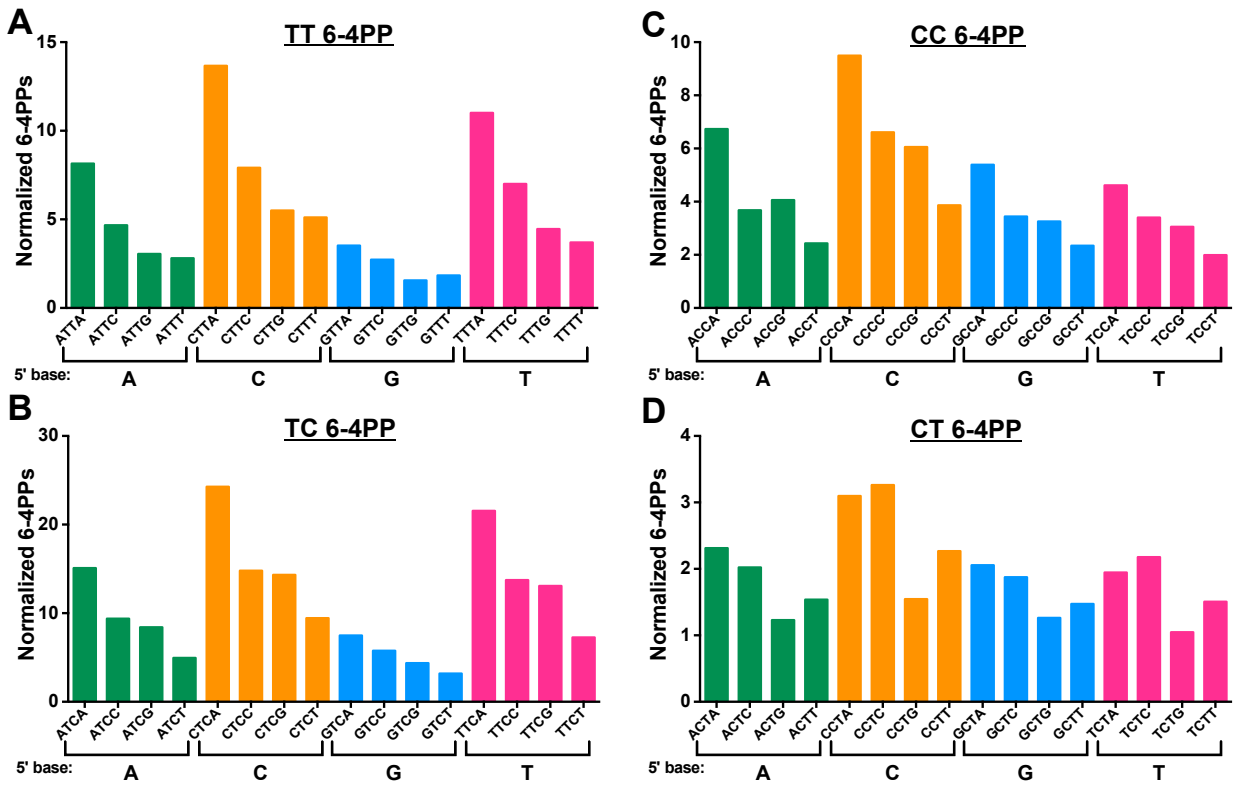

Figure S12. Normalized frequency of 6-4PP formation at different sequence contexts. (A-D) Frequency of UVDE-HS-seq reads in UV-irradiated yeast at different dipyrimidine-centered tetranucleotide sequences, normalized by the frequency of the tetranucleotide sequence in the yeast genome. Each panel shows the normalized frequency of one class of 6-4PP lesions (e.g., 6-4PP occurring at TT, TC, CC, or CT dinucleotides).

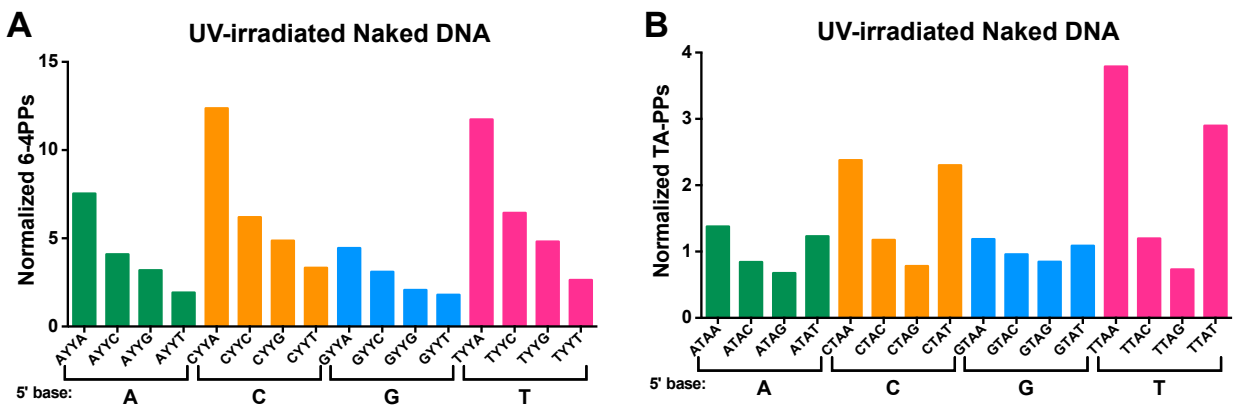

Figure S13. Normalized frequency of 6-4PP and TA-PP formation at different sequence contexts in UV-irradiated naked DNA. (A) Frequency of UVDE-HS-seq reads in UV-irradiated naked DNA for 6-4PP forming dipyrimidine sequences in different sequence contexts, normalized by the frequency of the tetranucleotide sequence in the yeast genome. Y = C or T nucleotide. (**B**) Same as panel A, except for UVDE-HS-seq reads associated with TA dinucleotides (i.e., TA-PP).

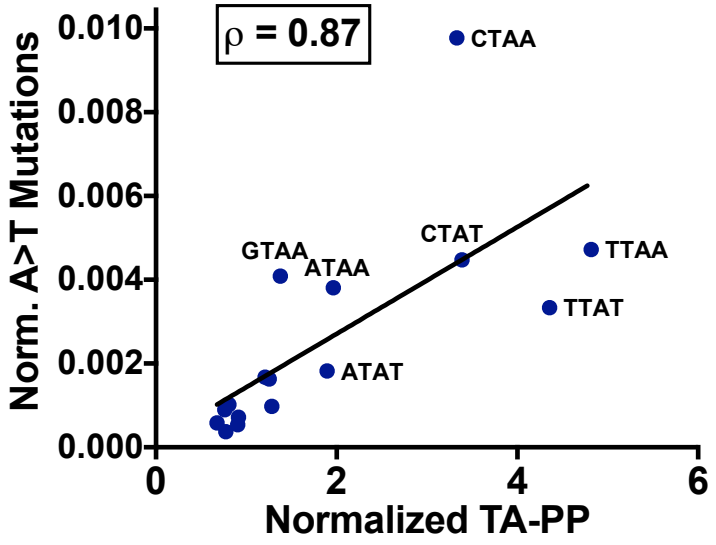

Figure S14. Frequency of UV-induced A>T substitutions in different sequence contexts is highly correlated with TA-PP formation. Frequency of UVDE-HS-seq reads in UV-irradiated cells at TA dinucleotides (i.e., TA-PP) in different sequence contexts, normalized by the frequency of each tetranucleotide sequence in the yeast genome is plotted relative to the frequency of A>T substitutions in UV-irradiated yeast, normalized in similar manner. Calculated Spearman correlation coefficient (ρ) is indicated, which is highly significant ($P < 0.0001$). Yeast UV mutation data is from (15).

References

1. Friedberg EC*, et al.* (2006) *DNA repair and Mutagenesis* (ASM Press, Washington, D.C.) 2nd Ed pp xxvii, 1118 p.
2. Morledge-Hampton B & Wyrick JJ (2021) Mutperiod: Analysis of periodic mutation rates in nucleosomes. *Computational and structural biotechnology journal* 19:4177-4183.
3. Pich O*, et al.* (2018) Somatic and Germline Mutation Periodicity Follow the Orientation of the DNA Minor Groove around Nucleosomes. *Cell* 175(4):1074-1087 e1018.
4. Rossi MJ*, et al.* (2021) A high-resolution protein architecture of the budding yeast genome. *Nature* 592(7853):309-314.
5. Rhee HS & Pugh BF (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* 483(7389):295-301.
6. Mao P, Smerdon MJ, Roberts SA, & Wyrick JJ (2016) Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proceedings of the National Academy of Sciences of the United States of America* 113(32):9057-9062.
7. Kasinathan S, Orsi GA, Zentner GE, Ahmad K, & Henikoff S (2014) High-resolution mapping of transcription factor binding sites on native chromatin. *Nature methods* 11(2):203-209.

8. Saldanha AJ (2004) Java Treeview--extensible visualization of microarray data. *Bioinformatics* 20(17):3246-3248.
9. Wang Y, Gross ML, & Taylor JS (2001) Use of a combined enzymatic digestion/ESI mass spectrometry assay to study the effect of TATA-binding protein on photoproduct formation in a TATA box. *Biochemistry* 40(39):11785-11793.
10. Park D, Morris AR, Battenhouse A, & Iyer VR (2014) Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic acids research* 42(6):3736-3749.
11. Weiner A*, et al.* (2015) High-resolution chromatin dynamics during a yeast stress response. *Molecular cell* 58(2):371-386.
12. Mao P, Smerdon MJ, Roberts SA, & Wyrick JJ (2020) Asymmetric repair of UV damage in nucleosomes imposes a DNA strand polarity on somatic mutations in skin cancer. *Genome research* 30(1):12-21.
13. Brogaard K, Xi L, Wang JP, & Widom J (2012) A map of nucleosome positions in yeast at base-pair resolution. *Nature* 486(7404):496-501.
14. Crooks GE, Hon G, Chandonia JM, & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome research* 14(6):1188-1190.
15. Laughery MF*, et al.* (2020) Atypical UV Photoproducts Induce Non-canonical Mutation Classes Associated with Driver Mutations in Melanoma. *Cell reports* 33(7):108401.