Continuous transcription initiation guarantees robust repair of all transcribed genes and regulatory regions

Anastasios Liakos1,2,4, Dimitris Konstantopoulos1,3,4, Matthieu D. Lavigne*1& Maria Fousteri*1

Inhibition of transcription caused by DNA damage-impaired RNA polymerase II (Pol II) elongation conceals a local increase in de novo transcription, slowly progressing from Transcription Start Sites (TSSs) to gene ends. Although associated with accelerated repair of Pol II-encountered lesions and limited mutagenesis, it is still unclear how this mechanism is maintained during genotoxic stress-recovery. Here we uncover a widespread gain in chromatin accessibility and preservation of the active H3K27ac mark after UV-irradiation. The concomitant increase in Pol II escape from promoter-proximal pause (PPP) sites of most active genes, PROMPTs and enhancer RNAs favors unrestrained initiation, as evidenced by the synthesis of nascent RNAs including start RNAs. Accordingly, drug-inhibition of PPP-release replenishes levels of pre-initiating Pol II at TSSs after UV. Our data show that such continuous engagement of Pol II molecules ensures maximal transcription-driven repair throughout expressed genes and regulatory loci. Importantly, revealing this unanticipated regulatory layer of UV-response provides physiological relevant traction to the emerging concept that Pol II initiation rate is determined by pause-release dynamics.
Pol II initiation at TSSs and release into productive elongation from PPP are ubiquitous and crucial steps regulating transcription of protein-coding genes and long non-coding RNAs\(^1,2\) (together called mRNAs in this paper). Similarly, Pol II is regulated for transcribing regulatory non-coding regions: enhancer RNAs (eRNAs) are expressed bidirectionally from eTSSs\(^3,4\), while PROMoter uPstream transcripts (or upstream antisense RNAs, collectively called PROMPTs) here are produced in the opposite direction to mRNA when two stable transcripts are not initiated in very close proximity and in opposite directions (bidirectional promoters)\(^5\). However, contrary to mRNAs, eRNAs and PROMPTs are short and unstable due to high early termination rates and increased susceptibility to degradation by the RNA exosome\(^6,7\), thus making their detection technically challenging.

Initiation of transcription in all the above regions depends on the efficient assembly of the pre-initiation complex (PIC) upstream of TSS and serine 5 phosphorylation (SSP) of Pol II C-terminal domain (CTD)\(^8\). After elongation of \(~30–60\) nucleotides of initiation-associated RNAs (start-RNAs)\(^9,10\), Pol II is paused at PPP sites by negative elongation factors DSIF and NELF\(^11,12\). Signal-regulated phosphorylation of these factors and of Pol II CTD serine 2 (S2P) by P-TEFb is required for productive elongation\(^11,13\). It recently emerged that, if this step does not occur rapidly, start-RNAs are terminated\(^12,14\), implying that Pol II turnover at PPP sites is high at steady state, and that replenishment of Pol II engaged in early transcription (initiation to PPP) is achieved through the continuous re-entry of pre-initiating Pol II into PICs\(^14,15\).

The integrity of the genetic information encoded in DNA sequence is persistently challenged by a variety of genotoxic perturbations\(^16\). A plethora of DNA damage response (DDR) mechanisms have evolved to guarantee the detection and removal of different types of DNA lesions, limiting the probability of mutagenesis by adjusting to the cell’s status and need for efficient recovery from DNA damage\(^16–18\). Nucleotide excision repair (NER) plays a vital role in sensing and removing a large panel of helix-distorting DNA adducts such as cyclobutane pyrimidine dimers (CPDs) induced by ultraviolet (UV) light, and benzo[a]pyrene guanine adducts induced by cigarette smoke\(^17\). Transcription-coupled NER (TC-NER) is promptly triggered by elongating Pol II molecules encountering DNA adducts and speeds up excision and repair in expressed loci\(^19–21\). In comparison, the second NER subpathway, Global Genome-NER (GG-NER), operates through the entire genome, but recognizes more stochastically helix distortions\(^17,21–23\). Importantly, given all the classes of transcripts defined above, it is estimated that the coverage of transcribed regions\(^24\) potentially scanned by TC-NER expands to more than 50% of the genome, thus qualifying transcription as a major driving force in safeguarding genomic stability.

Although TC-NER depends on lesion-sensing potential by elongating Pol II molecules, transcription elongation has been shown to be transiently inhibited after UV irradiation\(^25–27\) due to a proportion of Pol II molecules stalling at encountered DNA damages\(^25,28\). Moreover, depletion of the pre-initiating hypophosphorylated Pol II(hypo) isoform from chromatin shortly after UV irradiation\(^25,29,30\) has led to the assumption that new transcription initiation events are transiently and globally repressed\(^21,29–33\). On the other hand, recent reports\(^25,26,34\) have revealed a functionally essential stress-dependent increase in 5’ nascent RNA (nRNA) activity globally that depends on the UV-induced increase in active P-TEFb levels\(^35,36\), and on the rapid dissociation of the NELF complex\(^37\). The ensuing fast and global release of de novo Pol II elongation waves from PPP sites into gene bodies boosts lesion-sensing activity and accelerates removal of DNA adducts by TC-NER in virtually all active mRNA genes\(^25\). Together, these findings substantiate the possibility that initiation of transcription might not be as severely affected by UV as previously believed.

Taking also into consideration recent evidence that supports the model of disengagement of a given Pol II molecule from DNA template after damage recognition\(^21,31,38\), it is tempting to assume that ensuring continuity in transcription initiation may bring advantages in the repair process. We thus hypothesized that the apparent loss of pre-initiating RNAII may not be due to the absence of RNAII recruitment at TSSs, but rather due to a decrease in the dwell time of Pol II-hypo isoform at TSS, as justified by the concomitant increase in Pol II-ser5P and Pol II-ser2P downstream of TSS\(^25\). In this way, cells would be able to uninterruptedly feed the global release of scanning enzymes into transcribable sequences, and guarantee the detection of more lesions along genes’ template strand (TS). Herein, we decipher chromatin dynamics genome-wide upon UV damage, and find a significant gain in accessibility (ATAC-seq) at the TSSs of actively transcribed regions controlling mRNAs, PROMPTs, and eRNA expression. This phenomenon is underlined by the maintenance of active histone marks (H3K27ac), the lack of deposition of transcriptional silencing modifications (H3K27me3) at transcribed loci, and is correlated with the influx of Pol II into productive elongation. The paradoxical decrease in pre-initiating Pol II-hypo at these TSSs upon UV is elucidated by revealing that the presence of Pol II-hypo is rescued when PPP release is drug-inhibited. Accordingly, preserved production of start-RNAs after UV stress lies under the increased production of nRNA, and is prevented only after inhibition of transcription initiation. The identified genome-wide dependence of initiation rate on promoter-proximal pause-release dynamics explains the seamless recruitment/initiation of Pol II upon UV, in turn enabling efficient repair of the totality of the sequences encoding active regulatory regions and mRNAs.

**Results**

**Chromatin accessibility increases at active regulatory regions upon UV.** To characterize the impact that UV irradiation might have on the chromatin landscape of transcriptional regulatory regions during the early recovery times (from 0.5 to 4 h), and how this could be linked to the widespread PPP release of elongating Pol II and the local increase in mRNA production downstream of TSS\(^25,26,34\), we first determined the genome-wide changes in chromatin accessibility. The Omni-ATAC-seq protocol\(^39\) was implemented in our system involving UV-C irradiation (15 J/m\(^2\)) of human skin fibroblasts synchronized in G1 (see “Methods” and also ref.\(^25\)). We reproducibly measured chromatin accessibility before (NO UV) and 2 h after (+UV) irradiation (Supplementary Fig. 1a, b), performed peak calling and mapped a total of 106,052 accessible regions (ARs) across conditions (Supplementary Fig. 1b), and combined the replicates (see “Methods” for details).

ARs were enriched at promoters and intragenic or intergenic regions with transcriptional regulatory function (TSSs, TSS flanks, and enhancers according to ChromHMM annotation, Fig. 1a–c; Supplementary Fig. 1c–e, “Methods”). Interestingly, we reveal a widespread increase (up to 1.71 average fold change (FC)) in chromatin accessibility after stress at 97.9% of promoter-, 94.6% of intragenic-, and 94.4% of intergenic ARs (Fig. 1b–d; Supplementary Fig. 1e, f).

We then selected differentially accessible regions (DARs) by applying stringent thresholds both in terms of FC (Log2 FC > 1) and P value (P < 0.001), and found that 6410 loci showed particularly increased chromatin accessibility upon UV (DAR-gain) (Fig. 1e, top panel). DAR-gain found at promoter regions
Fig. 1 Increase in chromatin accessibility in response to mild doses of UV irradiation. a Classification of ARs according to ChromHMM annotation. The dashed line represents active regulatory loci. b (Left panel) Heatmap of ATAC-seq reads in genomic regions 1kb around ATAC-seq peak centers before (NO UV) and after UV (+UV, +2 h; 15 J/m²), categorized according to their genomic position relative to RefSeq genes (intergenic, intragenic, and promoter peaks) and sorted by increasing read density (as determined before UV). (Right panel) Heatmap showing the log2 fold change (log2 FC) between +UV and NO UV read densities calculated in genomic regions 1kb around ATAC-seq peak centers. c Average profile plots of ATAC-seq read densities of non-irradiated (solid line) and irradiated (dashed line) cells in intergenic (red), intragenic (green), and promoter (blue) regions. d Heat-density scatter plots comparing ATAC-seq read density before and after UV at all accessible regions (ARs) in intergenic, intragenic, and promoter regions, respectively. e (Upper panel) Volcano plot representing differentially accessible regions (DARs) between irradiated and non-irradiated cells. Regions with significantly increased (DAR-gain) or decreased (DAR-loss) accessibility are depicted in red and green, respectively. (Bottom panel) Proportion of DAR-gain loci in intergenic, intragenic, and promoter, and active promoters.

represented 13.3% of all promoter ARs (Fig. 1e, lower panel), thus pinpointing towards a potentially functionally relevant chromatin opening at TSS regions. DAR-gain located at intragenic and intergenic loci (Fig. 1e) were linked to genes if they overlapped functional enhancers defined in FANTOM5 (see “Methods”). We found that genes associated with DAR-gain loci (either identified on their promoter or enhancers) were representative (adjusted (adj.) P < 0.05) of a number of biological pathways previously associated with DDR processes, including cellular response to stress, DNA repair, transcription regulation by TP53, and cell cycle checkpoints (Supplementary Fig. 2). In addition, we identified a broad range of many other significant Gene Ontology (GO) categories (163 in total, Supplementary Fig. 2), a result in line with the previously reported global PPP release of elongating Pol II waves at all active gene bodies upon UV irradiation.

Chromatin marks linked to transcription remain stable after UV. A number of studies have demonstrated that the turnover, modification, and/or degradation of histones around damage sites represent essential steps in conserved pathways that help cells deal with genotoxic stress. However, especially in the case of UV-C-induced DNA damage, little is known about the post-translational modifications (PTMs) of histones around transcriptional regulatory regions. To better interpret the increase in chromatin accessibility and clarify its possible impact on genome-wide transcription dynamics, we studied the differential presence of two histone PTMs representative for the transcription status of associated chromatin: the silencing mark H3K27me3 and the activation mark H3K27ac.

We conducted ChIP-seq experiments with antibodies specific for both H3K27ac and H3K27me3 in NO UV and +UV (+2 h) conditions, and focused our analysis on TSSs of mRNAs and on a robust set of eTSSs, which are known to be functional and potentially transcribed in the investigated cell type according to the FANTOM5 database (see “Methods”). We used the ChIP-seq data generated in this study (H3K27ac and H3K27me3), as well as previously published ChIP-seq data (Pol II-ser2P) from the steady state (NO UV) condition, to determine subsets of active (presence of H3K27ac and Pol II-ser2P peaks over TSS), repressed (presence of H3K27me3 peaks over TSS), and inactive loci (no peak detected over TSS for H3K27ac, H3K27me3, and Pol II) (Fig. 2a, see “Methods”) in our cell system. We associated the changes in histone marks and Pol II observed in these regions upon UV with ATAC-seq results. The increase in chromatin accessibility was detected at all active TSSs, which correspond largely to the promoters identified above (compare Figs. 1a and 2a, and see “Methods”), as well as FANTOM5-annotated active eTSSs upon UV (Fig. 2b, ATAC, 95% confidence interval (CI) excludes 0). This opening was in sharp contrast to the UV-induced global loss of Pol II-hypo at TSSs and eTSSs (Fig. 2b, Pol II-hypo, 95% CI excludes 0) observed 1.5 h after UV irradiation (8 J/m²) (Fig. 2a, b; Supplementary Fig. 3).

Strikingly, we also found preservation (slight but not significant increase) of H3K27ac levels (Fig. 2a, b, 95% CI includes 0; Supplementary Fig. 3), and we observed no exchange of H3K27ac for H3K27me3 in response to UV at these active TSSs and eTSSs. Reciprocally, there was no loss of H3K27me3 for H3K27ac, and no gain of Pol II at repressed loci (Fig. 2a; Supplementary Fig. 3). Accordingly, the results of our genome-wide analysis were consistent with biochemical evidence obtained by histone acetic extraction followed by western blot analysis, showing that the global levels of H3K27me3 or H3K27ac remain fairly stable during the early period of recovery from UV stress (15 J/m²) (Supplementary Fig. 3c, d).

We therefore conclude that depletion of detectable Pol II-hypo at TSSs and eTSSs does not occur due to repression of these loci by trimethylation of H3K27me3, or because of the loss of the activating histone mark H3K27ac.
Chromatin opening parallels Pol II transition into elongation. To elucidate the functional advantage associated with increased chromatin accessibility in response to UV, we performed a thorough integrative analysis of our data in relation to previously published data sets (Pol II-ser2P from ref.25 and Cap Analysis Gene Expression (CAGE)-seq from ref.4, see "Methods"). First, we customized a genome annotation, which unambiguously pinpoints to the TSSs of mRNAs, PROMPTs, and eRNAs that do not overlap with regions possibly being transcribed through from neighboring/overlapping genes, promoters, or enhancers (see "Methods"). We then established three categories (Fig.3a–c), as per previously suggested models44: first, active bidirectional promoter regions, which include the TSSs of mRNA–mRNA pairs transcribed in opposite directions (Fig. 3a); second, active unidirectional promoters, which include the TSS of a given mRNA gene (+ or −) for which we could associate an expressed PROMPT in the antisense direction (Fig. 3b); third, active intergenic—as opposed to intragenic—enhancers to avoid potential contamination by interfering reads that derive from overlapping transcription of other active elements (Fig. 3c). Importantly, PROMPT and enhancer transcriptional activity were defined from available CAGE data for the skin and dermal fibroblasts (FANTOM5 consortium, see "Methods") that accurately determine transcript starting position (5′ end), abundance, and directionality of Pol II transcription in our model (Fig.3a–c, CAGE). TSS loci were sorted by inter-TSS distance, which we defined as the distance separating TSSs and/or the summits of CAGE signals detected on the (+) and (−) strands (Fig. 3a, b; "Methods"). This allowed us to identify regions with overlapping (convergent, CONV) or non-overlapping (divergent, DIV)
transcription (Fig. 3a, b). By focusing on the latter category, we could study the dynamics of transcription at play only in each direction, without having to deal with potential interferences from overlapping regions.

Using this setup, we discovered that the UV-dependent increase in chromatin accessibility (Fig. 3a–c, ATAC) was paralleled to the transition of Pol II into active elongation (Fig. 3a–c, Pol II-ser2P), not only at TSSs but also at adjacent PROMPTs and eRNA sequences (Fig. 3e, f), as shown by the loss in Pol II reads at TSSs and the gain of reads in downstream regions. These results were confirmed quantitatively by showing that escape index (EI) of elongating Pol II increased in the +UV condition in comparison with NO UV for 90.1% of bidirectional promoters (Fig. 3g, Chi-square test \( P = 5.1 \times 10^{-266} \)), as well as for 70.1% of PROMPTs (Fig. 3h, Chi-square test \( P = 4.5 \times 10^{-89} \)) and 68.6% of eRNAs (Fig. 3i, Chi-square test \( P = 2.5 \times 10^{-444} \)). We conclude that the PPP release of Pol II upon genotoxic stress is synchronously triggered at all active transcription units, and coincides with increased chromatin breathing. These data extend the previously characterized transcription-driven genome surveillance mechanism to essentially all active gene-regulatory regions, and give mechanistic insights into the synergy between the increase in chromatin accessibility and the transcriptional response observed upon UV.

**Fig. 3** Release of Pol II from pausing sites in coding and non-coding transcribed regions upon UV stress. a (Left) Scheme of convergent (“CONV”, overlapping) and divergent (“DIV”, non-overlapping) active bidirectional promoters expressing two mRNAs (blue arrows). (Right) Heatmap showing the distribution of CAGE (black; – and + strand separately) in the steady state, and ATAC-seq (blue) and Pol II-ser2P (purple) read densities, before (NO UV) and 2 h after UV (+UV) around +strand TSS (+2 kb). Loci are sorted by inter-TSS distance ((–strand TSS)–(+strand TSS)). Data for Pol II-ser2P are obtained from ref. 25 and CAGE data from ref. 4. Divergent loci correspond to inter-TSS >100 bp. b Same as in a but for active unidirectional genes TSSs, where PROMPTs (orange arrow) are transcribed in the opposite direction to the mRNA gene, from either the – strand (upper panel) or the + strand (bottom panel). Straight dashed lines indicate the position of mRNA TSS, and the sigmoidal dashed line indicates the variable relative position of CAGE PROMPT. Loci are sorted by inter-TSS distance (mRNA TSS - CAGE PROMPT) from the most convergent (mRNA and PROMPT overlapping, inter-TSS <100 bp) to the most divergent (mRNA and PROMPT non-overlapping, inter-TSS >100 bp) loci. c Same as in a, but for active intergenic enhancers expressing eRNAs (yellow arrow) in the opposite direction. d, e Average plots of Pol II-ser2P before (solid line) and after UV (dashed line) on divergent categories defined in a, b. Insets represent a zoomed view. f Same as in d, e, but for all active intergenic enhancers. g, i Comparison of escape index (EI), for indicated categories of d–f. Percentages of loci with increased EI after UV (dark purple) are shown. Chi-square test between active bidirectional and inactive unidirectional genes (g), and between active and inactive enhancers (i) was performed to determine whether the observed number of genes with \( \Delta EI > 1 \) (number of genes with EI after irradiation greater than EI before irradiation) differs from the expected value purely by chance (see “Methods” for details).
TSSs, PROMPTs, and eTSSs are hardly compatible with the previously suggested model of UV-induced global inhibition of transcription initiation. We thus searched for alternative reasons that could explain reduction of Pol II-hypo levels at active TSSs/eTSSs despite increased accessibility after UV.

We performed a set of experiments aiming to determine whether Pol II was actually recruited to TSSs upon UV (Pol II-hypo isoform used as a proxy). First, as depicted in Fig. 4a, we irradiated cells with a UV (15 J/m²) and left them to recover for 2 h, when the levels of Pol II-hypo are known to be severely depleted (Supplementary Fig. 4a, b (DMSO NO UV vs DMSO + UV +2 h)). We then applied, or not, an inhibitor of Pol II release into elongation from PPP sites (DRB, see "Methods"). Cells were crosslinked (X) 2 h after the addition of DRB (or DMSO for the control cells). In accordance with this, we observed an increase in Pol II-hypo levels at TSSs/eTSSs in the presence of DRB, as shown in the Western blot analysis (Fig. 4b). The quantification of Pol II-hypo levels, as compared with the NO UV/+DRB condition, showed a significant increase (P = 0.0055, n = 3).

Fig. 4 DRB rescues pre-initiating Pol II levels following DNA damage. a Experimental timeline showing times of UV irradiation (15 J/m²) and DRB treatment (see "Methods"). b Western blot analysis of chromatin extracts for Pol II-hypo levels as examined after employing the experimental strategy described in a. Lamin B1 and histone H3 were used as loading controls. Bar graph represents quantification of Pol II-hypo levels as compared with the NO UV/+DRB condition. Data shown reflect three independent experiments. Error bars represent S.E.M., and P values are calculated using two-sided Student’s t test. c Heatmap of Pol II-hypo ChIP-seq read densities in genomic regions 2 kb around TSS for categories defined in Fig. 3a-c after performing the combination of UV/DRB treatments described in a. d Average profile plots of read densities analyzed in c.
with the above mentioned previous reports, in cells that were crosslinked 2 h after UV irradiation in the absence of DRB (+UV/X 2 h), or in cells that were crosslinked 4 h after UV irradiation, and had been incubated with DMSO for the last 2 h (+UV/−DRB/ X 4 h), we detected only minimal levels of pre-initiating Pol II in total chromatin extracts or at TSSs, PROMPTs, and eTSSs, as revealed by Western blot analysis (Fig. 4b) and ChiP-seq (Fig. 4c, d), respectively. In contrast, when cells had been incubated with DRB for the last 2 h before being crosslinked at 4 h after UV irradiation (+UV/+DRB/X 4 h), we observed a significant rescue of pre-initiating Pol II (hypo) levels in total chromatin (Fig. 4b, two-sided Student’s t test P = 0.0055 compared with “+UV/−DRB/ X 4 h” and P = 0.0156 compared with “+UV/X 2 h”). The restoration of pre-initiating Pol II levels was even more pronounced when we focused on the occupancy on active TSSs, PROMPTs, and eTSSs, where average read densities detected by Pol II-hypo ChiP-seq after DRB treatment (+UV/+DRB/X 4 h) matched the control NO UV levels (NO UV/+DRB/ X 4 h) (Fig. 4c, d). Therefore, even by blocking the stress-triggered transition of Pol II molecules from PPP sites into elongation at 2 h post UV, when the prior-to-UV Pol II-hypo levels were almost completely depleted, we were able to reveal the underlying continuous de novo recruitment of Pol II-hypo molecules in PICs.

We also applied DRB just before and for 2 h after UV (15 J/m²), Supplementary Fig. 4a), and found a limited loss of pre-initiating Pol II in chromatin extracts upon UV (Supplementary Fig. 4b, c, two-sided Student’s t test P = 0.0145). This result was corroborated by ChiP-qPCR experiments (performed on the same chromatin extracts used above), as DRB prevented the UV-induced reduction in occupancy of Pol II-hypo at promoter/TSS-proximal regions of six active genes (Supplementary Fig. 4d, 2nd Student’s t test P = 0.003 for DMSO, while P = 0.2876 (non-significant) for DRB).

We thus conclude that the genome-wide UV-induced PPP release of Pol II molecules into elongation accelerates the transition into initiation of the next-to-be-recruited Pol II-hypo molecules, limiting the dwell time of this isoform at essentially all active TSSs, PROMPTs, and eTSSs.

Increased RNA synthesis from active TSSs upon UV. As UV irradiation does not inhibit neither the recruitment of Pol II-hypo into PICs nor Pol II escape into elongation, we next verified that newly synthesized RNA molecules were detected at the beginning of transcribed regions. We took advantage of our and other mRNA-seq data25-27, and we examined if the previously characterized global increase in EU- or Bru-labeled RNA reads at the beginning of genes (see Supplementary Fig. 4 in ref. 25) could originate from increasedPolII molecules20) and CAGE4 accurately locate active TSSs, transcript orientation. Upon correlation with CAGE, we found that onset of TC-NER coincided with CAGE reads location, confirming the fact that TC-NER (triggered by damage-arrested Pol II molecules20) and CAGE4 accurately locate active TSSs (Fig. 6b compare with Fig. 3a–c, left), Fig. 6c, d). As expected, repair efficiency was equal in each direction for bidirectional active promoters (Fig. 6b–c). This result was also in line with Pol II-hypo ChiP-seq data showing equivalent amounts of Pol II recruitment at PICs in both directions (Supplementary Fig. 6a, b), and CAGE data indicating strand-balanced production of capped mRNAs (Fig. 6e, CAGE, boxes centered around Log2 FC = 0).

Nevertheless, we note that the variability between both directions was strikingly less for TC-NER and Pol II-hypo than for CAGE (Fig. 6e (proportion of non-significant F-Tests: P = 0)), Supplementary Fig. 6b (top panel, proportion of non-significant F-Tests: P = 0)).

Next, we further investigated repair of PROMPTs and enhancers, a phenomenon previously observed, but hardly explained.24,46 We quantified strand-specific repair upstream and downstream of unidirectional promoters, and found that repair activity at unambiguously resolved divergent PROMPTs was stronger than expected from CAGE levels (Fig. 6b–d, DIV). Indeed, XR-seq read density was not correlated to the steady state levels of CAGE at those loci (Pearson correlation coefficient (PCC) = 0.1343). Also, FCs of TC-NER reads between mRNAs and PROMPTs were much smaller than those for CAGE (Fig. 6f, 95% CI excludes 0), thus matching the UV-independent Pol II-hypo uniformity (Supplementary Fig. 6a, b). Similarly, TC-NER levels on TS of eRNAs were higher than anticipated. Indeed, the densities of eRNA
XR-seq reads were similar to those of mRNAs (Fig. 6b–d), and contrasted with the very low CAGE signal detected at these loci (Fig. 3a–c, left). Therefore, balanced Pol II-hypo loading in PICs at all classes of transcripts, during steady state or upon stress, allows for equal initiation events and mirrors the homogeneous levels of XR-seq detected in these regions. Taken together, our results demonstrate that the widespread continual initiation and release into productive elongation of Pol II waves maximize repair activity.
regardless of prior-to-UV transcript expression level at all kinds of active regulatory regions (mRNAs, PROMPTs, and enhancers).

**Continuous initiation drives TC-NER to completion.** We next assessed the biological relevance of continuous transcription initiation from active regulatory regions during the UV-recovery period. We have reported previously that in the absence of a UV-triggered PPP release of elongating Pol II waves, the scanning activity of pri-elongating (e.g., already elongating prior to UV) Pol II molecules is not sufficient to enable recognition of damaged thymidine dimers (TpT sequence, abbreviated as TT) in the totality of the transcribed genome. In accordance, a recent study reports that inhibition of transcription elongation with DRB limits the extent of excision activity detected by CPD XR-seq. Thus, sending Pol II molecules from TSS into gene bodies after UV is of pivotal importance to maximize the detection of TTs on active TS.

We quantified the effect of various regimes of DRB (see scheme in Fig. 7a) on the dynamics and distribution of TC-NER (CPD XR-seq data from XP-C cells) around TTs and around TSSs of mRNA genes, PROMPTs, and enhancers (Fig. 7a–d, respectively, see “Methods” and ref. 23). We observe that when DRB was applied immediately after UV, repair of TTs located from TSS to upstream of the +1-h wave front (WF) of UV-released Pol II (as obtained in the non-DRB control) are seriously affected (Fig. 7a–c, DRB + UV + 1 h, clusters 0-II).
Fig. 6 TC-NER is homogeneous at all transcribed regions. a Scheme representing the orientation and nomenclature of the DNA strands for TC-NER-specific XR-seq analysis. b Heatmaps of re-analyzed TC-NER excision reads of GG-NER-deficient cells (XP-C), which are detected on template strand (TS) (+) or — (red) strand of the genome, 2 kb around TSSs for categories defined in Fig. 3a–c (“Methods”). Dotted gray lines denote the CAGE summits detected on each strand; the dark-blue dotted lines indicate positions 500 bases downstream of CAGE summits for the corresponding strand. c UCSC genome browser snapshots of representative loci for categories defined in b. d Average profiles of read densities derived from b only; the divergent (DIV) loci were considered. e (Left panel) Scheme representing the range used for calculating log2 FC of reads between + strand and —strand at divergent loci. (Right panel) Boxplots showing the ratio of reads between directions (indicated window sizes and borders) at bidirectional promoters for CAGE reads (shown in Fig. 3), and TC-NER-specific XR-seq reads shown in b. Boxplots show the 25th-75th percentiles, and error bars depict data range to the larger/smaller value no more than 1.5 × IQR (interquartile range, or distance between the first and third quartiles). Two-sample F-tests were conducted for each of 10,000 sampling pairs of 100 data points with replacement from each population to test for significant differences between sample variance. The calculated P expresses the percentage of the non-significant F-tests (F-test P > 0.05) out of all tests. f Comparison of CAGE and TC-NER-specific XR-seq reads as in e, but between divergent non-overlapping mRNA and PROMPTs (indicated window sizes and borders). In all, 95% confidence intervals (CI) of mean differences between log2 counts of tested conditions were calculated for 10,000 samplings of 100 data points with replacement from each population. Effect sizes of log2 counts between data sets were calculated using Cohen’s method (CES).

Fig. 7 TC-NER activity monitored with various regimes of DRB reveals the need for continuous transcription initiation. a Heatmap depicting the distribution of excised DNA fragments (XR-seq) derived from TC-NER activity (XP-C cells). Reads were aligned around TT loci of the transcribed strand, localized on active genes. Data were obtained from ref. 38. The box in blue border (left) illustrates the experimental timeline followed and the respective drug treatments. b Average plots of read densities (Rd) showed in a but for clusters upstream (0, I, and II) and downstream (III, IV, V, and VI) of Pol II-ser2P wave front (WF) as defined in ref. 23 for +UV (–1 h) condition. c Plot showing the percentage (%) of excised TT loci, as calculated by the difference between Rd at summit (S) and Rd at flanks (F) (%F score) from all clusters presented and analyzed in a and b. WF is indicated, and asterisk indicates where drastic decreases in the XR-seq signal occur in DRB2 experiments. d Average profile of read densities of XR-seq signal derived from XP-C cells. Dashed lines illustrate the XR-seq signal in + (blue) or — (red) strand, in the presence of DRB (see “Methods” for details).

Accordingly, allowing only a limited number of pri-elongating molecules to be launched just before the irradiation, while preventing de novo Pol II release after stress (Fig. 7a–c, DRB2 + UV + 1 h) highlights how an insufficient feed in Pol II impairs TC-NER activity at both distal and proximal transcribed regions (compare signal before and after asterisk positions in Fig. 7a, c).

We reasoned that the high extent of ongoing repair activity detected without DRB (UV + DMSO) is the result of the continuity in Pol II initiation. To confirm this directly, we mapped XR-seq reads around TTS located on TS 1, 4, and 8 h after UV in WT cells (data obtained from ref. 46, see “Methods”). We found that significant levels of transcription-dependent
excision activity were maintained at lesions located directly downstream of active genes TSSs at 4 and 8 h into the recovery process (compare Fig. 8a–c and Supplementary Fig. 7a–c). Notably, it also appeared that the bulk of the excision activity on TS of active genes shifts overtime (+8 h) from the proximal to the distal part of long genes (Fig. 8a–c, clusters III–VI; Supplementary Fig. 7d).

Overall, these results reveal that a large extent of the transcription-driven repair activity detected after UV is due to the ongoing entry of Pol II molecules at TSSs, which can scan farther and farther lesions along the gene bodies. Our analysis highlights the advantage of a continuously supplied transcription-dependent repair process (Fig. 8a and Supplementary Fig. 7d, TS of active genes) over slower CPD detection capabilities of GG-NER, which was detected at significantly lower levels across all regions on NTS of active genes and on both strands of inactive genes (Supplementary Fig. 7d).

**Discussion**

In this study, we provide quantitative insights into the molecular processes underlying the major transcription-coordinated cellular response that is activated in human cells upon genotoxic stress. The establishment of precise maps of chromatin state helped us to query in detail the impact of transcription on DNA repair activities at important functional regions, including PROMPT and eRNA loci. Our results support a model of continuous transcription initiation that can feed the widespread...
UV-triggered escape of Pol II into the elongation, enabling long-lasting efficient DNA lesion scanning of the whole transcribed genome.

The finding that an increase in chromatin accessibility parallels the conservation of H3K27ac-modified nucleosomes at the flanks of already opened regions in response to UV irradiation is compatible with reports showing that there can be a significant gain in DNA accessibility without changes in nucleosome occupancy during rapid transcriptional induction. Notably, the maintenance of H3K27ac at these sites prevents the imposition of repressive trimethylation at active loci (see Fig. 2), in accordance with the rule that H3K27ac and H3K27me3 are mutually exclusive. Moreover, finding more active transcription at these loci complies with prior reports, suggesting that an increase in gene expression is associated with surges in chromatin accessibility, and that the presence of nRNA inhibits the recruitment of H3K27me3-catalyzing polycomb-repressive complex 2 (PRC2) at active genes. Our results obtained using a range of mild UV doses (8–20 J/m²), and focusing on the early phase of recovery (1–4 h), support the claim that chromatin relaxation and maintenance of H3K27ac has an active role in the repair of transcription-blocking lesions, and gives substance to previously observed rather low increase in acetylation of histones in similar conditions. In contrast, drastic chromatin remodeling observed in murine cells at a later time during recovery (6 h) when much higher doses (80 J/m²) were used fits with findings signifying that cellular response to UV depends on the exposure dose. In other words, when cells deal with unmanageable levels of damages, they need to implement more radical expression changes, which are required for the associated fate of programmed death, a protective mechanism limiting the risk of malignant transformation.

Our analysis takes advantage of a high-resolution strand-specific map of TSSs for coding and non-coding (enhancers and PROMPTs) loci, and supports the idea that bidirectional transcription of divergent RNAs arises from two distinct hubs of transcription initiation (PICs), located within a single nucleosome-depleted region (NDR). Indeed, for bidirectional mRNAs and mRNA-PROMPTs, the binding of Pol II-hypo occurs at both edges of highly accessible regions (see ATAC-seq vs Pol II-hypo in Supplementary Fig. 6a, c), which correspond to single NDRs flanked by H3K27ac nucleosomes (see arrows in Supplementary Fig. 6a, c). These observations also extend the evidence supporting the claim that enhancers and PROMPT PICs are organized in a similar manner to gene PICs. Finally, the observed differences in transcript levels between PROMPTs and mRNAs (see Fig. 5) are probably not due to differences in Pol II-hypo recruitment (see Supplementary Fig. 6a, b, bottom), but rather due to differences in the frequency of premature termination at PPP sites and/or differences in degradation of PROMPT RNAs by the RNA exosome, which is known to be inhibited upon UV stress.

By uncoupling TSSs of mRNA genes from those of PROMPTs and enhancers, we reveal that P-TEFb-dependent release of elongating Pol II from PPP sites extends to all actively transcribed regions (see Fig. 3). Interestingly, a growing number of studies have reported data to suggest that (i) UV irradiation preferentially inhibits elongation, rather than transcription initiation, (ii) P-TEFb and NELF are important regulators of UV response, and (iii) although elongation gradually decelerates due to the encounter of Pol II with DNA lesions, significant initiation/early elongation activity is observed in the first thousand bases of actively transcribed regions, a characteristic that has also been used for the identification of active TSSs genome-wide after UV. These features are consistent with our finding that new Pol II molecules are constantly recruited to PICs post-UV (see Fig. 4), suggesting that a pool of non-engaged polymerases is still available, and that they promptly proceed into initiation of start-RNAs and subsequently into elongation of longer nRNAs (see Fig. 5). Considering that Pol II ChIP-seq density depends, among others, on the epoxide residence time at a given genomic locus, and that Pol II molecules recruited in the PIC are readily phosphorylated upon UV, we propose that the rapid exchange of Pol II isoforms after UV irradiation represents a perfectly plausible cause for the decreased ability to detect Pol II-hypo molecules at TSSs upon UV. It also accounts for the increase in EI, and could explain the gain in accessibility around TSS (see Figs. 2–4, Supplementary Fig. 6, and ref. 25). Such a model explains previously published data concerning the persistence of sufficiently important levels to be worthy of attention, of (i) PIC/basal transcription factors in nuclear extracts or upstream of genes (TFHJ) and (ii) nRNAs at the beginning of genes upon UV.

We note that excision fragments (from XR-seq) are distributed more homogeneously at sense (mRNA) and antisense (PROMPT) strands of unidirectional TSSs, and at enhancers, than it could be predicted from the CAGE levels (see Fig. 6). This finding reinforces the possibility that efficient repair at stable and unstable transcripts is primed by the uniform recruitment of Pol II-hypo at all classes of PICs in the steady state (see Supplementary Fig. 6). Remarkably, UV induces continuous and uniform transition into initiation (see Fig. 4), and constantly feeds a long-lasting PPP release of lesion-sensing transcribing Pol II into the transcriptome (see Figs. 3 and 5).

This concept was further validated by re-analyzing an experiment mapping excision repair of CDPs (XR-seq) after inhibition of PPP release, and thus a fortiori preventing the beneficial effects of Pol II initiation. Indeed, upon DRB treatment, we find a drastic impairment of CDP excisions at the beginning of all classes of active transcripts (see Fig. 7d), and in particular at TTSs located within the distance normally covered by de novo released Pol II molecules, without inhibition, at 1 h post-UV (low XR-seq signal for DRB vs DMSO in cluster 0-II, see Fig. 7a, c). In addition, the decrease in the percentage of excision in cluster II (between 10 kb and WF) in DRB2 experiment (see Fig. 7a, c) indicates that continuity in initiation post UV is crucial for the repair of sequences located toward the 3′ end of longer genes. Taking into consideration that only one Pol II molecule can be accommodated per PPP site at each actively transcribed loci at the time of irradiation, applying DRB after UV mainly restricts the benefits related to the continuous recruitment, initiation, and release into elongation of multiple polymerases throughout the recovery period.

Critically, such a scheme is compatible with the evidence that continuous nRNA-seq signal can be detected uninterruptedly between 2 and 12 h post UV from the start to the end of genes. We propose that keeping initiation active maximizes the probability to repair quickly all lesions located on the TS. Our model is compatible with the idea that even if a Pol II molecule is ubiquitylated and degraded after the recruitment of the NER complex, the amount of Pol II still normally recycling after termination from all the short genes (which contain less or no lesions) will provide sufficient trailing polymerases for scanning lesion-containing genes. In this way, this mechanism can sense efficiently one after the other the more distal lesions even in longer genes, and gives an alternative biological perspective to presently favored mode of action. In agreement, we detect TC-NER activity on TS directly downstream of TSSs and 8 h after UV. In addition, repair of distal damages in long genes is detected only when DRB is not applied after UV (compare clusters III–VI with DRB2 and DMSO in Fig. 7a–c), and intensities at 8 h in normal conditions (see Fig. 8 clusters III–VI; Supplementary Fig. 7d, distal).

Increase in TC-NER at regulatory regions has also been observed in E. coli and is compatible with the idea that the act of...
antisense transcription over regulatory regions exerts a meaningful biological function\textsuperscript{64} conserved through evolution. Indeed, these DNA sequences may serve as binding sites for transcription factors, or encode target sites for RNA-binding proteins, enabling accurate regulation of topologically associated mRNA genes\textsuperscript{58,65}. Given the effect of DNA repair on the landscape of somatic mutations in cancer tissues\textsuperscript{17,18}, surveillance of these vital sequences impacts on cell’s fitness. We propose that our model could account for the low levels of substitutions recently observed upstream of genes’ TSSs and around DNA-hypersensitive (DHS) sites\textsuperscript{66,67}. As such, transcription of non-coding regulatory loci could serve to keep their transcription factor-binding sequences (TFBS) in check.

Recent advances in the field of transcription regulation point to the fact that activation of paused genes is mediated through switching from a premature termination state of Pol II at PPP sites to a processive elongation state\textsuperscript{11,13,15}, implying that continuous pausing may be a prerequisite for Pol II promoter-proximal pausing has an inhibitory effect on switching from a premature termination state of Pol II at PPP sites required for fast transcriptional induction\textsuperscript{14}. Our results, showing that persistent initiation guarantees a prolonged transcription-coupled NER across the whole transcriptome, are functionally linked to the fact that DNA damage triggers widespread and the PPP release of a given Pol II is sufficient to drive immediate initiation of the next Pol II (see Figs. 4 and 5; Supplementary Figs. 4 and 5). In other words, the clearance rate of Pol II from TSSs highly depends on PPP status. These findings are favorable to the emerging concept that Pol II promoter-proximal pausing has an inhibitory effect on initiation\textsuperscript{68-70}, and highlight how this mechanism can function across the whole transcriptome. At the same time, our results provide a compelling physiological relevance to why cells could gain from firing initiation continuously, as a balance between promoter-proximal termination and escape into elongation allows efficient dynamic responses to stimuli or genotoxic stress.

### Methods

#### Cells culture and treatments.

Cells used in this study were VH10 hTERT-immortalized human skin fibroblasts, and were cultured with low-serum starvation, and released in full medium as described previously\textsuperscript{25}, unless stated differently. When applied, 5,6-dichloro-1-B-D-ribofuranosylbenzimidazole (DRB, Calbiochem) and triptolide (called TRP, Invivogen) were used in a final concentration of 100 μM and 125 nM, respectively, and they were added directly in group to media at the indicated times. Cells were irradiated with UV-C (254 nm, TUV Lamp, Philips) (8-20 mJ/cm\textsuperscript{2}, as indicated) and left to recover for indicated times (see figure legends and below).

#### ChIP-seq.

ChIP-seq was performed as previously described\textsuperscript{25} with minor changes (see Supplementary Methods for details). Cells were mock-treated (NO UV) or treated with UV (+UV) with 15 mJ/cm\textsuperscript{2} except if otherwise stated. Treated cells were left to recover for 2 h before harvesting or as indicated on the timeline (Fig. 4). The antibodies used for ChIP were the following: H3K27ac (ab7429, Abcam), H3K27me3 (ab7449, Millipore), and WDR5 (ab909, Millipore). The primers used for ChIP-qPCR experiments were the following (5′ to 3′, forward: F, reverse: R, ChiA neg was the negative primer): SSBP1_F: GTGAGGGAGGAAGGGATAGC, GATAACCGACAGAGGACATGG, MY C_F: ACTCAGCTGCTGAGGGAATCC, MY C_R: GGAGGAAATGATAGAGG CATTAAGGG, AKNA_F: CGGTTCACCATCTCCATCC, AKNA_R: TGGGAAAGAAGTTACACAG, APRT_F: GCCGGTCAACCAGCTTGT, APRT_R: TAG GGCAGGCTGAGTTTAAA, ChiA_neg_F: AGTCTGACGTCTTTGGGACGC, and ChiA_neg_R: GCTTCCGAGATACTAGTCC. Primer sequences: (i) ChIP-seq for Pol II-s2P and Pol II-hypo (Figs. 2 and 3; Supplementary Figs. 3, 6) were obtained from ref.\textsuperscript{25} (GSE83763) and obtained from ref.\textsuperscript{25} (GSE75398). Downloaded data were processed as described in ref.\textsuperscript{25} (Fig. 5; Supplementary Fig. 5).

#### ATAC-seq.

For H3K27ac and H3K27me3 ChIP-seq alignment files, peak calling was performed using MACS2\textsuperscript{25}. Because of the variability of ATAC-seq fragment lengths, several runs of the peak-calling algorithm were performed, using different parameters per run, in an attempt to maximize the sensitivity of the detection of open-chromatin regions. In particular, --nomodel --shift 100 --extsize 200 --shift 100 --extsize 200, --nomodel --shift 37 --extsize 73, --broad --shift 37 --extsize 73, --broad --nomodel --shift 37 --extsize 73, --broad --nomodel --shift 37 --extsize 73, --keep-dup all, --shift 37 --extsize 73, --keep-dup all and --nomodel --shift 37 --extsize 150 --keep-dup all were used for MACS2 peak calling. Only peaks present in five out of seven methods were kept for further analysis. Although the majority of peaks detected were common between conditions, a number of peaks were also detected only in NO UV and +UV. As these peaks were less in number than the common ones, and showed a rather low density of reads in another attempt to maximize the sensitivity of the detection of open-chromatin regions, and to unbiasedly investigate the changes occurring upon irradiation, we considered the union of the peaks in each condition to perform the rest of the analysis (Supplementary Table 1). To conduct differential accessibility analysis, diffBind R package (https://bioconductor.org/packages/2.10/bioc/html/diffBind.html) was used, with the filtered ATAC-seq peak set as a reference.

#### RT-qPCR.

Differential accessibility regions were detected and filtered by applying fold change (Log, FC ≥ 1) and P-value (p-val ≤ 0.001) thresholds.

---

**Nature Communications | 2020**

[https://doi.org/10.1038/s41467-020-14566-9](https://doi.org/10.1038/s41467-020-14566-9)
Read-density plots. ATAC-seq, CHIP-seq, RNA-seq, CAGE-seq, and XR-seq data were subjected to read-density analysis after read depth normalization of all samples. Heatmaps and average-density profiles were produced as described previously [1] around genomic regions of interest, as indicated in the figures. Heatmaps were generated directly using the software, from matrices of binned read densities (bin size is indicated in the figures) for all considered individual (n) items (metagenes). Read-density matrices were also imported in R and python using heatmap plotting average-density profiles (smoothing achieved by a moving window of the bin size as indicated) and (ii) for determination of read densities per genomic category.

Construction of mRNA-TSSs, PROMPT-TSSs, and eTSS annotation. To annotate transcription start sites (TSSs), all known protein-coding and non-coding RNA hg19 RefSeq transcripts release 86 was downloaded from UCSC table browser (http://genome-euro.ucsc.edu/cgi-bin/hgTables). For each transcript, a biotype was assigned using BioMart (www.biomart.org), and all the small non-coding RNAs were excluded. For all the gene models containing multiple alternative transcripts, TSS neighborhoods of a 100-bp window were clustered together, and only the longest TSS was kept, resulting in 30,473 transcripts. Transcripts were then separated into three groups, based on their transcriptional activity. TSS coordinates were extended to 2 kb on each direction, and were tested for overlap with the Pol II-ser2P-UV, H3K27ac-UV, and H3K27me3-UV peak sets. Regions overlapping with at least one Pol II-ser2P-UV and H3K27ac-UV peak were characterized as active, with a overlapping with a H3K27me3-UV peak, but not with a Pol II-ser2P-UV or with an H3K27ac-UV peak, were characterized as repressed, and those that did not overlap with any of the above peak sets were characterized as inactive. Any region overlapping with both H3K27ac-UV and H3K27me3-UV peaks was excluded from the rest of the analysis. The resulted in 15,819 active, 2943 repressed, and 7608 inactive transcripts (Supplementary Table 1). To further classify the active TSSs in terms of transcription directionality, the annotation was split up into unidirectional and bidirectional references. All active transcript pairs with opposite direction of transcription (forward(+) vs reverse(−)), where −2 kb ≤ TSSdistance ≤ 12 kb, were characterized as bidirectional, while the rest of the annotations were characterized as unidirectional. Bidirectional pairs were further categorized into two groups of annotations: convergent bidirectional transcript pairs with TSSdistance ≤ 100 bp, and divergent bidirectional transcript pairs with TSSdistance > 100 bp. To optimize the annotation of convergent TSS transcription pairs, coordinates were redefined by scanning at a radius of 250 bp, to detect the nucleotide occupied by the maximum-sense CAGE signal. Any bidirectional pair with a non-significant CAGE peak in the aforementioned region was excluded from the analysis. This finally resulted in 12,859 unidirectional transcripts and 2822 active bidirectional TSS pairs, 1806 of which were characterized as divergent and 1016 as convergent (Supplementary Table 1).

To gain a complete overview of the non-coding antisense transcription events occurring around mRNA-TSSs, we also annotated upstream antisense (uRNA) and downstream antisense (dRNA) transcripts (referred as an ensemble to PROMPT-TSSs in this context). Only active unique uRNA and dRNA-TSSs were used. For all the genes annotated with more than one mRNA transcript, only the leftmost TSS (for + strand genes), and rightmost TSS (for − strand genes) were considered for the rest of the analysis. The antisense CAGE peak with the highest sumit in the region that ranged from −2 kb upstream to +1 kb downstream of eTSSs was considered to be the upstreamTSS coordinate_forward reference strand (−TSSdistance) and downstreamTSS coordinate_reverse strand (TSSdistance). Forward (−) and reverse (T-) TSSs were extended to 1 kb, and sense and antisense maximum CAGE summit heights were detected at 100 bp downstream of TSS.

Promoter escape indices analysis. Promoter escape analysis was performed for a subset of active unidirectional and bidirectional transcripts, PROMPTs and active enhancers. In particular, to avoid the inclusion of Pol II-ser2P reads mapped in overlapping promoters and gene bodies, only active divergent unidirectional transcript—PROMPT pairs were considered, where TSSdistance > 100 bp, TSSdistance = TSS coordinate_forward reference strand − TSS coordinate_reverse reference strand, active divergent bidirectional transcript pairs with TSSdistance > 100 bp, and active divergent enhancers with no nearby transcripts within 10 kb and no nearby eTSSs within 2 kb. TSSs and PROMPT-TSS promoter escape indices (EI, inverse of pausing index) were calculated as previously defined [2], by taking the average coverage in rpm in the gene body (density in gene body was abbreviated as Db and ranged from 101 bp to 2 kb downstream of TSS or 101 bp downstream of TSS for genes larger or smaller than 2 kb, respectively) divided by the average coverage on the promoter-proximal region (Dp) ranged from 250 bp upstream to 100 bp downstream of TSS.

For enhancer escape analysis, EI was calculated as above, where density of reads at enhancer flanks (Df) is calculated for the regions ranging from −2 kb to −100 bp upstream of TSS and from +100 bp to +2 kb downstream of TSS, while density of reads on enhancer TSS (De) is calculated for the regions ranging from 100 bp upstream to 100 bp downstream of eTSS.

Nucleotide excision repair data meta-analysis. The strand-specific genome-wide maps of nucleotide excision repair of the UV-induced DNA damage (CPDs), available for XP-C mutants lacking the global genome nucleotide excision repair mechanism (GG-NER-deficient, TC-NER-proficient), were obtained from Hu et al. [3] (UV: 20/1 m², data used in Fig. 6, Gene Expression Omnibus (GEO) accession number GSE67941) and Chiao et al. [4] (UV: 20/2 m² data used in Fig. 7, GEO accession number GSE106823). XR-seq data for wild-type (WT) C57BL/6J (UV: 10/2 m², used for Fig. 8 and Supplementary Fig. 7, GEO accession number GSE76391) were obtained from Adar et al. [5]. Sequence read archive (SRA) data sets were downloaded from Gene Expression Omnibus using the sra toolkit prefetch (https://www.ncbi.nlm.nih.gov/sra/docs/sraadownload/) command, and converted to fastq files using fastq-dump. Fastq quality control, data filtering, and short read alignment were performed as above. Meta-analysis involved that read counts were normalized to equal read depth. Heatmap read-density matrices and average read-density plots of CPD XR-seq read around potential pyrimidine dimers. For XP-C data, we focus on TpT sequences (TT) as explained previously [3], and filtered the TTs overlapping with enhancers defined in ENCODE, available to avoid signals generated from eTSSs. For WT cells, the list of TTs was further curated to only consider TTs in active genes (or divergent genes) located on plus strand and filtering out TTs in the region between TSS and +2 kb for CONV and DIV bidirectional and unidirectional genes with inter-TSS < 100 bp (Supplementary Table 1). XR-seq reads were computed as described above. Read-density matrices were calculated for both strands (TSS and NTS) separately when indicated. For WT cells, only minus reads were used as these correspond to excised DNA from TS for plus genes. The ratio of XR-seq reads between directions, and calculation of variability between directions was performed as described in the legend of Fig. 6. F–F’ scores and quantification of reads around TT loci was performed as described in Lavigne et al. [7], with cluster borders defined previously.

FANTOM5 Cap-analysis of gene expression (CAGE) sequencing data meta-analysis. The FANTOM5 strand-specific CAGE-seq alignment files of normal dermal fibroblast primary cells (six donors with source code: 11269-1169G, 11346-1171GS, 11418-1181FS, 11450-1191A, 11454-119A5, and 11458-119A5) and normal skin fibroblasts (two donors with source code: 11653-120CS and 11651-120D4) were downloaded from ftp://ftp.biociedebc.ph/farchive/fantom5/datasets/ phase2_2/basic/human.primary_cell.HCAGE and were combined. Heatmap read-density matrices and average read-density plots were computed as described in the section “Read densities heatmaps and average plots”. Read-density matrices were calculated for both strands separately.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The data reported in this paper have been deposited with the Gene Expression Omnibus under accession code GSE125181.

Received: 11 July 2019; Accepted: 16 January 2020; Published online: 14 February 2020

References. 1. Haberre, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. Nat. Rev. Mol. Cell Biol. 19, 621–637 (2018).
7. Henriques, T. et al. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Nature* **507**, 455–461 (2014).
8. Kim, T.-K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **182**, 187–191 (2017).
9. Preker, P. et al. RNA exosome depletion reveals transcription upstream of active human promoters. *Science* **322**, 1851–1854 (2008).
10. Henriques, T. et al. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Nature* **507**, 455–461 (2014).
11. Lai, W. K. M. & Pugh, B. F. Genome-wide uniformity of human ‘open’ pre-initiation complexes. *Genome Res.* **27**, 15–26 (2017).
12. Nechaev, S. et al. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science* **327**, 335–339 (2010).
13. Cheng, B. & Price, D. H. Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *J. Biol. Chem.* **282**, 21901–21912 (2007).
14. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* https://doi.org/10.1101/gad.325142.119 (2019).
15. Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **13**, 1311–1320 (2014).
16. Krebs, A. R. et al. Genome-wide single-molecule footprinting reveals high turnover of initiating and promoter-paused RNA polymerase II. *Cell* **167**, 411–422.e4 (2017).
17. Steurer, B. et al. Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA polymerase II. *Proc. Natl Acad. Sci. USA* **115**, E4368–E4376 (2018).
18. Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071–1078 (2009).
19. Liakos, A., Lavigne, M. D. & Fousteri, M. Nucleotide excision repair: from damage response to transcription stress. *Nature* **439**, 462 (2015).
20. Li, J. et al. Principles of RNA metabolism and alternative transcription initiation within closely spaced promoters. *Nat. Genet.* **48**, 984–994 (2016).
21. Li, G. & Ho, V. C. p53-dependent DNA repair and apoptosis respond differently to high- and low-dose ultraviolet radiation. *Br. J. Dermatol.* **139**, 3–10 (1998).
22. Farrell, A. W., Halliday, G. M. & Lyons, J. G. Chromatin structure following UV-induced DNA damage—repair or death? *Int. J. Mol. Sci.* **12**, 8065–8085 (2011).
23. Rehentullaa, A., Hamilton, C. A., Chinnaiyan, A. M. & Dixit, V. M. Ultraviolet radiation-induced apoptosis is mediated by activation of CD-95 (Fas/APO-1). *J. Biol. Chem.* **272**, 25783–25786 (1997).
24. Caricchio, R., McPhie, L. & Cohen, P. L. Ultraviolet B radiation-induced cell death: critical role of ultraviolet dose in inflammation and lupus autoantigen redistribution. *J. Immunol.* **157**, 5778–5786 (2001).
25. Ibrahim, M. M. et al. Determinants of promoter and enhancer transcription directionality in metazoans. *Nat. Commun.* **9**, 4472 (2018).
26. Scruggs, B. S. et al. Bi directional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Cell Mol. Cell Biol.* **58**, 1101–1112 (2015).
27. Blasius, M., Wagner, S. A., Choudhary, C., Bartek, J. & Jackson, S. P. A quantitative 14–3–3 interaction screen connects the nuclear exosome targeting complex to the DNA damage response. *Genes Dev.* **28**, 1977–1982 (2014).
28. Tiedje, C. et al. p38MAPK/MK2-mediated phosphorylation of RBM7 regulates human nuclear exosome targeting complex. *RNA* **21**, 262–278 (2015).
29. Boeing, S. et al. Multiomics analysis of the UV-induced DNA damage response. *Cell Rep.* **15**, 1597–1610 (2016).
62. Ehrensberger, A. H., Kelly, G. P. & Svejstrup, J. Q. Mechanistic interpretation of promoter-proximal peaks and RNAPII density maps. *Cell* **154**, 713–715 (2013).
63. Adebali, O., Chiou, Y.-Y., Hu, J., Sancar, A. & Selby, C. P. Genome-wide transcription-coupled repair in *Escherichia coli* is mediated by the Mfd translocase. *PNAS* **114**, E2116–E2125 (2017).
64. Murray, S. C. & Mellor, J. Using both strands: the fundamental nature of antisense transcription. *Bioarchitecture* **6**, 12–21 (2016).
65. Seila, A. C., Core, L. J., Lis, J. T. & Sharp, P. A. Divergent transcription: a new feature of active promoters. *Cell Cycle* **8**, 2557–2564 (2009).
66. Haradhvala, N. J. et al. Mutational strand asymmetries in cancer genomes reveal mechanisms of DNA damage and repair. *Cell* **164**, 538–549 (2016).
67. Sabarinathan, R., Mularoni, L., Deu-Pons, J., Gonzalez-Perez, A. & López-Bigas, N. Nucleotide excision repair is impaired by binding of transcription factors to DNA. *Nature* **532**, 264–267 (2016).
68. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* **49**, 1045–1051 (2017).
69. Fitz, J., Neumann, T. & Pavri, R. Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation. *EMBO J.* **37**, e97965 (2018).
70. Gressel, S. et al. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *elife* **6**, e29736 (2017).
71. Xu, S., Grullon, S., Ge, K. & Peng, W. Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. *Methods Mol. Biol.* **1150**, 97–111 (2014).
72. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

Acknowledgements
We thank Pantelis Hatzis, Mihalis Verykokakis, and members of the Fousteri lab for critical discussions and reading of the paper. We thank Panagiotis Moulos for help and discussion on the underdevelopment computational pipelines. We thank Vladimir Benes and the Genecore facility (EMBL, Germany) for the special care they use in sequencing our NGS libraries. This work was funded by a European Research Council grant to M.F., Agreement-309612 (TransArrest) and Matching Funds to MF funded by National sources.

Author contributions
M.F. and M.D.L. designed the study and were responsible for interpretation of the results. M.F. directed the study and obtained financial support. M.D.L., M.F., and A.L. wrote the paper and all authors edited the paper. A.L. performed the experimental part of the study. D.K. performed the statistical and bioinformatics analyses. M.D.L. contributed significantly in the bioinformatics analysis of the data. All authors discussed the results, reviewed, commented on, and approved the final version of the paper. A.L and D.K. contributed equally to the paper.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-14566-9.

Correspondence and requests for materials should be addressed to M.D.L. or M.F.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020