The major mechanism of melanoma mutations is based on deamination of cytosine in pyrimidine dimers as determined by circle damage sequencing

Seung-Gi Jin¹, Dean Pettinga², Jennifer Johnson¹, Peipei Li³, Gerd P. Pfeifer¹*

Sunlight-associated melanomas carry a unique C-to-T mutation signature. UVB radiation induces cyclobutane pyrimidine dimers (CPDs) as the major form of DNA damage, but the mechanism of how CPDs cause mutations is unclear. To map CPDs at single-base resolution genome wide, we developed the circle damage sequencing (circle-damage-seq) method. In human cells, CPDs form preferentially in a tetranucleotide sequence context (S’-Py-T<>Py-T/A), but this alone does not explain the tumor mutation patterns. To test whether mutations arise at CPDs by cytosine deamination, we specifically mapped UVB-induced cytosine-deaminated CPDs. Transcription start sites (TSSs) were protected from CPDs and deaminated CPDs, but both lesions were enriched immediately upstream of the TSS, suggesting a mutation-promoting role of bound transcription factors. Most importantly, the genomic dinucleotide and trinucleotide sequence specificity of deaminated CPDs matched the prominent mutation signature of melanomas. Our data identify the cytosine-deaminated CPD as the leading premutagenic lesion responsible for mutations in melanomas.

INTRODUCTION

Cancer genome sequencing has identified millions of somatic mutations in different types of human cancer (1–4), in some cases pointing to a potential origin of these mutations. Cancers linked to environmental exposures show tumor-specific mutational signatures (5). For example, smoking-associated lung cancers carry a prevalent C/G to A/T mutation signature, as initially defined by Alexandrov et al. (1, 6). Another example is urothelial carcinomas linked to consumption of plants that contain the mutagen aristolochic acid, which produces characteristic A-to-T transversions (7). The theory that a particular exposure is linked to a cancer is particularly compelling when it is supported by epidemiological evidence and by mechanistic studies that attempt to recapitulate the mutagenesis process in vitro or in vivo.

Most melanomas and nonmelanoma skin cancers are linked to sunlight exposure. The ultraviolet B (UVB) component of sunlight (280 to 315 nm) is strongly carcinogenic in mouse models (8), and of particular relevance is the wavelength above 300 nm, which reaches the earth’s surface. For most sequenced melanoma genomes, C-to-T mutations at dipyrimidine sequences (for example, at TC or CC) are highly prevalent, suggesting that UVB radiation from solar exposure is involved in their formation (9, 10). Individual melanoma genomes may harbor thousands of such mutations. These same mutations are observed in various experimental systems when UV radiation is used as the mutagen (11, 12). However, it is unknown what specific mechanism is responsible for these melanoma mutations.

UVB radiation induces various forms of DNA damage in exposed cells (13). The most frequent damage type is the cis-syn cyclobutane pyrimidine dimer (CPD), which arises after absorption of photons by DNA bases and occurs through the formation of electronic excited states. Any two adjacent pyrimidine bases can dimerize via their 5,6 double bonds. The second most frequent lesion produced by UVB irradiation is the pyrimidine (6-4) photoprotect [6-4 photoprotect]. In comparative mutagenesis experiments, the CPD has been shown to be at least five times more mutagenic than the (6-4) photoprotects (14). One reason for this difference is the generally rapid global repair of (6-4) photoprotects (within minutes or hours) relative to CPDs, which can persist for days in irradiated cells or human skin. In addition to CPDs and (6-4) photoprotects, other DNA photoprotects can be produced in DNA at lower levels and may be relevant for atypical, non–C-to-T mutations in melanoma (15). In addition, (6-4) photoprotects are produced at even lower levels relative to CPDs when the irradiation wavelength is greater than 300 nm (16), while the most physiologically relevant part of the UV spectrum. Experiments with transgenic mice in which either the CPDs or the (6-4) photoprotects were removed by a specific repair enzyme have shown that removal of CPDs provides powerful protection against UVB-induced skin cancers (17).

The high frequencies of C-to-T and CC-to-CT double mutations in melanoma and nonmelanoma skin cancers are therefore best explained by proposing the CPD as a premutagenic lesion. However, it has remained unclear how CPD lesions induce the characteristic C-to-T mutations in skin cancer. There are two major models: The first one proposes that CPDs that contain cytosine are replicated by DNA polymerases that incorporate adenine across from cytosine in an error-prone pathway. This model has been referred to as the “A-rule” for polymerase bypass of non-instructional DNA lesions (18). When the CPD contains thymine, pairing with adenine will not lead to a mutation. The lesion-tolerant DNA polymerase eta (POLH) has been shown to bypass TT CPDs correctly by incorporating two adenines (19, 20). A competing model suggests that cytosines, when part of a CPD, are much more susceptible to hydrolytic deamination, a pathway that generates CPDs containing uracil. When these U-containing dimers are then bypassed by a polymerase with incorporation of adenine, a C-to-T mutation is the outcome, and it occurs because of the deamination reaction and not because of a polymerase error.

¹Department of Epigenetics, Van Andel Institute, Grand Rapids, MI 49503, USA.
²Bioinformatics and Biostatistics Core, Van Andel Institute, Grand Rapids, MI 49503, USA.
³Department of Neurodegenerative Science, Van Andel Institute, Grand Rapids, MI 49503, USA.
*Corresponding author. Email: gerd.pfeifer@vai.org

Jin et al., Sci. Adv. 2021; 7 : eabi6508 30 July 2021
In this study, we set out to investigate the CPD deamination mechanism with the goal of determining whether this specific pathway can be linked to the frequent C-to-T mutations found in melanoma. To perform this work, we first developed a new method for base resolution and whole-genome analysis of DNA damage. We used this new methodology, termed circle damage sequencing (circle-damage-seq), to separately map CPDs and deaminated CPDs in human UVB-irradiated cells at the genome scale and at the single-base level. Our data provide strong support for the CPD deamination model being the mechanistic cause of most melanoma mutations.

RESULTS
Development of the circle-damage-seq method
To determine mutational mechanisms at the whole-genome scale, it is critical to map the underlying DNA damage at single-base resolution and genome wide. However, these methods need to be specific and require high sensitivity. Because of the limitations often associated with currently used methods, we developed a new strategy, which we named circle-damage-seq. This method is based on DNA damage–specific repair enzymes and allows sensitive detection and genome-scale mapping of DNA base damage. The general outline of this method is depicted in Fig. 1. The cells are first exposed to a DNA-damaging agent and then harvested. DNA is purified and sonicated to an average size of 300 to 400 base pairs (bp), a suitable range for DNA circularization and downstream sequencing analysis (see Methods). After sonication and end repair, a DNA circularization step is carried out using a diluted ligation reaction with T4 DNA ligase. T4 ligase also repairs DNA single-strand breaks that would contribute to background signals. Any remaining noncircularized fragments are eliminated by applying an exonuclease mixture (Fig. 1A). Then, the lesion is excised with a lesion-specific DNA repair enzyme and abasic site–processing enzymes within the circularized DNA to introduce a 1-nucleotide (nt) gap. Next, a double-strand break is created at the DNA damage position by removing the single nucleotide on the opposite strand of the gapped DNA molecules with single-strand specific S1 nuclease, thus producing a double-strand break (Fig. 1B). The use of multiple enzymes has the potential

![Fig. 1. Outline of circle-damage-seq.](image-url)
to increase background of the procedure. We carefully optimized each enzymatic step including enzyme concentrations, incubation temperatures, and incubation times to minimize nonspecific DNA cleavage. Circle-damage-seq uniquely enables sequencing of both sides of the single cleavage site in one DNA molecule using paired-end sequencing. Closed circular DNA without any recognized damaged bases is excluded from the breakage and ligation steps, resulting in lesion-selective DNA library enrichment by polymerase chain reaction (PCR) (Fig. 1C). Notably, sequencing reads obtained from circle-damage-seq show a specific aligned read pattern, divergent reads with a single-base gap in the center representing the damaged DNA base, because each paired-end read is derived from a single opened circular DNA molecule (Fig. 1C). Thus, circle-damage-seq provides high selectivity for DNA damage mapping at single-nucleotide resolution with low background at reduced sequencing depth. This method should be applicable for mapping of any type of DNA damage for which a DNA excision repair or cleavage enzyme is available.

We first tested the feasibility of circle-damage-seq by analyzing the endogenous DNA modification N6-methyladenine (N6mA) in Escherichia coli dam+ cells (fig. S1). DNA double-strand breakage at N6mA bases within circularized DNA was achieved through Dpn I cleavage, which generates a blunt end by splitting the target DNA between 5′GN6mA and TC sequences. N6mA could readily be mapped in E. coli (fig. S1).

**Genomic mapping of UVB-induced CPDs**

To map UVB-induced CPDs (Fig. 1D), the major mutagenic UVB-induced lesion at dipyrimidine sequences (14), we irradiated human skin fibroblasts with UVB and immediately isolated the DNA. We cleaved DNA at CPDs with T4 endonuclease V (also known as T4-PDG) (21). This enzyme can also cleave abasic sites; therefore, it is important to minimize depurination during DNA isolation steps. CPD induction was confirmed by gel electrophoresis after cleavage of DNA with T4-PDG and S1 nuclease. As shown in Fig. 1E, to create double-strand breaks at the CPD positions after T4-PDG incision, we first used AP endonuclease 1 (APE1) to cleave the 5′ side of the sugar-phosphate backbone at pyrimidine dimers. We followed this treatment by incubation with E. coli photolyase under long-wave UVA light to remove dimerized pyrimidine bases that remain after the DNA glycosylase incision. Next, S1 nuclease treatment generated ligatable DNA double-strand breaks within the circularized DNA. The combined treatment creates a single-nucleotide gap representing the 5′ pyrimidine of the dimer. We aligned the sequencing reads from circle-damage-seq to the human reference genome. We observed the divergent paired reads with a single-base gap in the center and confirmed the matching bases in the gaps as pyrimidines that have another pyrimidine base in the 3′ position (Fig. 1F). We found this pattern with a high frequency at 5′TT and 5′TC sites and at lower frequency at 5′CT and 5′CC dinucleotide sequences, consistent with the known specificity of CPD formation at dipyrimidine sequences (10). Additional examples of genomic CPD mapping are shown in fig. S2.

**Sequence specificity of CPD formation**

Previous work has focused mainly on the dinucleotide specificity of genomic UV damage (10, 12, 13, 22–24). We now analyzed trinucleotide sequences that reflect the sequence specificity of CPD formation in human fibroblasts using ~100 million aligned, divergent read pairs with single-nucleotide gaps (Fig. 2). When normalized for the known trinucleotide content of the human genome, the most frequent sequence contexts considering the base flanking the

---

**Fig. 2. Sequence context of CPD formation in human fibroblasts irradiated with UVB.** (A) Distribution plot of trinucleotide sequences (PyPyN3′) undergoing CPD formation in UVB-irradiated human cells (green). NT, nontreated control (red). Some background is seen in the NT samples at non-dipyrimidine sequences (the four trinucleotides to the right). (B) Distribution plot of trinucleotide sequences (5′NPyPy) undergoing CPD formation in UVB-irradiated human cells (green). Data in (A) and (B) are normalized for the trinucleotide frequencies of the human genome.
dimerized pyrimidines at the 3′ side were 5′TTA and 5′TCT (CPD underlined), followed by 5′TCA (Fig. 2A). Thus, T is preferred in the first position of the dimer, and T and A are the preferred bases 3′ to the most abundant CPDs (5′TT and 5′TC). The most highly enriched trinucleotide sequences considering the 5′ neighboring base of the CPDs were 5′CTT and 5′CTC, followed by 5′TTC and 5′TTT (Fig. 2B) showing a preference for having another pyrimidine 5′ to a pyrimidine dimer. Divergent reads with single-base gaps were >300 times less frequent in nontreated (NT) cells than in UVB-irradiated cells with no particular enrichment of sequencing reads associated with such trinucleotides in the control samples [Fig. 2, A and B (bottom)].

We also analyzed the preferred tetranucleotide sequence contexts for CPD formation (fig. S3), which were generally consistent with the trinucleotide patterns. The sequence enrichments for the 5′NPyPyN tetranucleotide context (fig. S3C) were dominated by a T in the second position and by having another pyrimidine as the first base. This pattern is similar to one obtained after in vitro irradiation of DNA with UVB (25).

To identify even broader possible sequence motifs, we obtained a base enrichment on 20-nt regions around the gapped base pair using ~100 million aligned, divergent read pairs (Fig. 3A). Position 11 represents the nucleotide in the gap between divergent reads (the 5′ pyrimidine of the CPD). We observed a strong enrichment of 5′TT or 5′TC contexts at positions 11 and 12, representing the pyrimidine dimers. In this motif, T or C was enriched at position 10, and T and A bases were enriched at position 13, consistent with the tri- and tetranucleotide patterns. Unexpectedly, no further sequence enrichment was observed outside of this tetranucleotide context (positions 10 to 13). Our findings define a distinct tetranucleotide sequence preference, 5′-Py-T<>Py-T/A, for the major type of UVB damage in the human genome. This sequence is also found at sites hypersensitive to UV radiation (24).

**Gene level distribution of CPDs**

Next, using ~36 million aligned, divergent read pairs, we examined the global patterns of CPDs, with emphasis on their distribution along genes. CPD formation in cells is chiefly uniform along the genome and is largely dependent on DNA sequence, where 5′TY sequences show the strongest accumulation of CPDs. In the metagene profiles, we observed high levels of CPD formation immediately upstream of transcription start sites (TSSs) but a dip of CPDs right around the GC-rich TSS sequences themselves at many gene promoter regions (Fig. 3, B to D; see Fig. 3F and fig. S2 for specific gene examples). This finding suggests that transcription factors at promoter upstream regions may enhance CPD formation at many sites, consistent with previous studies (22, 24, 26–30), even in the absence of modulation by DNA repair. Melanoma mutations are also sharply enriched just upstream of the TSS (31).

The metagene profile of CPD formation reveals a notable spike of CPDs near the transcription end sites (TESs) of genes (Fig. 3E). Human and mouse genes have an AT-rich sequence context near the TES that contains polyadenylation signals (AATAAA) and their surrounding sequences (32), which may lead to an enrichment of CPDs containing thymine. However, a GC-rich sequence context is only inversely correlated with CPD frequency right at the TSS. Just upstream of the TSS, high GC content and CPD levels are correlated, directly pointing to a role of DNA-bound proteins in CPD induction rather than the DNA sequence itself.

**Genome-wide mapping of cytosine-deaminated CPDs**

CPDs are known to be the most prevalent and most mutagenic UVB-induced DNA lesions (14, 10). CPDs containing cytosine are unstable. Owing to saturation of the 5,6 double bond of cytosine upon dimer formation (Fig. 1D), they are prone to undergo hydrolytic cytosine deamination to form uracil, which could be a potentially mutagenic pathway (33–35). We next applied the circle-damage-seq method to analyze the extent of CPD cytosine deamination. We irradiated human fibroblast cells with UVB and harvested them 24 and 48 hours later to allow time for deamination. To specifically map the deaminated CPDs, we applied a photolyase-mediated reversal of the CPDs first, followed by the excision of U bases by uracil DNA glycosylase (UDG) in the circle-damage-seq method (Fig. 4A). This method successfully displays CPDs containing deaminated cytosines at single-base resolution and shows cytosine in the single-base gap, as predicted (Fig. 4B and fig. S4A). With this method, deamination of CPDs that contain 5-methylcytosine cannot be detected because the deaminated base is thymine. We note that we also observed double-base gaps at CC bases, indicating likely double deamination events, but these were difficult to quantify because of the unknown efficiency of UDG incision at UC sequences. Therefore, we focused on the single C gaps.

The patterns of deaminated CPDs were largely independent of total CPD patterns (fig. S4, B and C) because of their respective differential sequence requirements. At a genomic scale, deaminated CPDs were strongly depleted at many TSS regions and even more so than total CPDs (Figs. 4, C and D, and 5, A and B). Genes with CpG island (CGI) promoters showed a dip of total CPDs near the TSS, but this dip was missing in genes with non-CGI promoters (Fig. 5, A and B). However, a lower level of deaminated CPDs was still visible around the TSS, even in non-CGI promoters (Fig. 5B). Overall, levels of deaminated CPDs increased with time (Fig. 4C). Since deamination of CPDs is strictly a chemical reaction, substantial levels of deaminated CPDs were found at the 0-hour time point because the deamination reaction is expected to continue during DNA processing in vitro. The reasons for the lower levels of deaminated CPDs at the TSS could be the reduced propensity of TSS sequences to be transiently single stranded (which favors cytosine deamination), the shielding of these regions by general transcription factors, and/or the particularly rapid repair of these regions (28, 36–38), which may proceed to some extent during the 48-hour deamination time span even at the relatively high UV dose that we used. At the TES, there is a spike of total CPDs but a dip of deaminated CPDs, which is likely based on DNA sequence features, i.e., AT richness of these regions (Fig. 5C). The highest level of CPD formation and deaminated CPD formation was found at the regions immediately upstream of the TSS of CGI promoters (Fig. 5A). This finding suggests that bound protein factors can sensitize DNA to UV light and promote mutagenesis (26–31, 39).

**Sequence specificity of cytosine-deaminated CPDs and relationship to melanoma mutations**

Although TT and TC are the sequences most susceptible to CPD formation (Fig. 2, A and B, and fig. S3), the thymine positions of CPDs are not very mutagenic because of incorporation of adenine opposite to thymine by POLH (19) or other DNA polymerases. However, the mechanism of how CPDs containing cytosines cause C-to-T mutations has remained unknown. Existing models include error-prone DNA polymerases that incorporate adenine across
Fig. 3. CPD mapping by circle-damage-seq shows the distribution of CPDs at the sequence and gene level. (A) Nucleotide composition of mapped CPD positions. Position 11 and 12 represent UVB-damaged dipyrimidine sequences undergoing CPD formation. At each base position, the height of each letter represents the relative frequency of that nucleic acid base. (B) G+C sequence enrichment along human genes (hg19). (C) Heatmap and metagene profiles of CPD distribution along all genes of the hg19 human genome. CPD coverage is sorted from high (top, blue) to low (bottom, red). The CPD signals were mapped and binned in 50-bp windows from 1.5 kb upstream of the transcription start site (TSS) and then normalized relative to gene length over the gene bodies to the transcription end site (TES) and 1.5 kb downstream of the TES. (D) Heatmap and profile of CPD distribution around the TSS and 1.5 kb of flanking sequence of all hg19 genes. (E) Heatmap and profile of CPD distribution around the TES and 1.5 kb of flanking sequence of all hg19 genes. (F) Example of CPD sequence read distribution along several genes on chr2. There is a reduced frequency of CPDs near the TSS (blue arrows).
Fig. 4. Mapping of deaminated CPDs at the sequence and gene level. (A) Outline of the method used for mapping of deaminated CPDs. The red “=” symbol indicates a CPD. (B) Divergent paired reads of deaminated CPDs obtained by circle-damage-seq. Purple and lavender segments represent reads mapped to the plus and minus strands, respectively. Green arrows indicate single-base gaps at cytosine bases matching the deaminated base of a pyrimidine dimer. The trinucleotide context of the deaminated cytosine is indicated. (C) Heatmaps and metagene profiles of deaminated CPDs in UVB-irradiated human fibroblasts. Total signal is sorted from high to low (top to bottom). Metagene profiles are shown for untreated cells (NT) and at 0, 24, and 48 hours following UVB irradiation. (D) Browser views of total CPDs (top) and cytosine-deaminated CPDs (bottom) along the DKK3 gene on chromosome 11 in human fibroblasts. Note the reduced levels of deaminated CPDs at the TSS containing a CGI (red bar).
C-containing CPDs (the A-rule). A competing model proposes that cytosines within CPDs first undergo deamination to form uracil, before “error-free” polymerase bypass, e.g., by POLH, incorporating adenines across the deaminated, uracil-containing CPDs (Fig. 6A).

To evaluate the deamination model, we next determined the trinucleotide sequence preference of deaminated CPDs using ~100 million aligned reads and observed a prominent enrichment of TCC, TCA, TCT, CCT, and CCC sequences (Fig. 6, B to D). The trinucleotide data also suggest that cytosine needs to be in the second position of a CPD to score as a prominent deaminated CPD (Fig. 6, C and D; see for example the much higher levels of deaminated CPDs at trinucleotides beginning with 5’TC versus those beginning with 5’CT). A broader
Fig. 6. Sequence specificity of deaminated CPDs in the UVB-irradiated human genome and melanoma mutations. (A) Outline of CPD deamination and potential mutagenic consequences. (B) Sequence context of the occurrence of deaminated CPDs at 48 hours following UVB irradiation. Position 11 is the deaminated cytosine. (C) Trinucleotide sequence specificity of total CPD formation, formation of deaminated CPDs, and the prominent mutational signature 7 (SBS7) in melanoma. For total CPDs, we show the sequences in the context of 5’-dipyrimidine (NPyPy; blue line plot, ranked according to levels; CPDs at NPyPy are underlined). For deaminated CPDs and mutations, the NPyN (n = 32) trinucleotide context is shown (green and red bars, respectively). (D) Trinucleotide sequence specificity of total CPD formation, formation of deaminated CPDs, and the mutation signature SBS7 in melanoma genomes. We show the sequences in the context 5’-dipyrimidine-N (PyPyN) (n = 16). Data in (B) to (D) are not normalized for total genomic trinucleotide frequencies (see also fig. S6). (E) Heatmap showing the cosine similarity scores for total CPDs (NPyPy, n = 32), deaminated CPDs, and the prominent mutational signature 7 (SBS7/SBS7a/SBS7b) in melanoma along with the 30 COSMIC v2 mutational signatures. We only used the C-to-T and T-to-C mutation windows for comparisons with the COSMIC trinucleotide signatures. Dark blue colors indicate high similarity (see also fig. S7).
Our results are based on the development of a new method for mapping of DNA damage at base-level resolution in the human genome. Other methods are available to accomplish DNA damage mapping by next-generation sequencing (22, 24, 37, 43). The advantages of circle-damage-seq over comparable DNA damage mapping techniques are several-fold. (i) The single incision point of the modified base and the divergent reads emanating from this break position give a clear indication of where the damage is located. (ii) Background signals are minimized by removal of DNA molecules that did not undergo circularization. (iii) Most single-strand breaks that also contribute to background are removed during circular ligation. (iv) The method represents DNA damage–specific sequencing because DNA circles or linear molecules without damaged bases are not sequenced, thus contributing to lower background and lower sequencing costs.

The circle-damage-seq method depends on the ability to convert the modified base into a DNA single-strand break. Cleavage with S1 nuclease on the opposite DNA strand then creates a damage-specific ligatable double-strand break. DNA glycosylases, which operate at the initial step of the base excision repair pathway, can easily be adapted for this technique. By choosing the appropriate DNA glycosylase, for example, 8-oxoguanine DNA glycosylase (OGG1) for 8-oxoguanine or alkyladenine DNA glycosylase (AAG) for alkylated adenes, specific types of base damage can be mapped. The nucleotide excision repair (NER) complex could be used for more bulky DNA lesions (44). The longer single-stranded gap remaining after excision with NER enzymes should be a good substrate for S1 nuclease, leading to an 11-nt gap for each divergent read pair after circle-damage-seq, for example, when benzo[a]pyrene DNA adducts are excised with the UvrABC complex from DNA (45, 46). It should also be feasible to map rare endogenous DNA bases with respective cleavage enzymes, for example, 5-formylcytosine and 5-carboxycytosine with thymine DNA glycosylase or 6-methyladenine with Dpn I at GATC sequences (fig. S1).

Using circle-damage-seq, we derived a genome-wide tetranucleotide consensus sequence for CPD formation, (5’PyPy<>PyT/A) with unexpectedly no further sequence enrichment outside of this context. However, this consensus sequence pattern alone did not match with the sequences commonly mutated in melanomas. We therefore focused on the analysis of a mutagenesis pathway that proposes deamination of cytosines within CPDs as a major premutagenic mechanism. We were able to specifically detect and map cytosine-deaminated CPDs by developing a variation of circle-damage-seq that uses photolyase reversal of the CPD and uracil excision as the initial steps. Although uracil-containing CPDs are known to exist in vitro and in vivo after UV irradiation (33–35), their relevance for UV mutagenesis has remained unclear. In vitro, synthetic CPDs containing uracil can be bypassed with incorporation of adenine by the lesion-tolerant DNA polymerase eta (47). If a similar bypass reaction occurs in vivo, perhaps catalyzed also by other DNA polymerases, the mutation occurs by deamination and not by a polymerase error. Note, in this context, that both the yeast and human forms of the Pol Eta enzyme insert G opposite 5-methylcytosine in a CPD with greater than 100:1 selectivity in vitro, favoring the idea that the C to T deamination leads to cancer mutations.

DISCUSSION

Our results are based on the development of a new method for mapping of DNA damage at base-level resolution in the human genome. Other methods are available to accomplish DNA damage mapping by next-generation sequencing (22, 24, 37, 43). The advantages of circle-damage-seq over comparable DNA damage mapping techniques are several-fold. (i) The single incision point of the modified base and the divergent reads emanating from this break position give a clear indication of where the damage is located. (ii) Background signals are minimized by removal of DNA molecules that did not undergo circularization. (iii) Most single-strand breaks that also contribute to background are removed during circular ligation. (iv) The method represents DNA damage–specific sequencing because DNA circles or linear molecules without damaged bases are not sequenced, thus contributing to lower background and lower sequencing costs.

The circle-damage-seq method depends on the ability to convert the modified base into a DNA single-strand break. Cleavage with S1 nuclease on the opposite DNA strand then creates a damage-specific ligatable double-strand break. DNA glycosylases, which operate at the initial step of the base excision repair pathway, can easily be adapted for this technique. By choosing the appropriate DNA glycosylase, for example, 8-oxoguanine DNA glycosylase (OGG1) for 8-oxoguanine or alkyladenine DNA glycosylase (AAG) for alkylated adenes, specific types of base damage can be mapped. The nucleotide excision repair (NER) complex could be used for more bulky DNA lesions (44). The longer single-stranded gap remaining after excision with NER enzymes should be a good substrate for S1 nuclease, leading to an 11-nt gap for each divergent read pair after circle-damage-seq, for example, when benzo[a]pyrene DNA adducts are excised with the UvrABC complex from DNA (45, 46). It should also be feasible to map rare endogenous DNA bases with respective cleavage enzymes, for example, 5-formylcytosine and 5-carboxycytosine with thymine DNA glycosylase or 6-methyladenine with Dpn I at GATC sequences (fig. S1).

Using circle-damage-seq, we derived a genome-wide tetranucleotide consensus sequence for CPD formation, (5’PyPy<>PyT/A) with unexpectedly no further sequence enrichment outside of this context. However, this consensus sequence pattern alone did not match with the sequences commonly mutated in melanomas. We therefore focused on the analysis of a mutagenesis pathway that proposes deamination of cytosines within CPDs as a major premutagenic mechanism. We were able to specifically detect and map cytosine-deaminated CPDs by developing a variation of circle-damage-seq that uses photolyase reversal of the CPD and uracil excision as the initial steps. Although uracil-containing CPDs are known to exist in vitro and in vivo after UV irradiation (33–35), their relevance for UV mutagenesis has remained unclear. In vitro, synthetic CPDs containing uracil can be bypassed with incorporation of adenine by the lesion-tolerant DNA polymerase eta (47). If a similar bypass reaction occurs in vivo, perhaps catalyzed also by other DNA polymerases, the mutation occurs by deamination and not by a polymerase error. Note, in this context, that both the yeast and human forms of the Pol Eta enzyme insert G opposite 5-methylcytosine in a CPD with greater than 100:1 selectivity in vitro, favoring the idea that the C to T deamination leads to cancer mutations.

DISCUSSION

Our results are based on the development of a new method for mapping of DNA damage at base-level resolution in the human genome. Other methods are available to accomplish DNA damage mapping by next-generation sequencing (22, 24, 37, 43). The advantages of circle-damage-seq over comparable DNA damage mapping techniques are several-fold. (i) The single incision point of the modified base and the divergent reads emanating from this break position give a clear indication of where the damage is located. (ii) Background signals are minimized by removal of DNA molecules that did not undergo circularization. (iii) Most single-strand breaks that also contribute to background are removed during circular ligation. (iv) The method represents DNA damage–specific sequencing because DNA circles or linear molecules without damaged bases are not sequenced, thus contributing to lower background and lower sequencing costs.

The circle-damage-seq method depends on the ability to convert the modified base into a DNA single-strand break. Cleavage with S1 nuclease on the opposite DNA strand then creates a damage-specific ligatable double-strand break. DNA glycosylases, which operate at the initial step of the base excision repair pathway, can easily be adapted for this technique. By choosing the appropriate DNA glycosylase, for example, 8-oxoguanine DNA glycosylase (OGG1) for 8-oxoguanine or alkyladenine DNA glycosylase (AAG) for alkylated adenes, specific types of base damage can be mapped. The nucleotide excision repair (NER) complex could be used for more bulky DNA lesions (44). The longer single-stranded gap remaining after excision with NER enzymes should be a good substrate for S1 nuclease, leading to an 11-nt gap for each divergent read pair after circle-damage-seq, for example, when benzo[a]pyrene DNA adducts are excised with the UvrABC complex from DNA (45, 46). It should also be feasible to map rare endogenous DNA bases with respective cleavage enzymes, for example, 5-formylcytosine and 5-carboxycytosine with thymine DNA glycosylase or 6-methyladenine with Dpn I at GATC sequences (fig. S1).

Using circle-damage-seq, we derived a genome-wide tetranucleotide consensus sequence for CPD formation, (5’PyPy<>PyT/A) with unexpectedly no further sequence enrichment outside of this context. However, this consensus sequence pattern alone did not match with the sequences commonly mutated in melanomas. We therefore focused on the analysis of a mutagenesis pathway that proposes deamination of cytosines within CPDs as a major premutagenic mechanism. We were able to specifically detect and map cytosine-deaminated CPDs by developing a variation of circle-damage-seq that uses photolyase reversal of the CPD and uracil excision as the initial steps. Although uracil-containing CPDs are known to exist in vitro and in vivo after UV irradiation (33–35), their relevance for UV mutagenesis has remained unclear. In vitro, synthetic CPDs containing uracil can be bypassed with incorporation of adenine by the lesion-tolerant DNA polymerase eta (47). If a similar bypass reaction occurs in vivo, perhaps catalyzed also by other DNA polymerases, the mutation occurs by deamination and not by a polymerase error. Note, in this context, that both the yeast and human forms of the Pol Eta enzyme insert G opposite 5-methylcytosine in a CPD with greater than 100:1 selectivity in vitro, favoring the idea that the C to T deamination leads to cancer mutations.
occur before DNA replication. Since cytosine deamination in CPDs is strictly a chemical reaction, it will be difficult to prevent it from occurring, although CPD removal by skin-penetrating CPD-specific DNA repair enzymes appears to be a reasonable concept, as proposed earlier (49), and these enzymes work on uracil-containing CPDs. In summary, our data support a specific mutational mechanism of how a prevalent form of sunlight-induced DNA damage induces specific types of mutations found at high frequency in human skin cancer genomes.

METHODS

Cell culture
Human skin fibroblasts cells were obtained from the American Type Culture Collection (ATCC; catalog no. PCS-201-012) and grown using the Fibroblast Growth Kit (ATCC, catalog no. PCS-201-041) supplemented with 2% fetal bovine serum. The cells were used at passage number <5.

UVB irradiation and genomic DNA isolation
For mapping of UVB-induced CPDs, fibroblasts at 80 to 90% confluence on 10-cm culture plates were washed with 1× phosphate-buffered saline (PBS) and were irradiated in PBS with a UVB dose of 1000 J/m². The UVB lamp (Thermo Fisher Scientific, UVP 3UV lamp) had a peak spectral emission at 302 nm. The UVB dose was determined using a UVX radiometer with a UVB probe (Ultraviolet Products; Upland, CA). After irradiation, the cells were immediately trypsinized and collected, and then, genomic DNA was isolated using Quick-DNA Miniprep Plus Kit (Zymo Research; Irvine, CA) according to the manufacturer’s instruction manual. For mapping deaminated CPDs, we replenished the culture medium of the cells after UVB radiation with 4000 J/m² and incubated the cells for 24 or 48 hours at 37°C before harvesting and DNA isolation.

Circle-damage-seq mapping of CPDs

DNA preparation and circularization
To prepare fragmented genomic DNA, the DNA from UVB-irradiated cells was sheared at 4°C to an average length of 300 to 400 bp by sonication with a Covaris E220 sonicator (Covaris; Woburn, MA) under the following conditions: peak incident power (W), 140; duty factor, 10%; cycles per burst, 200 times, 80 s; then, the fragmented DNA was purified with DNA Clean & Concentrator-5 Kit (Zymo Research, DCC-5) according to the manufacturer’s protocol. DNA was eluted in 42 μl of 10 mM tris-HCl (pH 7.5). Next, the sonicated DNA was end-repaired to prepare blunt-ended DNA. One microgram of the sonicated DNA was incubated in 1× T4 DNA ligase buffer [New England Biolabs (NEB); Ipswich, MA] with the following components: 50 mM tris-HCl (pH 7.5), 10 mM MgCl₂, 1 mM adenosine triphosphate (ATP), 10 mM dithiothreitol (DTT), 1 μl (2 U/μl) of RNase H (NEB), 1 μl (3 U/μl) of T4 DNA polymerase (NEB), 1 μl (10 U/μl) of T4 polynucleotide kinase (NEB), and 1 μl of 10 mM deoxyadenosine triphosphates (dNTPs) in a final volume of 50 μl at 24°C for 30 min; then, the reaction was heat-inactivated by further incubation at 70°C for 20 min. The DNA was purified using DCC-5 Kit (Zymo Research) and eluted in 100 μl of 10 mM tris-HCl (pH 7.5). The purified DNA was quantitated using NanoDrop (Thermo Fisher Scientific).

To circularize the DNA, 1 μg of the blunt-ended DNA was incubated in 1× T4 DNA ligase buffer (NEB) with 4 μl (400 U/μl) of T4 DNA ligase (NEB) in a final volume of 200 μl at 16°C overnight. Then, the reaction mixture was cleaned up with DCC-5 kit, and the DNA was eluted in 42 μl of 10 mM tris-HCl (pH 7.5). To remove noncircularized linear DNA from the circularized DNA pool, the ligase-treated DNA was incubated with 2 μl (10 U/μl) of Plasmid-Safe ATP-Dependent DNAse (Lucigen; Middleton, WI) in 1× reaction buffer [33 mM tris-acetate (pH 7.5), 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT, and 1 mM ATP] in a final volume of 50 μl at 37°C for 30 min and then at 70°C for 20 min to stop the reaction. Then, the DNA was purified with 90 μl (1.8×) of AMPure XP beads (Beckman Coulter; Indianapolis, IN) and eluted in 35 μl of 10 mM tris-HCl (pH 8.0).

Cleavage of DNA at CPD sites
To generate DNA double-strand breaks at CPD sites, the circularized DNA was treated with the following DNA repair enzymes. First, to specifically incise DNA at CPD positions, the circularized DNA was incubated in 1× NEBuffer 4 reaction buffer (NEB) [50 mM potassium acetate, 20 mM tris-acetate (pH 7.9), 10 mM magnesium acetate, and 1 mM DTT] and 0.4 μl of bovine serum albumin (20 mg/ml) with 1 μl (10 U/μl) of T4-PDG (NEB) and 1 μl (10 U/μl) of APE1 (NEB) in a final volume of 40 μl at 37°C for 20 min. Then, the DNA was cleaned up with 72 μl (1.8×) of AMPure XP beads and eluted in 25 μl of 10 mM tris-HCl (pH 8.0). To revert the dimerized pyrimidines remaining after T4-PDG incision, the DNA from above was treated with 3 μl (0.25 μg/μl) of E. coli phrB photolyase (Novus Biologicals; Centennial, CO) in 1× reaction buffer [50 mM tris-HCl (pH 7.0), 50 mM NaCl, and 10 mM DTT] in a final volume of 50 μl. The reaction tube was placed at a distance of 15 cm under a UVA lamp (Thermo Fisher Scientific) that has a peak spectral emission at 365 nm and incubated under UVA illumination for 90 min at room temperature. Then, the DNA was cleaned up with 90 μl (1.8×) of AMPure XP beads and eluted in 32 μl of 10 mM tris-HCl (pH 8.0). To cleave the opposite strand of the nicked DNA, the DNA from above was incubated with 1 μl (5 U/μl) of single-strand specific S1 nuclease (Thermo Fisher Scientific) in 1× reaction buffer [40 mM sodium acetate (pH 4.5), 300 mM NaCl, and 2 mM ZnSO₄] in a final volume of 40 μl for 4 min at room temperature. We stopped the reaction by adding 2 μl of 0.5 M EDTA and 1 μl of 1 M tris-HCl (pH 8.0) to the reaction mixture and further incubated the mixture for 10 min at 70°C. The DNA sample was cleaned up with 72 μl (1.8×) of AMPure XP beads and eluted in 48 μl of 10 mM tris-HCl (pH 8.0).

Cleavage of DNA at deaminated CPD sites
To generate double-strand breaks at deaminated CPD sites, the circularized DNA was initially treated with 3 μl (0.25 μg/μl) of E. coli phrB photolyase, as described above to revert the dimerized pyrimidines. Then, the DNA was cleaned up with 90 μl (1.8×) of AMPure XP beads and eluted in 35 μl of 10 mM tris-HCl (pH 8.0). To specifically incise DNA at deaminated CPD positions, the circularized DNA was incubated in 1× NEBuffer 4 reaction buffer (NEB) with 1 μl (5 U/μl) of UDG (NEB) and 1 μl (10 U/μl) of APE1 (NEB) in a final volume of 40 μl at 37°C for 20 min. Then, the DNA was cleaned up with 72 μl (1.8×) of AMPure XP beads and eluted in 32 μl of 10 mM tris-HCl (pH 8.0). To cleave the opposite strand of the nicked DNA, the DNA was then incubated with 1 μl (5 U/μl) of single-strand specific S1 nuclease and cleaned up as described above.

Adapter ligation
The double-strand cleaved DNA was subsequently A-tailed in 1× NEBNext dA-tailing reaction buffer (NEB) [10 mM tris-HCl (pH 7.9),

Jin et al., Sci. Adv. 2021; 7 : eabi6508  30 July 2021
10 mM MgCl₂, 50 mM NaCl, and 1 mM DTT) and 0.2 mM dATP, with 2 µl (5 U/µl) of Klenow fragment exo- (NEB) in a final volume of 55 µl by incubation at 37°C for 30 min, and then, the reaction was heat-inactivated by further incubation at 70°C for 30 min. To perform adaptor ligation, 30 µl of NEBNext Ultra II Ligation Master Mix, 1 µl of NEBNext Ligation Enhancer (NEB), and 3 µl of 1.5 µM NEBNext adaptors for Illumina sequencing (NEB) [5′-phos-GATC-GGAAGAGACACAGTCTGAACCTCAGTC/ideoxyU/ACACTCTTCCTACACGAGCTTCTCCGATC*’-3’ and 5′-phos-GATCGAGGAGCAGCAGTCTGAACCTCAGTCTCAGTCT/ideoxyU/ACACTCTTCCTACACGAGCTTCTCCGATC*C-3’ (*; phosphorothioate linkage)] were added to the dA-tailing reaction mixture in a total volume of 93 µl and incubated at 20°C for 15 min, and this reaction step was followed by incubation with 3 µl (1 U/µl) of USER enzyme (NEB) at 37°C for 20 min. Then, the adaptor-igated circle-damage-seq library was cleaned up with 96 µl (1×) of AMPure XP beads and size-selected (300 to 1000 bp) with 0.5 volumes of AMPure XP beads, and the DNA was eluted in 25 µl of 10 mM tris–HCl (pH 8.0).

**Library amplification and sequencing**

To amplify the circle-damage-seq library by PCR, the eluted library was mixed with 25 µl of NEBNext Ultra II Q5 Master Mix (2×), 2.5 µl of 10 µM NEBNext i5, and 2.5 µl of 10 µM NEBNext i7 primers (NEB) in a final volume of 50 µl. The amplification reaction mixture was incubated at 98°C for 30 s, and then, 12 cycles of PCR at 98°C for 10 s and 65°C for 75 s were performed, followed by a final extension step at 65°C for 5 min. The PCR product was purified with 50 µl (1×) of AMPure XP beads and eluted in 20 µl of 10 mM tris–HCl (pH 8.0). The purified library was quantitated using a Qubit fluorometer and dsDNA HS Assay kit (Thermo Fisher Scientific). The size distribution of the circle-damage-seq library was determined using Bioanalyzer (Agilent; Santa Clara, CA), and the circle-damage-seq library was quantified using KAPA Library Quantification kit (Kapa Biosystems) according to the manufacturer’s instruction. The circle-damage-seq library was sequenced as 150-bp paired-end sequencing runs on an Illumina HiSeq 2500 platform to obtain between 10 million and 800 million reads. From the 800 million read samples, approximately 200 million paired aligned reads were retained specifically as divergent read pairs with single-base gaps after removal of duplicates.

**Circle-damage-seq mapping of Dpn I cuts in the E. coli genome**

As a pilot experiment, we initially performed the mapping of Dpn I cut sites to optimize the circle-damage-seq method. To test the sensitivity of circle-damage-seq toward different percentages of Dpn I–cleaved sites, we mixed E. coli DNAs from two different strains, dam(+) and dam(−) to prepare 0, 2, and 10% of dam(+) DNA in a background of dam(−) input DNA. We subjected this mixture to the circle-damage-seq procedure, as described above. Briefly, to prepare fragmented DNA samples, Rsa I and Alu I restriction enzymes (NEB) that leave blunt ends were used to digest 1 µg of E. coli genomic DNA. The cleaved DNAs were subjected to A-tailing using a NEBNext Ultra II End Prep kit (NEB) and followed by ligation of an annealed hairpin adaptor (5′-phos-CGGTGGAGGATCCATGAC/ideoxyU/ATCGGTCCACCG/T-3’) using NEBNext Ultra II Ligation Master Mix. The adaptor-ligated DNAs were treated with USER enzyme and circularized with T4 DNA ligase, followed by treatment with Plasmid-Safe ATP-Dependent DNase. To generate double-strand breaks at Dpn I cleavage sites, the circularized E. coli DNAs were incubated in 1× CutSmart buffer (NEB) with 1 µl (20 U/µl) of Dpn I restriction enzyme (NEB) in a final volume of 20 µl at 37°C for 1 hour. Then, circle-damage-seq library preparation and amplification by PCR were performed, as described above. The circle-damage-seq library was sequenced as 75-bp paired-end reads on an Illumina NextSeq500 platform to obtain ~15 million reads.

**Bioinformatics analysis**

All libraries were hard-clipped from the 3’ end to 2 × 80 nucleotide length using Trim_Galore (www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Reads were then adapter-trimmed and quality-trimmed (phred < 20) in a single pass using Trim_Galore default parameters. Trimmed reads were aligned to the hg19 reference genome using BWA-MEM version 0.7.17 (50) with default parameters. Duplicate alignments were marked and removed with the “removeDups” SAMBLASTER option (51). SAMtools version 1.11 (52) was then used to remove reads with low mapping quality (MAPQ < 20) and to sort reads by genomic coordinates. Divergently aligned read pairs (pairs facing away from each other) with a single-nucleotide gap between reads were selected by selecting SAM records with TLEN of 3. TLEN equals the distance between the mapped end of the template and the mapped start of the template. Records with TLEN = 3 represent divergently aligned reads with one nucleotide between mates. Sequences for the three nucleotides representing the last nucleotide of each read and the gapped base were identified using BEDtools (53). To visualize the reads, bam files were produced by selecting SAM records with TLEN of 3 and visualized with the Integrative Genomics Viewer version 2.8.9. The tracks were displayed with a “view as pairs” option to show pairs together with a line joining the ends.

Single-base substitution mutational signatures associated with skin cancer (SBS7/SBS5a/SBS5b) were obtained from the COSMIC database (http://cancer.sanger.ac.uk/cosmic/signatures/SBS/). Normalized forms of trinucleotide sequence signatures were presented by dividing the numbers by the relative abundance of the hg19 genome trinucleotide frequencies.

When treating cells with UVB to produce CPDs, if the gapped nucleotide was T or C in the (+) strand, the damage would have occurred on the (+) strand, whereas a gapped A or G in the (+) strand would indicate that the damage must have occurred on the (−) strand. For deaminated CPDs, the gapped nucleotide would be a cytosine on the (+) strand that is flanked by a pyrimidine on either side. When there is a G as the gapped nucleotide, the damage would have occurred on the (−) strand and another purine is seen as a flanking base. The Dpn I treatment protocol probes 6mA formed at 5’GATC sequences from dam(+) E. coli, producing sequencing libraries with no gap between divergently aligned read pairs (fig. S1). Hence, alignment records with TLEN = 2 were selected.

Chromosome name, start position, stop position, and inferred strand (+/−) for all selected read pair loci were written to a BED file, sorted, and indexed with tabix (54) for downstream analyses. Logo plots were drawn using gsqseqlogo (55) or Seq2Logo (www.cbs.dtu.dk/biotools/Seq2Logo/) for fig. S6. “bamCoverage,” “computeMatrix,” and “plotHeatmap” programs in deepTools (56) were used to generate heatmap plots of genomic regions. Gene-centered analysis included 1.5 kb upstream of the TSS and 1.5 kb downstream of the TES with average coverage calculated in non-overlapping 50-nt bins and normalized to gene length.
Cosine similarity analyses were performed using R (version 4.0.2) package “coop” www.R-project.org/ (57). Since we have only 32 or 16 combinations of trinucleotides for CPDs and deamination available, we could only use the C-to-T and T-to-C mutation windows for the NPyPy (n = 32) and the C-to-T windows for the PyPyN (n = 16) to compare with the COSMIC trinucleotide signatures. Other mutations, such as C to G, C to A, T to G, and T to A, were ignored. This explains some level of similarity between, for example, SBS4 and SBS7, which are otherwise mostly unrelated (SBS4 is dominated by C-to-A mutations).

**Data availability**
All data are available from the GEO database (accession number GSE159807).

**Code availability**
All code for this project is available at https://github.com/GSE159807).

All data are available from the GEO database (accession number 36. Y. Tu, S. Tornaletti, G. P. Pfeifer, DNA repair domains within a human gene: Selective dimers are responsible for the vast majority of mutations induced by UVB irradiation in mammalian cells. J. Biol. Chem. 276, 44688–44694 (2001).

M. F. Laughery, A. J. Brown, K. A. Bohn, S. Sivapragasam, H. S. Morris, M. Tchmola, and F. R. de Bruil, A. J. Brown, S. Esaki, S. Lockwood, G. M. K. Poon, M. J. Smerdon, S. A. Roberts, J. J. Wyrrick, ETS transcription factors induce a unique UV damage signature that drives repair of sequences near the transcription initiation site. J. Mol. Biol. 354, 618–622 (2016).

T. A. Rosenquist, A. P. Grollman, Mutational signature of aristolochic acid: Glue to the recognition of a global disease. DNA Repair (Amst) 44, 205–211 (2015).

F. R. de Bruil, H. Rebel, Early events in UV carcinogenesis—DNA damage, target cells and mutant p53 foci. Photochem. Photobiol. 84, 382–387 (2008).

E. D. Pleasance, R. K. Cheetham, P. J. Stephens, D. J. McBride, S. J. Humphray, C. D. Greenman, I. Varela, M. L. Lin, G. R. Ordóñez, G. R. Bignell, K. Ye, J. Alipaz, M. J. Bauer, D. Beare, A. Butler, R. J. Carter, L. Chen, A. J. Cox, S. Edkins, P. I. Kokkos-Gonzales, N. A. Gormley, R. J. Grocock, C. D. Haudenschild, M. M. Hims, T. James, M. Jia, Z. Kingsbury, C. Leroy, J. Marshall, A. Menzies, L. J. Mudge, Z. Ning, T. Royce, O. B. Schulz-Trieglaff, A. Spirioud, L. A. Stebbings, L. Szakowski, J. Teague, D. Williamson, L. Chen, M. T. Ross, P. J. Campbell, D. R. Bentley, P. A. Futreal, M. R. Stratton, A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 463, 191–196 (2010).

P. G. Pfeifer, Mechanisms of UV-induced mutations and skin cancer. Genome Instability Disease 1, 99–113 (2020).

P. G. Pfeifer, A. Baranovskaya, UV wavelength-dependent DNA damage and human non-melanoma and melanoma skin cancer. Photochem. Photobiol. Sci. 11, 90–97 (2012).

D. E. Brash, UV signature mutations. Photochem. Photobiol. 91, 15–26 (2015).

J. Cadet, A. Grand, T. Douki, Solar UV radiation-induced DNA bipyrimidine photoproducts: Formation and mechanistic insights. Top. Curr. Chem. 356, 249–275 (2015).

Y. H. You, D. H. Lee, J. H. Yoon, S. Nakajima, A. Yasui, G. P. Fiefer, Cycllobutane pyrimidine dimers are responsible for the vast majority of mutations induced by UVB irradiation in mammalian cells. J. Biol. Chem. 276, 44688–44694 (2001).

M. F. Laughery, A. J. Brown, K. A. Bohn, S. Sivapragasam, H. S. Morris, M. Tchmola, and F. R. de Bruil, A. J. Brown, S. Esaki, S. Lockwood, G. M. K. Poon, M. J. Smerdon, S. A. Roberts, J. J. Wyrrick, Atypical UV photoproduces induce non-canonical mutation classes associated with driver mutations in melanoma. Cell Rep. 33, 108401 (2020).

A. Besaratia, J. I. Yoon, C. Schroeder, S. E. Bradforth, M. Cockburn, G. P. Fiefer, Wavelength dependence of ultraviolet radiation-induced DNA damage as determined by laser irradiation suggests that cycllobutane pyrimidine dimers are the principal DNA lesions produced by terrestrial sunlight. FASEB J. 25, 3079–3091 (2011).

J. Jans, W. Schul, Y.-G. Sert, Y. Rijksen, H. Rebel, A. P. M. Eker, S. Nakajima, H. van Steeg, N. A. Gormley, R. J. Grocock, C. D. Haudenschild, M. M. Hims, T. James, M. Jia, Z. Kingsbury, C. D. Greenman, I. Varela, M. L. Lin, G. R. Ordóñez, G. R. Bignell, K. Ye, J. Alipaz, M. J. Bauer, D. Beare, A. Butler, R. J. Carter, L. Chen, A. J. Cox, S. Edkins, P. I. Kokkos-Gonzales, N. A. Gormley, R. J. Grocock, C. D. Haudenschild, M. M. Hims, T. James, M. Jia, Z. Kingsbury, C. Leroy, J. Marshall, A. Menzies, L. J. Mudge, Z. Ning, T. Royce, O. B. Schulz-Trieglaff, A. Spirioud, L. A. Stebbings, L. Szakowski, J. Teague, D. Williamson, L. Chen, M. T. Ross, P. J. Campbell, D. R. Bentley, P. A. Futreal, M. R. Stratton, A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 463, 191–196 (2010).

P. G. Pfeifer, Mechanisms of UV-induced mutations and skin cancer. Genome Instability Disease 1, 99–113 (2020).

P. G. Pfeifer, A. Baranovskaya, UV wavelength-dependent DNA damage and human non-melanoma and melanoma skin cancer. Photochem. Photobiol. Sci. 11, 90–97 (2012).

D. E. Brash, UV signature mutations. Photochem. Photobiol. 91, 15–26 (2015).

J. Cadet, A. Grand, T. Douki, Solar UV radiation-induced DNA bipyrimidine photoproducts: Formation and mechanistic insights. Top. Curr. Chem. 356, 249–275 (2015).

Y. H. You, D. H. Lee, J. H. Yoon, S. Nakajima, A. Yasui, G. P. Fiefer, Cycllobutane pyrimidine dimers are responsible for the vast majority of mutations induced by UVB irradiation in mammalian cells. J. Biol. Chem. 276, 44688–44694 (2001).
39. R. Sabarinathan, L. Mularoni, J. Deu-Pons, A. Gonzalez-Perez, N. Lopez-Bigas, Nucleotide excision repair is impaired by binding of transcription factors to DNA. Nature 532, 264–267 (2016).

40. V. J. Cannistraro, J. S. Taylor, Acceleration of 5-methylcytosine deamination in cyclobutane dimers by G and its implications for UV-induced C-to-T mutation hotspots. J. Mol. Biol. 392, 1145–1157 (2009).

41. V. J. Cannistraro, S. Pondugula, Q. Song, J. S. Taylor, Rapid deamination of cyclobutane pyrimidine dimer photoproducts at TCG sites in a translationally and rotationally positioned nucleosome in vivo. J. Biol. Chem. 290, 26597–26609 (2015).

42. J. Drost, R. van Boxtel, F. Blokzijl, T. Mizutani, N. Sasaki, V. Sasselli, J. de Ligt, S. Behjati, J. E. Groeleman, T. van Wezel, S. Nik-Zainal, R. P. Kuiper, E. Cuppen, H. Clevers, Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science 358, 234–238 (2017).

43. A. Canela, S. Sridharan, N. Sciascia, A. Tubbs, P. Meltzer, B. P. Sleckman, A. Nussenzweig, DNA breaks and end resection measured genome-wide by end sequencing. Mol. Cell 63, 898–911 (2016).

44. A. Sancar, Mechanisms of DNA excision repair. Science 266, 1954–1956 (1994).

45. M. F. Denissenko, A. Pao, M. Tang, G. P. Pfeifer, Preferential formation of benzo[alpha]pyrene diol epoxide-DNA adduct formation at methylated versus unmethylated CpG sites in the p53 gene. Carcinogenesis 20, 1085–1089 (1999).

46. K. Takasawa, C. Masutani, F. Hanaoka, S. Iwai, Chemical synthesis and translesion replication of a cis-syn cyclobutane thymine–uracil dimer. Nucleic Acids Res. 32, 1738–1745 (2004).

47. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin. 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

48. A. K. Quinlan, BEDTools: The Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinformatics 47, 11.12.1–34 (2014).

49. H. Li, Tabix: Fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics 27, 718–719 (2011).

50. V. J. Cannistraro, J. S. Taylor, Acceleration of 5-methylcytosine deamination in cyclobutane dimers by G and its implications for UV-induced C-to-T mutation hotspots. J. Mol. Biol. 392, 1145–1157 (2009).

51. V. J. Cannistraro, S. Pondugula, Q. Song, J. S. Taylor, Rapid deamination of cyclobutane pyrimidine dimer photoproducts at TCG sites in a translationally and rotationally positioned nucleosome in vivo. J. Biol. Chem. 290, 26597–26609 (2015).

52. J. Drost, R. van Boxtel, F. Blokzijl, T. Mizutani, N. Sasaki, V. Sasselli, J. de Ligt, S. Behjati, J. E. Groeleman, T. van Wezel, S. Nik-Zainal, R. P. Kuiper, E. Cuppen, H. Clevers, Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. Science 358, 234–238 (2017).

53. A. Canela, S. Sridharan, N. Sciascia, A. Tubbs, P. Meltzer, B. P. Sleckman, A. Nussenzweig, DNA breaks and end resection measured genome-wide by end sequencing. Mol. Cell 63, 898–911 (2016).

54. A. Sancar, Mechanisms of DNA excision repair. Science 266, 1954–1956 (1994).

55. M. F. Denissenko, A. Pao, M. Tang, G. P. Pfeifer, Preferential formation of benzo[alpha]pyrene adducts at lung cancer mutational hotspots in P53. Science 274, 430–432 (1996).

56. M. S. Tang, J. B. Zheng, M. F. Denissenko, G. P. Pfeifer, Y. Zheng, Use of UvrABC nuclease to quantify benzo[alpha]pyrene diol epoxide-DNA adduct formation at methylated versus unmethylated CpG sites in the p53 gene. Carcinogenesis 20, 1085–1089 (1999).

57. K. Takasawa, C. Masutani, F. Hanaoka, S. Iwai, Chemical synthesis and translesion replication of a cis-syn cyclobutane thymine–uracil dimer. Nucleic Acids Res. 32, 1738–1745 (2004).

58. Q. Song, S. M. Sherrer, Z. Suo, J. S. Taylor, Preparation of site-specific Trp[cis-cis]CG cis-syn cyclobutane dimer-containing template and its error-free bypass by yeast and human polymerase n. J. Biol. Chem. 287, 8021–8028 (2012).

59. D. B. Yarosh, A. Rosenthal, R. Moy, Six critical questions for DNA repair enzymes in skincare products: A review in dialog. Clin. Cosmet. Investig. Dermatol. 12, 617–624 (2019).

60. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

61. G. G. Faust, I. M. Hall, SAMBLASTER: Fast duplicate marking and structural variant read extraction. Bioinformatics 30, 2503–2505 (2014).

62. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

63. A. R. Quinlan, BEDTools: The Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinformatics 47, 11.12.1–34 (2014).

64. H. Li, Tabix: Fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics 27, 718–719 (2011).

65. V. J. Cannistraro, J. S. Taylor, Acceleration of 5-methylcytosine deamination in cyclobutane dimers by G and its implications for UV-induced C-to-T mutation hotspots. J. Mol. Biol. 392, 1145–1157 (2009).

66. A. Canela, S. Sridharan, N. Sciascia, A. Tubbs, P. Meltzer, B. P. Sleckman, A. Nussenzweig, DNA breaks and end resection measured genome-wide by end sequencing. Mol. Cell 63, 898–911 (2016).

67. A. Sancar, Mechanisms of DNA excision repair. Science 266, 1954–1956 (1994).

68. M. F. Denissenko, A. Pao, M. Tang, G. P. Pfeifer, Preferential formation of benzo[alpha]pyrene adducts at lung cancer mutational hotspots in P53. Science 274, 430–432 (1996).

69. M. S. Tang, J. B. Zheng, M. F. Denissenko, G. P. Pfeifer, Y. Zheng, Use of UvrABC nuclease to quantify benzo[alpha]pyrene diol epoxide-DNA adduct formation at methylated versus unmethylated CpG sites in the p53 gene. Carcinogenesis 20, 1085–1089 (1999).

70. K. Takasawa, C. Masutani, F. Hanaoka, S. Iwai, Chemical synthesis and translesion replication of a cis-syn cyclobutane thymine–uracil dimer. Nucleic Acids Res. 32, 1738–1745 (2004).

71. Q. Song, S. M. Sherrer, Z. Suo, J. S. Taylor, Preparation of site-specific Trp[cis-cis]CG cis-syn cyclobutane dimer-containing template and its error-free bypass by yeast and human polymerase n. J. Biol. Chem. 287, 8021–8028 (2012).

72. D. B. Yarosh, A. Rosenthal, R. Moy, Six critical questions for DNA repair enzymes in skincare products: A review in dialog. Clin. Cosmet. Investig. Dermatol. 12, 617–624 (2019).

73. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

74. G. G. Faust, I. M. Hall, SAMBLASTER: Fast duplicate marking and structural variant read extraction. Bioinformatics 30, 2503–2505 (2014).