Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution

Jinchuan Hu,[1,4] Sheera Adar,[2,4] Christopher P. Selby,[1] Jason D. Lieb,[3] and Aziz Sancar[1]

[1]Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599, USA; [2]Integrative Program for Biological and Genome Sciences, University of North Carolina, Chapel Hill, North Carolina 27599, USA; [3]Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA

We developed a method for genome-wide mapping of DNA excision repair named XR-seq (excision repair sequencing). Human nucleotide excision repair generates two incisions surrounding the site of damage, creating an ~30-mer. In XR-seq, this fragment is isolated and subjected to high-throughput sequencing. We used XR-seq to produce stranded, nucleotide-resolution maps of repair of two UV-induced DNA damages in human cells: cyclobutane pyrimidine dimers (CPDs) and (6-4) pyrimidine–pyrimidone photoproducts [(6-4)PPs]. In wild-type cells, CPD repair was highly associated with transcription, specifically with the template strand. Experiments in cells defective in either transcription-coupled excision repair or general excision repair isolated the contribution of each pathway to the overall repair pattern and showed that transcription-coupled repair of both photoproducts occurs exclusively on the template strand. XR-seq maps capture transcription-coupled repair at sites of divergent gene promoters and bidirectional enhancer RNA (eRNA) production at enhancers. XR-seq data also uncovered the repair characteristics and novel sequence preferences of CPDs and (6-4)PPs. XR-seq and the resulting repair maps will facilitate studies of the effects of genomic location, chromatin context, transcription, and replication on DNA repair in human cells.

Nucleotide excision repair is the sole mechanism for removing bulky DNA base lesions, which are caused by a variety of genotoxic agents, including UV radiation (Sancar 1996; Wood 1997; Reardon and Sancar 2005). Nucleotide excision repair consists of two pathways, global repair and transcription-coupled repair, that differ primarily in the damage recognition step (Mellon et al. 1987; Hanawalt and Spivak 2008). Mutations that inactivate global genome repair cause the genetic diseases xeroderma pigmentosum (XP), while those that inactivate transcription-coupled repair cause Cockayne syndrome (CS) (Cleaver et al. 2009). In global repair, damage recognition is accomplished by XPC together with RPA and XPA (Sugasawa et al. 1998; Wakasugi and Sancar 1998; Reardon and Sancar 2003), and the ultimate dual-incision complex contains all core excision repair factors except XPC (Wakasugi and Sancar 1998). In transcription-coupled repair, damage recognition is mediated by a stalled elongating RNA polymerase II (Pol II), which, with the aid of the CSB translocase, recruits the core excision repair factors except XPC. The subsequent steps of the two pathways consist of the "core" nucleotide excision repair pathway.

The core mechanism of human DNA excision repair is currently well understood and consists of dual incisions bracketing the lesion, which results in removal of a single strand that is nominally 30 nucleotides (nt) in length (24- to 32-nt range, hereafter referred to as "30 nts") (Huang et al. 1992; Svoboda et al. 1993; Kemp et al. 2012). The resulting gap in the duplex is filled by DNA polymerases, and the newly synthesized repair patch is then ligated (Wood 1997; Reardon and Sancar 2005). The excision reaction is carried out by the coordinated activities of six repair factors: RPA, XPA, XPC, TFIIH, XPG, and XPF-ERCC1. The core excision reaction has been successfully reconstituted in vitro (Mu et al. 1995, 1996; Wood 1997), although the transcription-coupled repair has not. It should be noted that even though general and

[4]These authors contributed equally to this work.
Corresponding authors: aziz_sancar@med.unc.edu, jdlieb@uchicago.edu

transcription-coupled repairs recognize damage by different mechanisms, the compositions of the ultimate dual-incision complexes are identical. Both contain TFIIH, and neither contains XPC.

Repair in vivo is influenced by factors aside from the six proteins that comprise the core excision repair machinery, such as chromatin organization, genomic location, transcription, and DNA replication (Fong et al. 2013; Adam et al. 2014). To assess the contribution of these factors to excision repair, it would be useful to compare each of these variables with high-resolution maps of repair sites across the genome. However, several factors make genome-wide studies of DNA repair challenging. The main obstacles are that (1) the relevant entity for detection is a single damaged nucleotide or a dinucleotide; (2) the specific sites of damage vary widely from cell to cell, and any result obtained from a cell culture represents a projection of all repair occurring in the cellular population; and (3) at a specific point in time, only a small fraction of the damages is in the process of being repaired. Thus, any method used to map DNA repair must yield a high signal to noise ratio and high sensitivity. Although genome-wide maps of UV damage distribution (Teng et al. 2011; Bryan et al. 2014; Zavala et al. 2014; Powell et al. 2015) and single-gene analyses of excision repair in human cells and yeast are available (Pfeifer et al. 1991; Tornaletti and Pfeifer 1994; Denissenko et al. 1996; Li et al. 2000, 2014), methods for genome-wide detection of excision repair at nucleotide resolution have not been reported.

Here we extended a recent method for capturing the excised 30-mer released in vivo during nucleotide excision repair (Hu et al. 2013; Choi et al. 2014; Kemp et al. 2014) to overcome many of the obstacles for high-resolution analysis of human excision repair. XR-seq (excision repair sequencing) uses strand-specific sequencing of captured 30-mers to generate a genome-wide map of human excision repair at single-nucleotide resolution. Using XR-seq, we mapped repair of both cyclobutane pyrimidine dimers (CPDs) and (6-4) pyrimidine–pyrimidone photoproducts [(6-4)PPs] throughout the human genome. Furthermore, for both damages, we separated the activities of global genome repair and transcription-coupled repair by using an XP-C mutant cell line that lacks general repair (Venema et al. 1991; van Hoffen et al. 1995) and a CS-B mutant cell line lacking solely transcription-coupled repair (Venema et al. 1990).

Results

### Characterization of nucleotide excision repair kinetics in wild-type and mutant cells

In mammalian cells, the UV-induced CPD and (6-4)PP are removed by dual incision of the phosphodiester bonds 20 nt ± 5 nt upstream of and 6 nt ± 3 nt downstream from the photoproduct (Fig. 1A). This dual incision generates a 24- to 32-nt-long oligomer carrying the photoproduct, which is referred to here as the "nominal 30-mer." The nominal 30-mer is released in complex with the transcription/repair factor TFIIH in both general and transcription-cou-

pled repair (Kemp et al. 2012; Hu et al. 2013; Choi et al. 2014). We performed experiments in three cell types: the wild-type skin fibroblast cell line NHF1, which is proficient in both transcription-coupled and general excision repair; XP-C, a mutant skin fibroblast cell line that lacks general repair but is proficient in transcription-coupled repair (Venema et al. 1991; van Hoffen et al. 1995); and CS-B, a mutant skin fibroblast cell line lacking transcription-coupled repair but proficient in general repair (Venema et al. 1990). Prior to initiating XR-seq experiments, we sought to characterize the kinetics of repair in each cell line so that we could choose appropriate time points for analysis.

Cells were irradiated with 20 J/m^2 UVC and collected following incubation times ranging from 20 to 240 min. Cells were lysed, and the excision products containing either CPD or (6-4)PP were isolated with anti-photoproduct antibodies (Hu et al. 2013). Analyzed by autoradiography, the primary excision products from the three cell lines were 24–32 nt in length and were processed rapidly to smaller fragments with a median size of ∼20 nt (Fig. 1B). The (6-4)PPs are repaired with higher efficiency than CPDs in wild-type cells because they are recognized with higher affinity by the general excision repair system (Mu et al. 1997; Reardon and Sancar 2003; Hu et al. 2013). Thus, in wild-type cells, measurable CPD repair was observed only after (6-4)PP repair was nearly complete. In contrast, in XP-C cells that lack global repair, a stalled RNA polymerase is the only recognition signal for the core excision repair complex (Selby and Sancar 1993; Lindsey-Boltz and Sancar 2007; Hanawalt and Spivak 2008), and therefore the rate of repair is proportional to the abundance of transcription-blocking photoproducts. Because CPDs are approximately five times more abundant than (6-4)PPs (Mitchell 1988), in XP-C cells, they are excised at a rate that is approximately fivefold faster than (6-4) PPs. The converse situation occurs in the CS-B cell line that lacks transcription-coupled repair, wherein the preferential repair of (6-4)PP relative to CPDs is even greater than is observed in wild-type cells. These results suggest that in wild-type NHF1 cells, CPD repair at early time points is due predominantly to transcription-coupled repair. Therefore, in the CS-B cells that lack transcription-coupled repair, the earliest repair events are predominantly directed to (6-4)PPs. With this in mind, we used the 1-h time point as the basis for comparison in all data analysis that followed because it allowed us to make comparisons between CPDs and (6-4)PPs under similar cellular conditions.

The XR-seq procedure

To perform XR-seq, cells were irradiated with 20 J/m^2 UVC and collected following incubation for repair (Fig. 1C,D), and lysate was prepared from the cells. The primary excision product (nominal 30-mer) was isolated by TFIIH immunoprecipitation followed by ligation of 5′ and 3′ adapters compatible with the Illumina TruSeq small RNA protocol. Following adapter ligation, oligomers carrying the CPD or (6-4)PP were
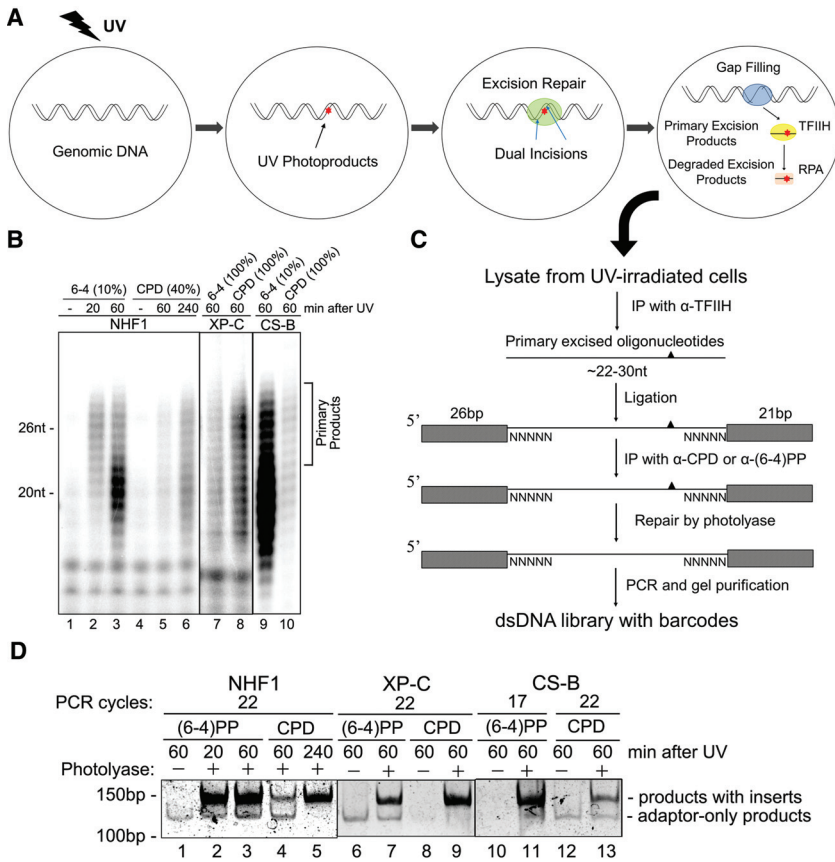
Figure 1. The XR-seq method. (*A*) Schematic of the procedure to isolate the nominal 30-mer generated by nucleotide excision repair. UV-induced photoproducts are removed from the genome by dual incisions, releasing the primary excision product in complex with TFIIH. The primary product is degraded with a half-life of ~2 h to ~20-nt-long fragments that are bound to RPA. For XR-seq, the primary products are isolated by TFIIH immunoprecipitation. (*B*) Excision patterns of photoproducts in NHF1 (wild-type), XP-C (deficient in global repair), and CS-B (deficient in transcription-coupled repair) cells. The excised oligonucleotides were immunoprecipitated with either anti-(6-4)PP antibodies or anti-CPD antibodies, and then the indicated fraction of purified DNAs was radiolabeled at the 3′end with ^{32}P-cordycepin and analyzed on sequencing gels. (*C*) Procedure for preparation of the dsDNA library for the Illumina HiSeq 2000 platform. (*D*) Analysis of dsDNA libraries of the excised nominal 30-mer by polyacrylamide gel electrophoresis. One percent of the ligation products were PCR-amplified with the indicated cycles.

immunoprecipitated by monoclonal antibodies specific to one damage or the other. To allow downstream DNA amplification of the damaged templates, the photoproducts were repaired by either CPD photolyase or (6-4)PP photolyase (Selby and Sancar 2006). The repaired products were then amplified by PCR using 50- and 63-nt-long primers that introduce specific barcodes compatible with the Illumina TruSeq small RNA kit. The PCR products containing excised oligonucleotides were ~145 base pairs (bp) in length, and the "empty" products were 118 bp in length (Fig. 1D). The appearance of ~145-bp products only after photolyase repair shows that the resulting libraries were comprised exclusively of previously damaged DNA fragments, and there is no background from undamaged DNA. PCR products were purified by PAGE, and samples from the 1-h time point were sequenced using the Illumina HiSeq 2000 platform, producing single-end 50-nt reads.

CPD repair occurs preferentially at transcribed regions, while (6-4) photoproduct repair is distributed uniformly throughout the genome

We mapped the XR-seq reads from wild-type NHF1 cells, obtaining strand-specific, genome-wide DNA repair signal across the human genome. At a chromosome-wide scale (50 Mb), (6-4)PP repair is relatively evenly distributed. CPD repair is more heterogeneous, with regions of relatively higher and lower repair. We compared our DNA repair tracks with ENCODE stranded total RNA-seq tracks obtained from NHDF skin fibroblast cells (The ENCODE Project Consortium 2012) and found that regions of elevated CPD repair correspond to regions with higher levels of transcription (chromosome 7 is depicted in Fig. 2A; all human chromosomes in Supplemental Fig. 1). Zooming in on a 1.5-Mb region, the relationship between CPD repair levels and RNA levels becomes clearer, with higher strand-specific XR-seq signal at transcribed genes. By convention, the RNA signal is mapped to the "sense" or "nontemplate" strand, which corresponds to the "+" strand for genes transcribed from left to right and the "−" strand for genes transcribed from right to left. XR-seq signal is enriched on the transcribed template strand, consistent with the fact that the signal for excision repair recruitment is stalled RNA Pol II (Selby et al. 1997; Lindsey-Boltz and Sancar 2007; Hanawalt and Spivak 2008). Over 68% of all CPD XR-seq reads overlap the transcribed regions of genes (±5000 bp). In contrast, only 56% of (6-4)PP reads are derived from the same regions, which is still higher than the 46% that one would expect for uniform repair throughout the genome ($P < 0.02$; Materials and Methods) (Fig. 2B; Supplemental Table 1). Strand-specific signal of biological replicates in which independent cell populations were UV-irradiated and subjected to XR-seq was highly correlated across the genome (Fig. 2C; Supplemental Figs. 2, 3), with even greater correlation over exons (Supplemental Fig. 4).

Figure 2. Genome-wide maps of CPD and (6-4)PP excision repair in NHF1 wild-type cells. (*A*) Distribution of the XR-seq signal, separated by strand, for CPD and (6-4)PP over the entire chromosome 7 (chr7; *top*) or focused on a 1.5-Mb region (*bottom*). ENCODE total stranded RNA-seq tracks in black are plotted *above* the XR-seq tracks for comparison. Arrows on the *bottom* depict the direction and length of annotated genes. (*B*) Distribution of the aligned reads between annotated genes (UCSC hg19 genes; green), the 5000 bp upstream of the gene (light green), the 5000 bp downstream from the gene (gray), and intergenic regions (blue). For comparison are the average results of 50 permutations of a random set of 26mers from the hg19 genome. (*C*) Spearman's correlation coefficient ρ calculated between biological replicates (experiments conducted in two independently UV-treated populations of cells) and between CPD and (6-4)PP XR-seq in NHF1 cells. Samples are ordered by hierarchal clustering. Darker box shades indicate higher correlation. (*D*) Average profile of CPD XR-seq and (6-4)PP XR-seq over all University of California at Santa Cruz (UCSC) reference genes. Genes and XR-seq signal were separated based on their direction to allow differentiation between template (purple) and nontemplate (turquoise) strand repair. Signals over the gene body were normalized to a 3000-bp window to allow for comparison.

We next tested whether there is preferential repair of the template strand throughout gene bodies, genome-wide. We plotted the average repair profile for template and nontemplate strands over genes and the regions 1000 bp upstream of and downstream from them. We observed the preference of template strand repair of CPD beginning at the transcription start site (TSS), persisting until the annotated transcription end site (TES) and, to a lesser extent, beyond, probably due to inaccurate annotations of some termination sites (Fig. 2D). In the regions upstream of the genes, the nontemplate strand (relative to the coding gene) is preferentially repaired. This is consistent with the observation that ~80% of active human genes exhibit divergent initiation from their promoters, producing upstream RNAs (uaRNAs) (Core et al. 2008, 2014). In contrast to CPD repair, the repair of (6-4)PPs is not elevated over the gene body and has no strand preference (Fig. 2D).

Genetic separation of transcription-coupled and global excision repair reveals distinct genome-wide patterns

We separated the transcription-coupled and global genome repair mechanisms by using mutant cell lines. XP-C cells are proficient only in transcription-coupled

repair (Venema et al. 1991; van Hoffen et al. 1995), while CS-B cells are proficient only in general (transcription-uncoupled) excision repair (Venema et al. 1990). Thus, we conducted XR-seq experiments for CPD and (6-4)PP in XP-C and CS-B cells, allowing us to analyze the contribution of each of the pathways in the repair of both damages genome-wide. The repair patterns in the two mutant cell lines are completely distinct (Fig. 3A; Supplemental Fig. 5A,B).

In XP-C cells, the enrichment of CPD repair at transcribed regions that is seen in wild-type cells is amplified. Not only is there an elevated relative level of repair over the transcribed regions of genes, but repair of intergenic regions and nontemplate strands is essentially absent. As in wild-type cells, CPD repair also occurs on the nontemplate strand upstream of the TSS, likely as the result of divergent transcription occurring at the promoters (Fig. 3C; Core et al. 2008, 2014). In the absence of general excision repair, like CPD repair, (6-4)PP repair is completely dependent on transcription-coupled repair and is observed solely on the template strand (Fig. 3A; Supplemental Fig. 5C). In XP-C cells, <10% of total reads fall outside of annotated genes and the 5000 bp flanking them (54% expected by chance) (Fig. 3B; Supplemental Table 1).

In contrast, in CS-B cells in which transcription-coupled repair is absent, CPD repair appears relatively uniform throughout the genome and no longer occurs preferentially over transcribed regions (Fig. 3A,B,D). In the absence of transcription-coupled repair in CS-B cells, (6-4)PP repair is spread throughout the genome in a pattern similar to that seen in wild-type cells, consistent with the fact that (6-4)PPs are normally repaired by the general excision pathway in wild-type cells (Fig. 3A; Supplemental Fig. 5C). There is a very slight elevation in repair of both damages around the TSS. Thus, although the two damage types have very different repair patterns and kinetics in wild-type cells, in the two mutant cell lines, the pattern of (6-4)PP repair was very similar to the CPD repair pattern (Fig. 3A,B).

Detection of nucleotide excision repair at enhancers

While the vast majority of DNA repair signal in XP-C cells maps to annotated genes and the regions immediately upstream of and downstream from them, discrete repair

events also occur on the nontemplate strand and in intergenic regions (Fig. 3A [yellow shade], B). Given the high specificity of the assay to transcribed regions, we speculated that many of these signals could be explained as sites of enhancer RNA (eRNA) transcription (Kim et al. 2010; Wang et al. 2011; Hah et al. 2013). Similar to genes, eRNAs are driven by bidirectional promoters (Core et al. 2014). However, they are generally shorter and less stable and are therefore more difficult to detect by standard RNA-seq.

We compared the transcription-coupled repair XR-seq signal from XP-C cells with ENCODE data sets from NHDF skin fibroblasts. We found that sites of repair outside of genes coincide with DNase I hypersensitivity signal and H3K27ac ChIP-seq (chromatin immunoprecipitation [ChIP] coupled with deep sequencing) signal, which correlated with active enhancers (Creyghton et al. 2010; Zentner et al. 2011). Many of the repair peaks also align with enhancers called by the chromHMM segmentations (Ernst and Kellis 2010) in lung fibroblasts (Fig. 3E, orange/yellow sites; Supplemental Fig. 5D). A distinct pattern of XR-seq signal arose near putative strong enhancers, consisting of two distinct peaks (for plus and minus strand repair, respectively) on each side of the enhancer (Fig. 3F; Supplemental Fig. 5E). This pattern is consistent with repair occurring on the template strand of divergently transcribed eRNAs. In contrast, there was essentially no XR-seq signal around intergenic DNase peaks that are not accompanied by enhancer marks.
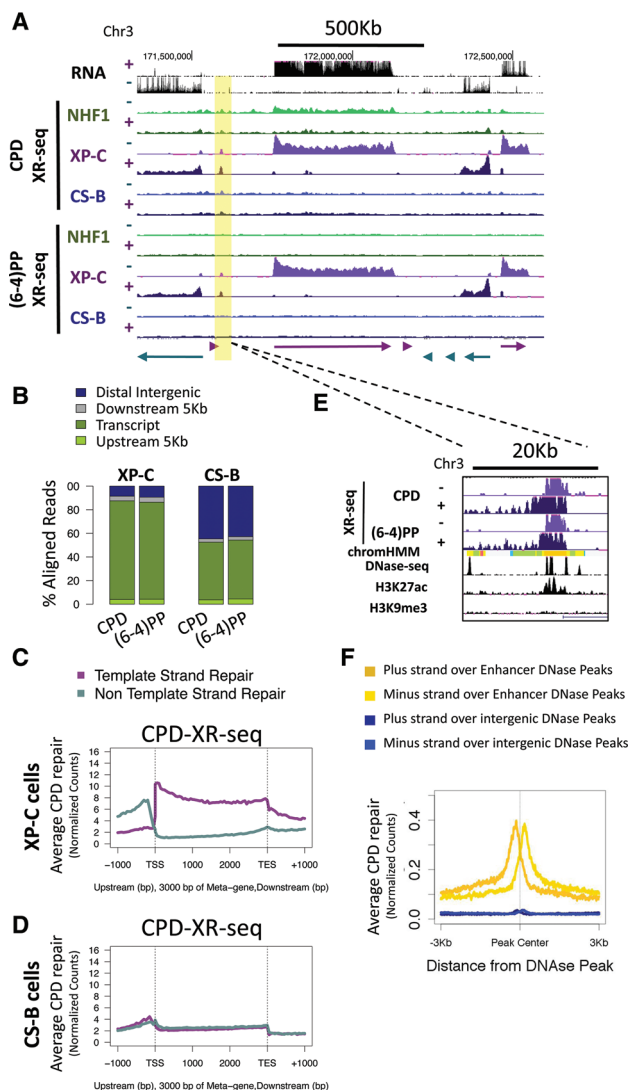
Figure 3. Mapping transcription-coupled and global excision repair. (*A*) Distribution of the XR-seq signal, separated by strand, for CPD (*top*) and (6-4)PP (*bottom*) over a 1.5-Mb region of chromosome 3. Shown is signal from NHF1 wild-type cells (green), XP-C cells that are proficient in only transcription-coupled repair (purple), and CS-B cells that are proficient only in global excision repair (blue). ENCODE total stranded RNA-seq tracks in black are plotted *above* the XR-seq tracks for comparison. Arrows on the *bottom* depict the direction of annotated genes. (*B*) Genomic distribution of aligned reads between genes (green), the 5000 bp upstream of the gene (light green), the 5000 bp downstream from the genes (gray), and intergenic regions (blue) shows that XR-seq signal for both damages is highly enriched over genes in XP-C cells but is more evenly distributed in CS-B cells. (*C*,*D*) Average profile of transcription-coupled (*C*) and global (*D*) excision repair of CPD over all UCSC reference genes. Genes and XR-seq signal were separated based on their direction to allow differentiation between template (purple) and nontemplate (turquoise) strand repair. Signals over gene body were normalized to a 3000-bp window to allow for comparison. (*E*) Focusing in on the intergenic region highlighted in yellow in *A*, transcription-coupled repair of CPD and (6-4)PP in XP-C cells occurs on both strands and overlaps DNase-seq (DNase sequencing) and H3K27ac ChIP-seq (chromatin immunoprecipitation [ChIP] coupled with deep sequencing) signal and chromHMM strong enhancer prediction (yellow/orange). (*F*) Average plus strand (dark lines) and minus strand (lighter lines) CPD transcription-coupled repair in XP-C cells around the center of intergenic DNase peaks. DNase peaks were divided into "enhancer peaks," which overlap H3K27ac peaks or chromHMM strong enhancers (golds), and "intergenic peaks," which do not (blues).

Thus, XR-seq appears to detect transcription-coupled repair at all sites of RNA Pol II transcription.

The level of transcription-coupled repair is highly correlated with RNA levels

To further investigate the relationship between the excision repair pathways and transcription, we integrated the DNA repair signal over exons with the available ENCODE fibroblast RNA-seq data. Transcription-coupled repair of CPD in wild-type cells or of either damage in XP-C cells is highly correlated to the RNA levels (Spearman's correlation coefficient of ~0.8) (Supplemental Fig. 4). Genes were stratified based on their RNA expression levels (FPKM [fragments per kilobase of exons per million bases mapped]) (Supplemental Table 2). We plotted the average repair profile for CPD over each of these groups for the transcribed and nontemplate strands separately, relative to the TSS (Fig. 4A,B) and TES (Supplemental Fig. 6). In NHF1 wild-type cells, higher levels of transcription are associated with higher levels of repair in the template strand (Fig. 4A), beginning at the TSS and continuing into the gene body. CPD repair upstream of the TSS also appears to scale with expression of the downstream gene, although to a lesser extent (also seen by plotting relative to the nontemplate strand) (Fig. 4B). In XP-C cells, at the TSS, there is a steep rise in CPD repair levels that is highly associated with levels of RNA (Fig. 4A) and persists into the gene body and even beyond the termination site (Supplemental Fig. 6). On the nontemplate strand, the situation is reversed, with no CPD repair in the gene body but higher repair rates in the upstream regions that correspond to the RNA levels of the annotated gene (Fig. 4B). The border between the repaired and nonrepaired states is not as clear on the nontemplate strand, with some level of nontemplate CPD repair occurring in the first 1000 bp of the gene. In the CS-B cell line, there is a slight elevation of CPD repair associated with expressed genes on both the template and nontemplate strands, with repair on the template strand slightly elevated toward the region upstream of the TSS (Fig. 4A,B). Association of repair with expression for the (6-4)PP is only observed in XP-C cells (Supplemental Figs. 7, 8).

Excised fragments reveal sequence preferences for damage formation and excision sites

The short length of the excised oligomer allowed it to be completely sequenced within the 50-nt reads, which enabled us to determine the precise length of the sequenced excised fragments. Consistent with the autoradiograph results in the NHF1 wild-type cell line (Fig. 1B), for both CPD and (6-4)PP, most of the fragments fall between 20 and 30 nt, and the mean length of the oligomers was ~26 nt. (Fig. 5A; Supplemental Fig. 9).

The photoproduct is expected to be 6 nt ± 3 nt from the 3′ end based on in vitro data (Huang et al. 1992). Indeed, compared with the 67% expected by chance, >98% of excised fragments contain dipyrimidines at the position 3–9

nt from the 3′ end. For the detailed analysis that follows in this section, we focused on excised fragments that were exactly 26 nt in length (Fig. 5B; Supplemental Figs. 10–12). We calculated the frequency of the possible dipyrimidines (TT, TC, CT, and CC) at each of the positions along the 26mer. For both CPD and (6-4)PP, there is a strong enrichment of dipyrimidines at the 3′ end, peaking at positions 19–20 and 21–22 for CPD and 18–19 and 19–20 for (6-4)PP, which is 5–6 nt and 6–7 nt from the 3′ end, respectively. In both experiments, >80% of the reads had dipyrimidines at the respective positions. These enrichments were significant compared with the distribution of dipyrimidines in the human genome ($P < 0.02$) (Materials and Methods; Supplemental Table 3). The most abundant dipyrimidine in CPD XR-seq was TT, and in (6-4)PP-XR-seq fragments the most abundant was TC, as previously reported (Mitchell et al. 1992; Douki and Cadet 2001). These patterns were consistent for 26–30mer reads and were also observed in the two mutant cell lines (Supplemental Figs. 10, 11). In addition, there is a partial depletion of dipyrimidines around position 9–10 from the 5′ end. This depletion is maintained at an equal distance from the 5′ end regardless of fragment length, suggesting a sequence preference in determining the 5′ incision event (Fig. 5B; Supplemental Fig. 10). Finally, there is a depletion of TT and TC at the first 5′ position, but this can be explained as a bias introduced in the molecular biology procedures. Adapter ligation is dependent on annealing of the excised fragments to the adapter oligomer. Therefore, this depletion of T may be a consequence of preferential annealing of G/Cs over T/As.

There is a preference for C upstream of and A downstream from (6-4) photoproducts

To examine sequence context preferences around the UV damage itself, we measured the frequency of the nucleotides flanking the dipyrimidines (Fig. 5C; Supplemental Fig. 12; Supplemental Table 4). For TT dinucleotides at position 19–20 in the CPD XR-seq fragments, there is a preference for C 5′ to the putative photoproduct site and a preference for T concomitant with a depletion of A and G 3′ to it. For TC at position 19–20 in the (6-4)PP XR-seq fragments, there is a pronounced preference for C 5′ and A 3′ to the putative photoproduct site ($P < 0.02$) (Materials and Methods; Supplemental Table 4). These preferences are consistent with previous reports on sequence effects on photoproduct formation (Mitchell et al. 1992; Bryan et al. 2014).

To rule out that the observed sequence context preference of A downstream from a (6-4)PP in the excised oligomer is the result of preferential repair by the photolyases during XR-seq library preparation, we performed in vitro repair of oligonucleotides. Because synthetic TC-(6-4)PP is not available and the same sequence preference is observed for TT-(6-4)PP (Supplemental Fig. 12A), we performed in vitro repair of oligonucleotides containing either a T(6-4)TA or T(6-4)TG. Both are repaired at similar efficiencies (Fig. 5D,E). Taken together with the fact that immunoprecipitation of these oligomers was essentially
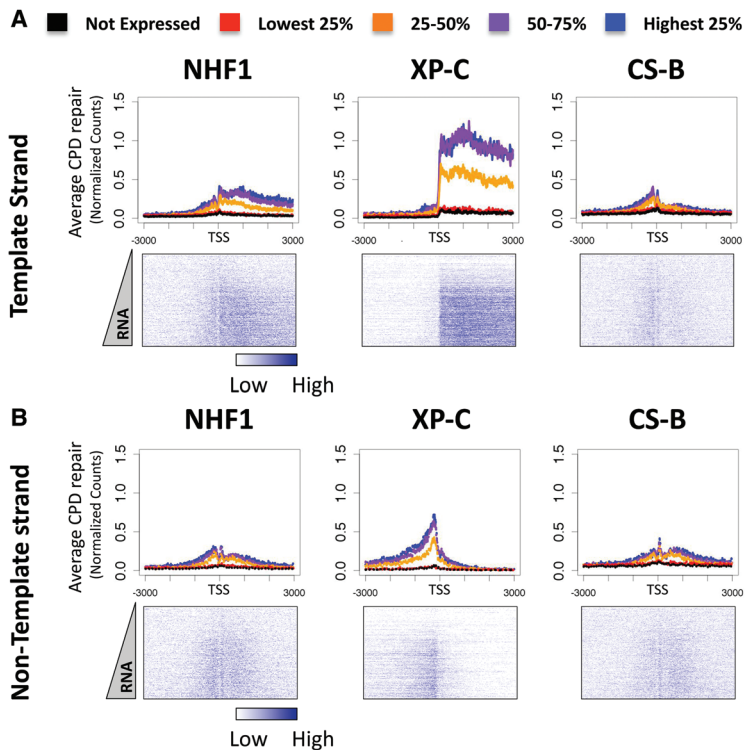
Figure 4. Strong association of transcription-coupled repair with RNA levels. CPD repair profile around the TSS is plotted for the template strand (*A*) or nontemplate strand (*B*). (*Top* row) Average profile for five gene groups. Genes were divided into five groups based on expression level and include nonexpressed (black), lowest (red), low (orange), high (purple), and highest (blue) based on the calculated FPKM from RNA-seq in NHDF cells. (*Bottom* row) Corresponding heat map of repair over all expressed genes, which are ordered by ascending FPKM.

identical (Materials and Methods), we conclude that (6-4) PP forms preferentially in the TCA sequence context.

Discussion

XR-seq produces single-nucleotide-resolution genome-wide maps of DNA excision repair

Technological improvements in genomics along with our recent ability to isolate the nominal 30-mer released during nucleotide excision repair (Kemp et al. 2012; Hu et al. 2013; Choi et al. 2014) have enabled us for the first time to create high-resolution, stranded, genome-wide maps for excision repair in human cells. We validated XR-seq by showing that the obtained sequence lengths are, on average, 26 nt long and mostly span between 20 and 30 nt. Analysis of the sequences of excised fragments obtained by XR-seq results are consistent with the 5′ and 3′ incision events that generated the excised oligonucleotides. The sequences 6 nt from the 3′ end are highly enriched for pyrimidine dinucleotides. Furthermore, we found that, analyzing sequences of different lengths, the distance of pyrimidine dinucleotides from the 3′ end is maintained (Supplemental Fig. 10). Taken together, these findings indicate that the integrity of the excised oligonucleotides that we analyzed was preserved throughout the molecular biology steps.

XR-seq analysis uncovered novel sequence context preferences for (6-4)PP formation

The sequence context preferences that we observed could be the product of preferential damage formation or prefer-

ential repair or a result of preferential binding by the anti-CPD or anti-(6-4)PP antibody or preferential repair by the CPD or (6-4) photolyase. The following seems to rule out damage recognition as the cause of the sequence context of the excised oligonucleotides: (1) The same nominal 30-mer sequence pattern is observed in transcription-dependent and transcription-independent repair (Supplementary Figs. 11, 12B), in which damage recognition mechanisms are distinct, and (2) all dipyrimidine photoproducts are equally efficient at blocking transcription and therefore are equally susceptible to transcription-coupled repair. Preferential repair by photolyase during sample preparation for XR-seq was ruled out because, under our experimental conditions, the (6-4)PP in the two sequence contexts (TTA and TTG) were repaired equally (Fig. 5D,E). Finally, a recent study examining photoproduct distribution in the yeast genome by a method that did not rely on photoproduct immunoprecipitation (Bryan et al. 2014) also found TCA preference for (6-4)PP formation, thus ruling out selective immunoprecipitation during our sample preparation for XR-seq. Our immunoprecipitation experiments with synthetic oligomers with (6-4)PP in TTA and TTG contexts revealed nearly identical immunoprecipitation efficiencies (Materials and Methods). Thus, the preferences seem to be inherent to the formation of the damages themselves.

Contribution of general and transcription-coupled repair to the removal of CPD and (6-4) photoproducts

In wild-type NHF1 fibroblasts, (6-4)PPs are preferentially recognized by global repair, while CPDs are recognized
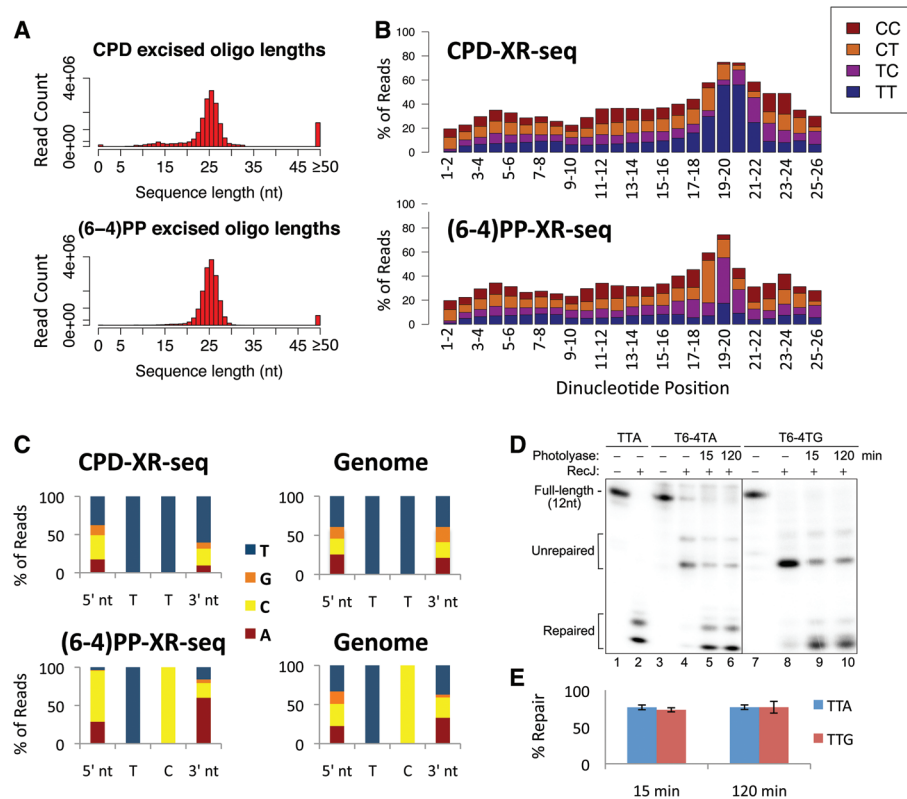
Figure 5. Single-nucleotide resolution of excision repair in NHF1 wild-type cells. (*A*) Distribution of excised oligonucleotide sequence lengths, calculated after removal of flanking adapter sequences from sequenced 50-nt reads. (*B*) Analysis of the frequency of each of the possible dipyrimidines along reads of 26-nt length shows enrichment 5–7 nt and 6–7 nt from the 3′ end for CPD XR-seq and (6-4)PP XR-seq, respectively. (*C*) Analysis of the nucleotides flanking the putative damaged pyrimidines at position 19–20 of the 26-nt-long excised fragments reveals sequence context preferences. Depicted are TT for CPD XR-seq and TC for (6-4)PP XR-seq. For comparison, the expected frequencies from an average of 50 random permutations of 26mers in the hg19 reference genomes are shown. (*D*) The oligonucleotides containing (6-4)PP were first repaired by photolyase and then digested by RecJ exonuclease. Only repaired or undamaged DNA could be completely degraded by RecJ, which is blocked by (6-4)PP. Locations of completely degraded products (repaired) and partially degraded products (unrepaired) are indicated by brackets. (*E*) Quantification of *D*. Values are the average of three independent experiments and are shown with SD.

by the transcription-coupled repair mechanism. Strikingly, in XP-C or CS-B mutant cells, in the absence of the competing mechanism, the differences between CPD and (6-4)PP patterns are lost, revealing that CPDs and (6-4)PPs can both be recognized and repaired by the available remaining pathway. Global repair of CPD is slightly higher surrounding the TSSs of genes, which could be due to higher levels of underlying damage in that region. In addition, the presence of TFIIH at the active promoters as part of the transcription machinery may facilitate damage recognition by general excision repair. XPC is indeed recruited to promoters in the absence of damage treatment (Drapkin et al. 1994; van Vuuren et al. 1994; Le May et al. 2010; Fong et al. 2013).

XR-seq of transcription-coupled repair at annotated genes and sites of divergent transcription

The results highlight the near exclusive activity of transcription-coupled repair on the template strand. This is consistent with the signal for transcription-coupled repair

being stalled RNA Pol II as it encounters the lesion (Selby et al. 1997; Hanawalt and Spivak 2008). We also observed a bias of repair toward the 5′ of genes, in agreement with the documented patterns of RNA Pol II enrichment near the TSS (Kim et al. 2005; Guenther et al. 2007; Gyenis et al. 2014). While the RNA-seq data that we used for our analysis were obtained from a different skin fibroblast cell line than the one used in our XR-seq experiments, they displayed high correlation to the transcription-coupled repair profiles that we observed.

Studies on divergent initiations from promoters have shown that transcription starts are sharp at the annotated TSSs (Core et al. 2014). For XR-seq, this pattern is observed for DNA repair on the template strand and leading into the annotated gene. However, on the nontemplate strand in XP-C cells, there is a more gradual rise in the levels of XR-seq signal starting ~1000 bp downstream from the annotated TSS. This could reflect unannotated alternative transcription initiation sites, potentially XPC- and transcription-independent repair due to the general accessibility, and the presence of DNA repair complexes

recruited by RNA Pol II on the nearby template strand (Le May et al. 2010; Fong et al. 2013).

XR-seq as a tool for discovery of novel transcripts

We observed sites of transcription-coupled repair in intergenic regions (~8%–9% of reads). Comparison with existing ENCODE histone modification data from NHDF cells as well as the chromHMM predictions from NHLF fibroblasts (Ernst and Kellis 2010) suggest that many of these sites, but not all, are sites of predicted eRNA transcription. The rest may be sites of eRNAs that are specific for our cell lines or to the UV response. As opposed to RNA-seq methods, which are limited by the stability of the RNA, in XR-seq, excised fragments from all transcribed regions in the genome are expected to have an equal stability. As a result, in a single assay under the conditions described here, XR-seq can capture all Pol II transcription events, including divergent transcription, and eRNAs.

Conclusion

We developed the XR-seq method, which maps excision repair genome-wide at single-nucleotide resolution. This method should prove useful in determining the effects of chromatin and other factors on nucleotide excision repair, for example, through treating cells with histone deacetylases or PARP inhibitors, which can sensitize the cells to chemotherapeutic drugs. The study of UV damage repair is important in understanding carcinogenesis by UV and UV-mimetic chemicals. However, our assay can also be applied to study the response of cancer cells to chemotherapy by performing XR-seq for excision of damages induced by chemotherapy such as platinum adducts. Genomic approaches are rapidly uncovering gene mutations that drive carcinogenesis and disease. XR-seq will aid in quantifying how DNA damage and repair efficiencies vary with respect to genomic position and chromatin status, information that will be valuable to incorporate into models of carcinogenesis, cancer risk, and genome stability.

Materials and methods

Cell lines

NHF1 cells, telomerase-immortalized normal human fibroblast monolayers derived from foreskin of a normal newborn, were obtained from W.K. Kaufmann (University of North Carolina, Chapel Hill) (Heffernan et al. 2002). XP-C (XP4PA-SV-EB, GM15983) and CS-B (CS1ANps3g2, GM16095) mutant human skin fibroblasts were purchased from the National Institute of General Medical Sciences Human Genetic Cell Repository (Coriell Institute). Mutant cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum at 37°C in a 5% CO_2 humidified chamber. The NHF1 cells were maintained under the same conditions with the addition of 2 mM glutamine.

Oligonucleotides and adaptors

The oligonucleotides for the ligation adaptors used were A5F (5′-GTTCAGAGTTCTACAGTCCGACGATC-3′), A5R (5′-NNNNNGATCGTCGGACTGTAGAACTCTGAAC-SpC3-3′), A3F (5′-phos-TGGAATTCTCGGGTGCCAAGG-SpC3-3′), and A3R (5′-CCTTGGCACCCGAGAATTCCANNNNN-SpC3-3′). NNNNN indicates five random nucleotides. A5R, A3R, and A3F were 3′-blocked by Spacer-C3, and A3F was also 5′-phosphorylated. These oligonucleotides were synthesized by IDT. To prepare the 5′ adaptor or 3′ adaptor, A5F and A5R or A3F and A3R were annealed, respectively. Ten nanomoles of A3F or A5F and 12 nmol of A3R or A5R were mixed together in 50 µL of hybridization buffer (10 mM Tris-HCl at pH7.5, 100 mM NaCl, 0.1 mM EDTA), incubated for 5 min at 95°C, and then slowly cooled down to 25°C.

The oligonucleotides for the PCR primers used were RP1 (5′-AATGATACGGCGACCACCGAGATCTACACGTTCAGAG TTCTACAGTCCGA-3′) and RPIn (5′-CAAGCAGAAGACGG CATACGAGATXXXXXXGTGACTGGAGTTCCTTGGCACCC GAGAATTCCA-3′). The underlined XXXXXX indicates different index sequences in accordance with those in the Illumina True-Seq small RNA kit. These oligonucleotides were synthesized by Sigma.

The oligonucleotides for the photoreaction and exonuclease digestion assay were UM (AGGAATTAAGGA), 6-4A (AGGAAT-TAAGGA), and 6-4G(AGGAATTGAGGA). The underlined TTs were (6-4)PP. These oligonucleotides were synthesized by the Synthetic Organic Chemistry Core Laboratory, University of Texas Medical Branch, Galveston, TX.

UV irradiation

UV irradiation was performed as described previously (Gaddameedhi et al. 2010). Briefly, cells were grown to ~80% confluence before UV irradiation. Culture medium was removed, and cells were then placed under a GE germicidal lamp emitting primarily 254-nm UV light (1 J/m^2/sec) connected with a digital timer for 20 sec (20 J/m^2 in total). Following irradiation, fresh culture medium was added to the cells, which were further incubated for the indicated time. Cells were washed with ice-cold PBS, harvested with a cell scraper in PBS, and collected by centrifugation at 2500 rpm for 4 min.

Purification and detection of excision products

Excision products containing either (6-4)PP or CPD were purified, radiolabeled, and subjected to electrophoresis as described previously (Hu et al. 2013). Briefly, low-molecular-weight DNA isolated by a modified Hirt method were subjected to immunoprecipitation with either anti-(6-4)PP (Cosmo Bio, NM-DND-002) or anti-CPD (Kamiya Biomedical, MC-062) antibody. Purified oligonucleotides were 3′ end-labeled by terminal deoxynucleotidyl transferase (New England Biolabs) and [α-^{32}P]-3′-deoxyadenosine 5′-triphosphate (cordycepin 5′-triphosphate, Perkin Elmer). After phenol-chloroform extraction and ethanol precipitation, labeled DNAs were resolved in 10% denaturing sequencing gels.

Photoreactivation and exonuclease digestion assay

Synthesized oligonucleotides with a single (6-4)PP were 3′ end-labeled as described above and purified through G25 gel filtration columns (GE Healthcare). About 0.1 pmol of radiolabeled substrates and 0.5 pmol of the same oligonucleotides unlabeled were incubated with 1.25 pmol of Escherichia coli (6-4)PP photolyase (Selby and Sancar 2012) in 50 µL of PLR buffer (50 mM Tris-HCl at pH7.5, 100 mM NaCl, 1 mM EDTA, 10 mM DTT) for 15 min or 2 h on ice with the irradiation of 18 mW/cm^2 366 nm light from two black-light bulbs (F15T8-BLB, General Electric) filtered

through one plate of glass. After phenol-chloroform extraction and ethanol precipitation, the oligonucleotides were incubated with 1 µL of RecJ$_f$ (New England Biolabs) for 1 h at 37°C in 10 µL of 1× NEB2 buffer. Two microliters of reaction mixture was mixed with 10 µL of formamide loading buffer and incubated for 5 min at 95°C before separation with 20% denaturing sequencing gels. The gels were quantified by PhosphorImager (GE Healthcare).

XR-seq assay

Cells were lysed, and immunoprecipitation was carried out with anti-TFIIH as described previously (Hu et al. 2013) with modification. Cell pellets from one 150-mm tissue culture dish were resuspended in 400 µL of ice cold buffer A (25 mM Hepes at pH 7.9, 100 mM KCl, 12 mM MgCl$_2$, 0.5 mM EDTA, 2 mM DTT, 12.5% glycerol, 0.5% NP-40) and incubated for 10 min on ice. Resuspended cells were transferred to an ice-cold Dounce homogenizer/tissue grinder and lysed on ice with 15 strokes using a tight plunger. The chromatin fraction was then pelleted by centrifugation for 30 min at 16,873g at 4°C. The supernatants were collected, and 2 µg of anti-p89 (XPB, Santa Cruz Biotechnology, sc293), 1 µg of anti-p62 (Santa Cruz Biotechnology, sc292), and 200 µg of RNase A (Sigma, R4642) were added. The reactions were rotated for 3–5 h at 4°C and then incubated overnight with 15 µL of recombinant protein A/G Plus-agarose (Santa Cruz Biotechnology, sc2003). The reaction could be scaled up (Supplemental Table 5). After washing three times with 1 mL of buffer A and once with 1 mL of buffer B (25 mM Hepes at pH 7.9, 100 mM KCl, 12 mM MgCl$_2$, 0.5 mM EDTA, 2 mM DTT, 12.5% glycerol, 1% NP-40), DNA was eluted from immunoprecipitates with 100 µL of buffer C (10 mM Tris-Cl at pH 7.5, 1 mM EDTA, 1% SDS) for 15 min at 65°C. The eluted DNA was then isolated by phenol-chloroform extraction followed by ethanol precipitation. DNA pellets were resuspended in 45 µL of water and then incubated with 5 µL of RNase A/T1 mixture (Thermo, EN0551) for 1 h at 37°C. After phenol-chloroform extraction, purification through G50 filtration columns (GE Healthcare), and ethanol precipitation, the DNA was then used for ligation.

To add double-stranded adaptors to both ends, purified excised oligomers were incubated with 40 pmol of 5′ adaptor and 100 pmol of 3′ adaptor in 10 µL of 2× hybridization buffer (20 mM Tris-HCl at pH7.5, 200 mM NaCl, 0.2 mM EDTA) for 10 min at 60°C and then for 5 min at 16°C in a thermal cycler. To perform ligation, 4 µL of 5× ligase buffer, 1 µL of T4 DNA ligase HC (Life, 15224-041), 1 µL of 50% PEG8000 (New England Biolabs), and 4 µL of H$_2$O were added to each reaction. The reactions were incubated overnight at 16°C. After phenol-chloroform extraction and ethanol precipitation, ligation products were subjected to immunoprecipitation with anti-CPD or anti-(6-4)PP antibodies as previously described (Hu et al. 2013). For NHF1 and CS-B samples, the ligation products were first immunoprecipitated with anti-CPD, and then the supernatant was subjected to immunoprecipitation with anti-(6-4)PP. For XP-C samples, the order was reversed: first immunoprecipitation with anti-(6-4)PP and then the flowthrough immunoprecipitation with anti-CPD. Because pilot experiments indicated enrichment of (6-4)PP in TCA and TTA contexts, we carried out immunoprecipitation experiments with synthetic substrates with (6-4)PP in TTA and TTG sequence contexts to ascertain that there was no preferential immunoprecipitation of (6-4)PP in a PyrPyrA context. We found that, under our experimental conditions, immunoprecipitation efficiencies for PyrPyrA and PyrPyrG (6-4)PP were 0.75 ± 0.04 and 0.77 ± 0.01, respectively.

Purified DNA were repaired by (6-4)PP photolyase or CPD photolyase as described previously (Selby and Sancar 2006). DNA containing (6-4)PP or CPD were incubated with 1.25 pmol of *E.*

coli (6-4)PP photolyase (Selby and Sancar 2012) or 20 pmol of *Drosophila melanogaster* CPD photolyase (Selby and Sancar 2006) in 50 µL of PLR buffer for 2 h on ice and irradiated with 18 mW/cm^2 366-nm light. One percent of unrepaired or repaired samples was used for a quality check. After phenol-chloroform extraction and ethanol precipitation, repaired DNA were PCR-amplified by Kapa Hifi HotStart ReadyMix with RP1 and RPIn (n means different index sequences, compatible with Illumina TrueSeq small RNA kit) for the indicated cycles in Supplemental Table 5. The PCR products were extracted with phenol-chloroform, precipitated with ethanol, and subjected to electrophoresis in a 10% native polyacrylamide gel in 1× TBE. Gel slices corresponding to 130- to 155-nt fragments were excised. The DNA was eluted in 0.3 M NaCl for 6 h at room temperature, precipitated with ethanol, and resuspended in buffer EB (10 mM Tris-HCl at pH 8.5). DNA concentration was determined by Pico Green. Libraries from all NHF1 samples (four samples) were pooled together and sequenced in one HiSeq 2000 lane (1 × 50), and libraries from all XP-C and CS-B samples (eight samples) were pooled together and sequenced in one HiSeq 2000 lane (1 × 50) by the University of North Carolina High-Throughput Sequencing Facility.

Sequencing and genome alignment

At least 17 million reads were obtained for each sample. Due to low amounts of starting material, especially for the CPD XR-seq samples, we observed relatively high levels of redundant reads that were eliminated from further analysis (18%–59%). Flanking adapter sequences were removed from the reads using trimmomatic (Bolger et al. 2014). Reads were aligned to the hg19 human genome using bowtie (Langmead et al. 2009) with the command options -q -nomaqround -phred33-quals -m 4 -n 2 -e 70 -l 20 -best -S. Following alignment, the files were split into "plus" and "minus" strands for subsequent analyses. We obtained at a total of at least 2.6 million mapped reads for each strand in each experiment type (Supplemental Table 5). The raw data and aligned data files are available on Gene Expression Omnibus (GEO), accession number GSE67941. For comparison of the DNA repair signal, we normalized all of the count data by the sequencing depth, and data are available for viewing on the University of California at Santa Cruz (UCSC) genome browser by pasting the link http://trackhubs.its.unc.edu/sancarlb/XRseq/hub.txt as the track hub URL in "My hubs."

ENCODE data

NHDF long total stranded RNA-seq (ENCODE DCC accession ENCSR00CUH), H3K27ac (accession ENCSR000APN), H3K9me3 (accession ENCSR000ARX), and DNase-seq (accession ENCSR000EMP) fastq, aligned reads .bam files, and peak files as well as the NHLF chromHMM chromatin state segmentation (UCSC accession wgEncodeEH000792) were downloaded from the ENCODE portal (http://genome.ucsc.edu/ENCODE) or viewed on the UCSC browser.

Genomic distribution of reads

The UCSC refGene.txt gene annotation (downloaded from the Illumina iGenome Web site) was used to obtain annotation of TSSs and TESs. We calculated overlap of aligned reads sequentially to gene bodies and 5 kb upstream of and 5 kb downstream from genes, with the remaining reads classified as intergenic. For comparison and to calculate *P*-values, we conducted the same analysis on 50 random data sets of 15 million 26mers from the hg19 human genome assembly.

Statistical analysis

Correlation between samples was calculated either genome-wide (counts over 3000-bp windows) or over gene exons. Total counts over exons for both RNA and XR-seq data were calculated with htseq-count (Anders et al. 2014), counting reads mapping to the coding or template strand, respectively. Spearman's correlations between sample count data were calculated and plotted using the R corrplot package. *P*-values for comparison of distributions were calculated as [number of times distribution of experimental and control data overlapped]/[total number of control data tests].

Plotting average XR-seq profiles

To calculate RNA levels, ENCODE RNA-seq reads were mapped to the hg19 genome using TopHat version 2.0.13 (Kim et al. 2013). We calculated FPKM for the two replicates using Cufflinks (Trapnell et al. 2010) and the UCSC hg19 genes.gtf. For average XR-seq profiles relative to the annotated TSS or TES, we limited the gene list to genes that do not have overlapping or neighboring genes for at least 6000 bp upstream or downstream on either strand. Genes were divided into 3586 nonexpressed and 1868 expressed, which were further divided into four quartiles based on their RNA expression level. Heat maps were generated using matrix2png (Pavlidis and Noble 2003). For enhancers, we obtained a list of 18,561 ENCODE DNase peaks that overlapped H3K27ac ChIP-seq peaks and chromHMM strong enhancer prediction segments but did not overlap genes or the 5-kb flanking regions. As a control, we used a set of 42,733 DNase peaks that did not overlap H3K27ac ChIP-seq peaks, chromHMM enhancer segments (strong or weak), or genes and their 5-kb flanking regions (Supplemental Table 6). Read counts were calculated from the aligned .bam files using BEDTools coverage and were normalized compared with the sequencing depth of NHF1 CPD XR-seq. For plotting average profiles along gene bodies, we used normalized base count wiggle files, the full USCS annotated gene list, and the CEAS package (Shin et al. 2009).

Sequence analysis of reads

Dinucleotide frequencies at each position of 26mer reads were calculated (Supplemental Table 3). The frequency of each of the 4 nt at the 5′ and 3′ positions was calculated for dipyrimidines at position 19–20 of the 26mer reads (Supplemental Table 4). For comparison and to calculate *P*-values for dinucleotide enrichment or depletion, we generated 50 random data sets of 15 million 26mers from the hg19 human genome assembly.

Acknowledgments

References

Adam S, Polo SE, Almouzni G. 2014. How to restore chromatin structure and function in response to DNA damage—let the chaperones play. Delivered on 9 July 2013 at the 38th FEBS Congress in St Petersburg, Russia. *FEBS J* **281**: 2315–2323.

Anders S, Pyl PT, Huber W. 2014. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Bryan DS, Ransom M, Adane B, York K, Hesselberth JR. 2014. High resolution mapping of modified DNA nucleobases using excision repair enzymes. *Genome Res* **24**: 1534–1542.

Choi JH, Gaddameedhi S, Kim SY, Hu J, Kemp MG, Sancar A. 2014. Highly specific and sensitive method for measuring nucleotide excision repair kinetics of ultraviolet photoproducts in human cells. *Nucleic Acids Res* **42**: e29.

Cleaver JE, Lam ET, Revet I. 2009. Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity. *Nat Rev Genet* **10**: 756–768.

Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**: 1845–1848.

Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**: 1311–1320.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci* **107**: 21931–21936.

Denissenko MF, Pao A, Tang M, Pfeifer GP. 1996. Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. *Science* **274**: 430–432.

Douki T, Cadet J. 2001. Individual determination of the yield of the main UV-induced dimeric pyrimidine photoproducts in DNA suggests a high mutagenicity of CC photolesions. *Biochemistry* **40**: 2495–2501.

Drapkin R, Reardon JT, Ansari A, Huang JC, Zawel L, Ahn K, Sancar A, Reinberg D. 1994. Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature* **368**: 769–772.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**: 817–825.

Fong YW, Cattoglio C, Tjian R. 2013. The intertwined roles of transcription and repair proteins. *Mol Cell* **52**: 291–302.

Gaddameedhi S, Kemp MG, Reardon JT, Shields JM, Smith-Roe SL, Kaufmann WK, Sancar A. 2010. Similar nucleotide excision repair capacity in melanocytes and melanoma cells. *Cancer Res* **70**: 4922–4930.

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**: 77–88.

Gyenis A, Umlauf D, Ujfaludi Z, Boros I, Ye T, Tora L. 2014. UVB induces a genome-wide acting negative regulatory mechanism that operates at the level of transcription initiation in human cells. *PLoS Genet* **10**: e1004483.

Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. 2013. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* **23**: 1210–1223.

Hanawalt PC, Spivak G. 2008. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* **9**: 958–970.

Heffernan TP, Simpson DA, Frank AR, Heinloth AN, Paules RS, Cordeiro-Stone M, Kaufmann WK. 2002. An ATR- and Chk1-dependent S checkpoint inhibits replicon initiation

following UVC-induced DNA damage. *Mol Cell Biol* **22**: 8552–8561.

Hu J, Choi JH, Gaddameedhi S, Kemp MG, Reardon JT, Sancar A. 2013. Nucleotide excision repair in human cells: fate of the excised oligonucleotide carrying DNA damage in vivo. *J Biol Chem* **288**: 20918–20926.

Huang JC, Svoboda DL, Reardon JT, Sancar A. 1992. Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5′ and the 6th phosphodiester bond 3′ to the photodimer. *Proc Natl Acad Sci* **89**: 3664–3668.

Kemp MG, Reardon JT, Lindsey-Boltz LA, Sancar A. 2012. Mechanism of release and fate of excised oligonucleotides during nucleotide excision repair. *J Biol Chem* **287**: 22889–22899.

Kemp MG, Gaddameedhi S, Choi JH, Hu J, Sancar A. 2014. DNA repair synthesis and ligation affect the processing of excised oligonucleotides generated by human nucleotide excision repair. *J Biol Chem* **289**: 26574–26583.

Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, Richmond TA, Wu Y, Green RD, Ren B. 2005. A high-resolution map of active promoters in the human genome. *Nature* **436**: 876–880.

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, et al. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**: 182–187.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**: R36.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Le May N, Mota-Fernandes D, Velez-Cruz R, Iltis I, Biard D, Egly JM. 2010. NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. *Mol Cell* **38**: 54–66.

Li S, Waters R, Smerdon MJ. 2000. Low- and high-resolution mapping of DNA damage at specific sites. *Methods* **22**: 170–179.

Li W, Selvam K, Ko T, Li S. 2014. Transcription bypass of DNA lesions enhances cell survival but attenuates transcription coupled DNA repair. *Nucleic Acids Res* **42**: 13242–13253.

Lindsey-Boltz LA, Sancar A. 2007. RNA polymerase: the most specific damage recognition protein in cellular responses to DNA damage? *Proc Natl Acad Sci* **104**: 13213–13214.

Mellon I, Spivak G, Hanawalt PC. 1987. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* **51**: 241–249.

Mitchell DL. 1988. The relative cytotoxicity of (6-4) photoproducts and cyclobutane dimers in mammalian cells. *Photochem Photobiol* **48**: 51–57.

Mitchell DL, Jen J, Cleaver JE. 1992. Sequence specificity of cyclobutane pyrimidine dimers in DNA treated with solar (ultraviolet B) radiation. *Nucleic Acids Res* **20**: 225–229.

Mu D, Park CH, Matsunaga T, Hsu DS, Reardon JT, Sancar A. 1995. Reconstitution of human DNA repair excision nuclease in a highly defined system. *J Biol Chem* **270**: 2415–2418.

Mu D, Hsu DS, Sancar A. 1996. Reaction mechanism of human DNA repair excision nuclease. *J Biol Chem* **271**: 8285–8294.

Mu D, Tursun M, Duckett DR, Drummond JT, Modrich P, Sancar A. 1997. Recognition and repair of compound DNA lesions (base damage and mismatch) by human mismatch repair and excision repair systems. *Mol Cell Biol* **17**: 760–769.

Pavlidis P, Noble WS. 2003. Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **19**: 295–296.

Pfeifer GP, Drouin R, Riggs AD, Holmquist GP. 1991. In vivo mapping of a DNA adduct at nucleotide resolution: detection of pyrimidine (6-4) pyrimidone photoproducts by ligation-mediated polymerase chain reaction. *Proc Natl Acad Sci* **88**: 1374–1378.

Powell JR, Bennett MR, Evans KE, Yu S, Webster RM, Waters R, Skinner N, Reed SH. 2015. 3D-DIP-Chip: a microarray-based method to measure genomic DNA damage. *Sci Rep* **5**: 7975.

Reardon JT, Sancar A. 2003. Recognition and repair of the cyclobutane thymine dimer, a major cause of skin cancers, by the human excision nuclease. *Genes Dev* **17**: 2539–2551.

Reardon JT, Sancar A. 2005. Nucleotide excision repair. *Prog Nucleic Acid Res Mol Biol* **79**: 183–235.

Sancar A. 1996. DNA excision repair. *Annu Rev Biochem* **65**: 43–81.

Selby CP, Sancar A. 1993. Molecular mechanism of transcription-repair coupling. *Science* **260**: 53–58.

Selby CP, Sancar A. 2006. A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity. *Proc Natl Acad Sci* **103**: 17696–17700.

Selby CP, Sancar A. 2012. The second chromophore in *Drosophila* photolyase/cryptochrome family photoreceptors. *Biochemistry* **51**: 167–171.

Selby CP, Drapkin R, Reinberg D, Sancar A. 1997. RNA polymerase II stalled at a thymine dimer: footprint and effect on excision repair. *Nucleic Acids Res* **25**: 787–793.

Shin H, Liu T, Manrai AK, Liu XS. 2009. CEAS: *cis*-regulatory element annotation system. *Bioinformatics* **25**: 2605–2606.

Sugasawa K, Ng JM, Masutani C, Iwai S, van der Spek PJ, Eker AP, Hanaoka F, Bootsma D, Hoeijmakers JH. 1998. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell* **2**: 223–232.

Svoboda DL, Taylor JS, Hearst JE, Sancar A. 1993. DNA repair by eukaryotic nucleotide excision nuclease. Removal of thymine dimer and psoralen monoadduct by HeLa cell-free extract and of thymine dimer by Xenopus laevis oocytes. *J Biol Chem* **268**: 1931–1936.

Teng Y, Bennett M, Evans KE, Zhuang-Jackson H, Higgs A, Reed SH, Waters R. 2011. A novel method for the genome-wide high resolution analysis of DNA damage. *Nucleic Acids Res* **39**: e10.

Tornaletti S, Pfeifer GP. 1994. Slow repair of pyrimidine dimers at p53 mutation hotspots in skin cancer. *Science* **263**: 1436–1438.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515.

van Hoffen A, Venema J, Meschini R, van Zeeland AA, Mullenders LH. 1995. Transcription-coupled repair removes both cyclobutane pyrimidine dimers and 6-4 photoproducts with equal efficiency and in a sequential way from transcribed DNA in xeroderma pigmentosum group C fibroblasts. *EMBO J* **14**: 360–367.

van Vuuren AJ, Vermeulen W, Ma L, Weeda G, Appeldoorn E, Jaspers NG, van der Eb AJ, Bootsma D, Hoeijmakers JH, Humbert S, et al. 1994. Correction of xeroderma pigmentosum repair defect by basal transcription factor BTF2 (TFIIH). *EMBO J* **13**: 1645–1653.

Venema J, Mullenders LH, Natarajan AT, van Zeeland AA, Mayne LV. 1990. The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. *Proc Natl Acad Sci* **87**: 4707–4711.

Venema J, van Hoffen A, Karcagi V, Natarajan AT, van Zeeland AA, Mullenders LH. 1991. Xeroderma pigmentosum complementation group C cells remove pyrimidine dimers selectively from the transcribed strand of active genes. *Mol Cell Biol* **11**: 4128–4134.

Wakasugi M, Sancar A. 1998. Assembly, subunit composition, and footprint of human DNA repair excision nuclease. *Proc Natl Acad Sci* **95**: 6669–6674.

Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, et al. 2011. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**: 390–394.

Wood RD. 1997. Nucleotide excision repair in mammalian cells. *J Biol Chem* **272**: 23465–23468.

Zavala AG, Morris RT, Wyrick JJ, Smerdon MJ. 2014. High-resolution characterization of CPD hotspot formation in human fibroblasts. *Nucleic Acids Res* **42**: 893–905.

Zentner GE, Tesar PJ, Scacheri PC. 2011. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* **21**: 1273–1283.