Check for updates



MOLECULAR BIOLOGY

# XPF interacts with TOP2B for R-loop processing and DNA looping on actively transcribed genes

**Georgia Chatzinikolaou[1]†, Kalliopi Stratigi[1]†, Athanasios Siametis[1,2], Evi Goulielmaki[1], Alexia Akalestou-Clocher[1,2], Ioannis Tsamardinos[3], Pantelis Topalis[1], Caroline Austin[4], Britta A. M. Bouwman[5], Nicola Crosetto[5,6], Janine Altmüller[7], George A. Garinis[1,2]***

Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

**Co-transcriptional RNA-DNA hybrids can not only cause DNA damage threatening genome integrity but also regulate gene activity in a mechanism that remains unclear. Here, we show that the nucleotide excision repair factor XPF interacts with the insulator binding protein CTCF and the cohesin subunits SMC1A and SMC3, leading to R-loop–dependent DNA looping upon transcription activation. To facilitate R-loop processing, XPF interacts and recruits with TOP2B on active gene promoters, leading to double-strand break accumulation and the activation of a DNA damage response. Abrogation of TOP2B leads to the diminished recruitment of XPF, CTCF, and the cohesin subunits to promoters of actively transcribed genes and R-loops and the concurrent impairment of CTCF-mediated DNA looping. Together, our findings disclose an essential role for XPF with TOP2B and the CTCF/cohesin complex in R-loop processing for transcription activation with important ramifications for DNA repair–deficient syndromes associated with transcription-associated DNA damage.**

## INTRODUCTION

Transcription requires the concerted action of basal transcription factors, sequence-specific DNA binding proteins, chromatin remodeling, and modification enzymes to enable the synthesis of the primary transcript (*1*). Besides transcription-blocking DNA insults, the process of mRNA synthesis leads to transcription-associated recombination or rearrangements that occur during robust shifts in transcription demands threatening cell viability (*2*, *3*). To ensure that genome integrity is preserved and that transcription is not compromised, cells use a battery of partially overlapping DNA repair systems aimed at counteracting DNA damage and restore DNA to its native form (*4*).

ERCC1-XPF is a two-subunit structure–specific endonuclease where XPF contains the nuclease domain of the complex and ERCC1 is required for DNA binding and the subsequent nuclease activity (*5*). The complex is essential for incising DNA 5′ to the DNA lesion during nucleotide excision repair (NER) (*6*, *7*), a highly conserved mechanism that removes helical distortions throughout the genome, i.e., global genome NER, or selectively from the transcribed strand of active genes, i.e., transcription-coupled NER (*7*–*10*). Besides NER, ERCC1-XPF is required for the repair of DNA interstrand cross-links (*11*, *12*) and for removing nonhomologous 3′ single-stranded tails from DNA ends during double strand break (DSB) repair by homologous recombination (HR) or by alternative nonhomologous end joining (NHEJ), where short stretches of homology are used to join two broken DNA ends (*13*–*17*). Furthermore, ERCC1-XPF is involved in telomere maintenance (*18*, *19*), and recently, it was shown to play a role in a subpathway of long-patch base excision repair involving 5′ gap formation (*20*).

In humans, mutations in ERCC1-XPF lead to xeroderma pigmentosum (XP; affected proteins: XPA through XPG), Cockayne syndrome (CS; affected proteins: CSA, CSB, UVSSA, XPB, XPD, XPF, TTDA, and certain mutations in the gene encoding XPG), or Fanconi anemia, whose clinical outcomes are exceptionally diverse (*21*–*25*). For instance, patients with mutations in ERCC1 manifest a severe form of CS named cerebro-oculo-facio-skeletal syndrome (*6*, *26*). Instead, the great majority of XPF patients present with mild symptoms of XP, including sun sensitivity, freckling of the skin, and basal or squamous cell carcinomas that typically occur at later stages in life (*6*). Mice carrying inborn defects in *Ercc1* and *Xpf* genes fully recapitulate the severe growth retardation and premature onset of heterogeneous pathological symptoms seen in patients with defects in the corresponding genes (*27*, *28*).

Recent studies have shown that ERCC1-XPF plays a role in the regulation of gene expression (*29*, *30*), chromatin looping (*31*, *32*), and the fine-tuning of growth-promoting genes during postnatal development (*33*). However, no solid evidence exists on how the ERCC1-XPF endonuclease complex is functionally involved in these processes. By using an in vivo biotinylation tagging strategy, coupled with functional genomics and proteomics approaches to map DNA DSBs, we have found that XPF interacts with TOP2B, CCCTC-binding factor (CTCF), and the cohesin subunits SMC1A and SMC3 on active promoters. This interaction facilitates efficient R-loop processing and CTCF-mediated DNA looping in actively transcribed genes, which has substantial implications for transcription-associated DNA damage and gene regulation.

[1]Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology–Hellas, GR70013, Heraklion, Crete, Greece. [2]Department of Biology, University of Crete, Heraklion, Crete, Greece. [3]Computer Science Department of University of Crete, Heraklion, Crete, Greece. [4]Biosciences Institute, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. [5]Division of Microbiology, Tumor and Cell Biology, Karolinska Institutet and Science for Life Laboratory, Stockholm 17177, Sweden. [6]Human Technopole, Viale Rita Levi-Montalcini 1, 22157 Milan, Italy. [7]Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin Institute of Health at Charité–Universitätsmedizin Berlin, Core Facility Genomics, Charitéplatz 1, 10117 Berlin, Germany.
*Corresponding author. Email: garinis@imbb.forth.gr
†These authors contributed equally to this work

Downloaded from https://www.science.org on February 13, 2025

## RESULTS

### Transcription activation and UVC irradiation differentially recruit XPF to DNA

To assess the recruitment of XPF to DNA, genome-wide, we crossed homozygous av*Xpf*$^{+/+}$ knockin mice expressing XPF fused with a 15–amino acid biotinylatable sequence and a 3× FLAG tag with mice broadly expressing the hemagglutinin (HA)–tagged bacterial BirA biotin ligase transgene (BirA) (Fig. 1A) (*32*). BirA is a bacterial ligase that specifically biotinylates the 15–amino acid avidin within the tag, allowing the isolation of biotin-tagged XPF (bXPF)–bound genome targets and protein complexes by binding to streptavidin. Streptavidin pulldowns followed by high-throughput sequencing (bXPF-Seq) were performed on primary bXPF and BirA mouse embryonic fibroblasts (MEFs) under basal conditions, upon

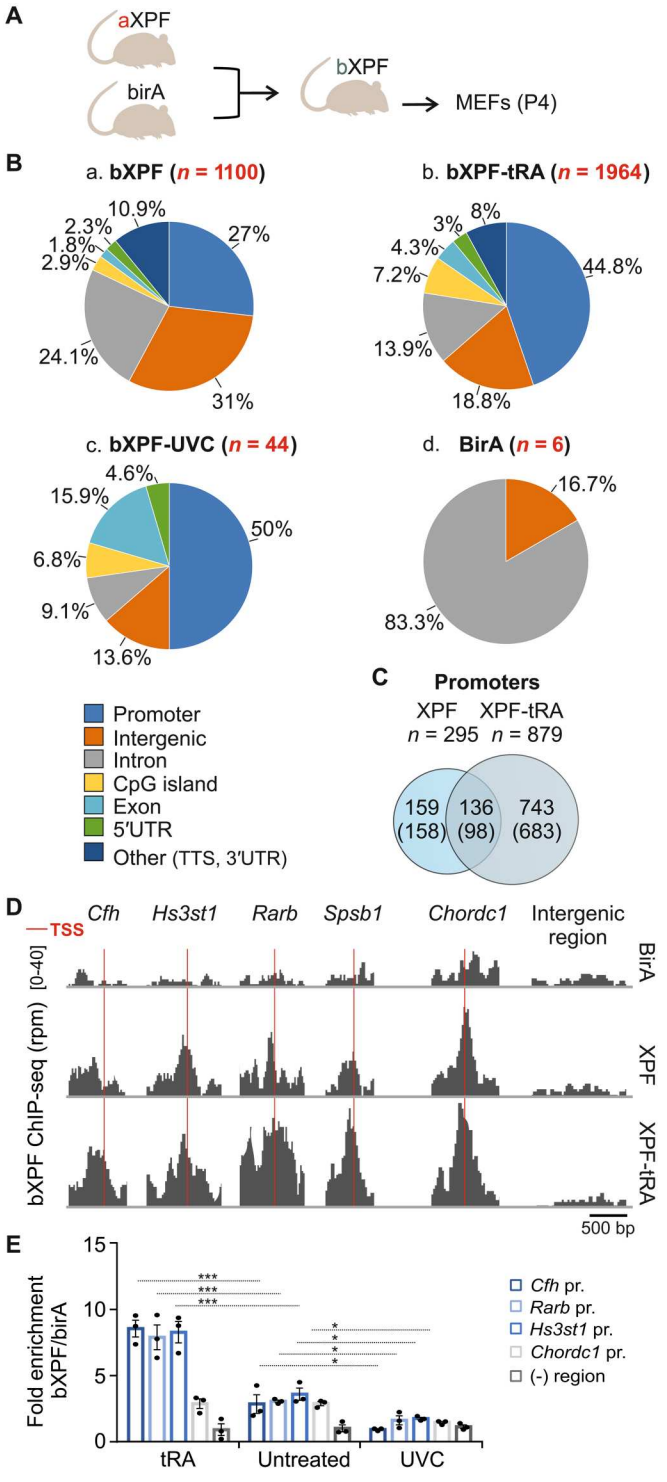


**Fig. 1. Genome-wide ChIP-Seq analysis of XPF occupancy in mouse embryonic fibroblasts (MEFs).** (**A**) Schematic representation of homozygous av*Xpf*$^{+/+}$ knockin mice expressing the XPF subunit of the nucleotide excision repair (NER) structure–specific endonuclease XPF-ERCC1 fused with a 15–amino acid tandem affinity purification–tag biotinylatable sequence and a 3× FLAG tag crossing with mice broadly expressing the hemagglutinin (HA)–tagged bacterial BirA biotin ligase transgene (BirA) to generate bXPF MEFs used for the ChIP-Seq analysis. (**B**) Pie charts illustrating the genomic distribution of bXPF binding sites in untreated, tRA-treated (10 μμ, 16 hours), and ultraviolet C (UVC)–treated (10 J/m$^2$) bXPF and BirA (control) MEFs. Peaks occurring within ±2 kb of the transcription start site (TSS) were considered promoter. (**C**) Venn diagram of XPF and XPF-tRA ChIP-Seq peaks mapped on promoters and corresponding number of unique genes (parenthesis). (**D**) Genome browser views depicting bXPF ChIP-Seq signals on ±2 kb genomic regions flanking the TSS of representative trans retinoic acid (tRA)–responsive genes (e.g., *Cfh*, *Hs3st1*, *Rarb*, and *Spsb1*), tRA nonresponsive gene (e.g., *Chordc1*), and the nontranscribed genomic region (intergenic region) in untreated (bXPF) and tRA-treated (bXPF-tRA) MEFs. A black line sets the scale at 500 bp. (**E**) bXPF ChIP signals on the promoters of tRA-induced *Cfh*, *Rarb*, and *Hs3st1* genes, the tRA-noninduced *Chordc1* gene, and on an intergenic nontranscribed (-) region.

Downloaded from https://www.science.org on February 13, 2025

transcription stimulation or upon exposure to ultraviolet C (UVC) irradiation. An irreproducible discovery rate filtering across three biological replicates [false discovery rate (FDR) ≤ 0.05] revealed that most of the identified bXPF-Seq peaks were mapped to intronic (24.1%), promoter (27%), and intergenic (31%) sequences under native conditions (Fig. 1, Ba, Bd, and D). Under transcription activation by all-trans retinoic acid (tRA), a pleiotropic factor known to activate transcription during cell differentiation and embryonic development (*34*), the number of bXPF-Seq peaks increased by 78% (i.e., 1964) genome-wide (Fig. 1, Bb and D, and fig. S1A) and almost twice as much (198%) at promoters corresponding to 683 unique, well-annotated genes (Fig. 1C). Instead, the number of bXPF-Seq peaks in bXPF MEFs exposed to 10 J/m$^2$ of UVC irradiation (i.e., 44) was comparable to that seen in untreated BirA

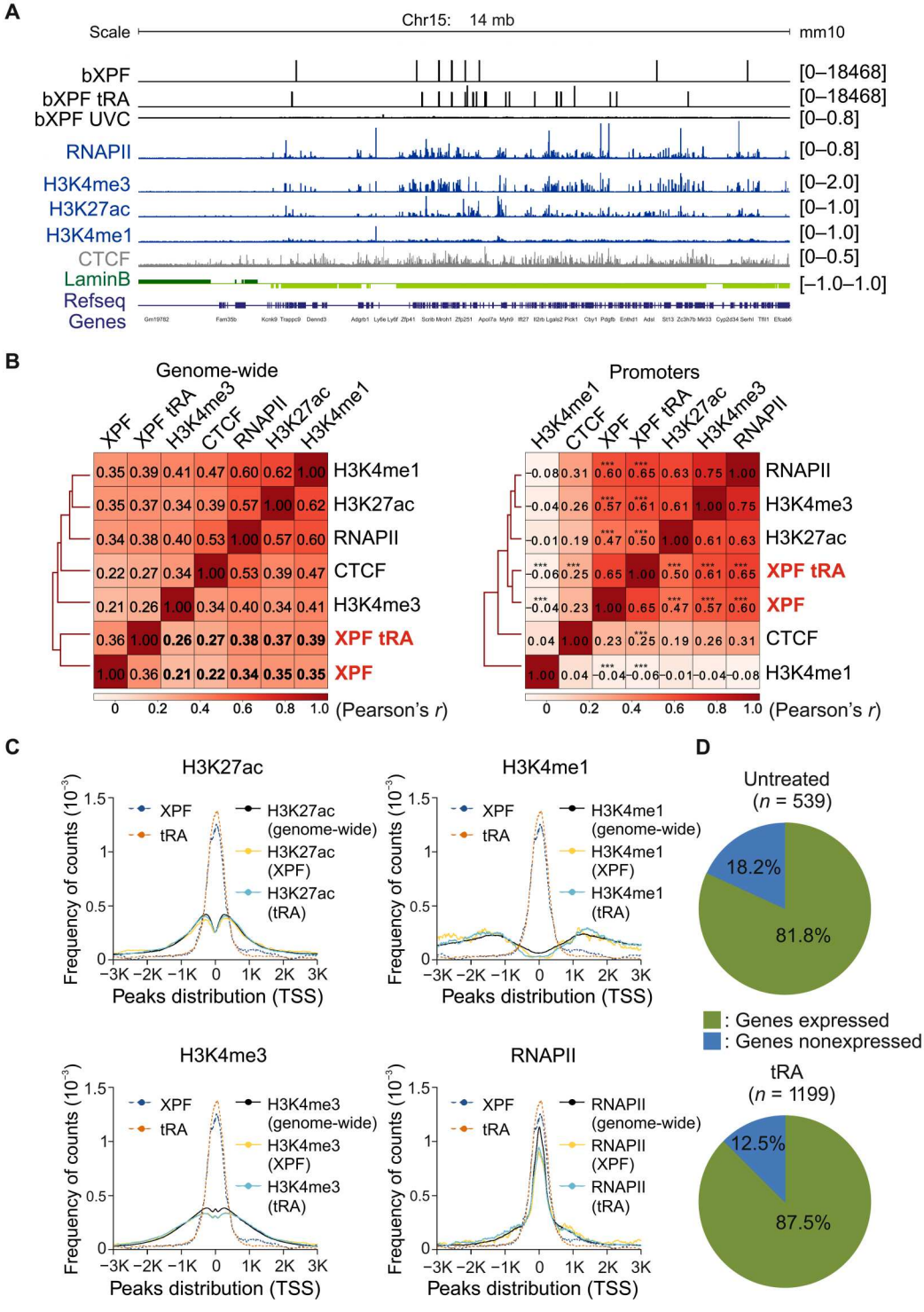


**Fig. 2. Chromatin state of XPF recruitment sites.** (**A**). IGV overview of ChIP-Seq profiles for bXPF [untreated mouse embryonic fibroblasts (MEFs)], bXPF-tRA [trans retinoic acid (tRA)–treated MEFs], bXPF-UV (UV-irradiated MEFs), RNAPII, H3K4me3, H3K27ac, H3K4me1, CTCF, and Dam-ID profile of lamin B on a representative 14-Mb (chromosome 15) genomic region in MEFs. (**B**). Genome-wide (left) or gene promoter–wide (right) heatmap representation of Pearson's *r* correlation analysis of XPF (untreated bXPF MEFs), XPF-tRA (tRA-treated MEFs), RNAPII, H3K27ac, H3K4me3, and H3K4me1 ChIP-Seq profiles. For promoters, the *P* value (***$P$ = 0.001 to 0.0001 and **$P$ = 0.05 to 0.001) is based on the comparison of Pearson's *r* correlations (single-sided test) from independent samples, in this case, between the correlations of genome-wide and promoter-associated ChIP-Seq signals. (**C**) Average count frequencies on ±3-kb genomic regions flanking the transcription start site (TSS) for RNAPII, H3K27ac, and H3K4me3 activating histone marks, H3K4me1 repressive histone modification, and bXPF- and bXPF-tRA–bound gene targets. Continuous lines depict the profile frequencies genome-wide (black), only for gene targets bound by XPF (yellow), or only for gene targets bound by XPF upon tRA (light blue). Dotted lines depict the genome-wide profiles of bXPF (blue) and bXPF-tRA (orange). (**D**) Pie charts depicting the RNA-Seq gene expression status (blue: nonexpressed; green: expressed) of bXPF-bound genes [5′UTR, promoter-TSS, exon, intron, transcription termination site (TTS), and 3′UTR] in untreated (top pie chart) and tRA-treated (10 μμ, 16 hours) (bottom pie chart) MEFs.

Downloaded from https://www.science.org on February 13, 2025

transgenic control cells (Fig. 1, Bc and Bd, and fig. S1A). A series of follow-up chromatin immunoprecipitation (ChIP) assays coupled with quantitative polymerase chain reaction (qPCR) on peak sequences flanking the transcription start sites (TSSs) of *Cfh*, *Rarb*, and *Hs3st1* gene promoters were conducted. These promoters were selected from RNA-sequencing (RNA-Seq) profiling in untreated and tRA-treated bXPF MEFs (fig. S1E). The ChIP assays confirmed the recruitment of bXPF on promoters in untreated MEFs, the significantly higher bXPF ChIP signals in tRA-treated MEFs, and the substantial reduction of bXPF ChIP signals in UVC-irradiated MEFs (Fig. 1E and fig. S1B, as indicated). We also find that bXPF is recruited minimally to the promoter regions of the tRA nonresponsive genes, e.g., *Chordc1*, *Dcaf10*, and *Dhx16*, or in the nontranscribed genomic regions in tRA-treated MEFs compared to untreated control cells (Fig. 1, D and E; as indicated, fig. S1, B and C). In addition, bXPF ChIP signals on *Cfh*, *Rarb*, and *Hs3st1* gene promoters were significantly reduced in UVC-irradiated, tRA-treated (tRA/UVC) cells compared to non–UVC-irradiated, tRA-treated control cells (fig. S1D). In line, we find a significant reduction in the *Cfh*, *Rarb*, and *Hs3st1* mRNA levels in tRA/UVC-treated MEFs compared to the non–UVC-irradiated, tRA-treated control cells (fig. S1, E and F). Thus, under conditions that favor transcription, XPF is primarily recruited to promoters. However, the protein is randomly distributed throughout the genome when cells are exposed to UVC radiation. The latter is consistent with the indiscriminate distribution of DNA lesions following exposure to genotoxic agents.

## bXPF recruitment on DNA coincides with RNAPII and active histone PTMs

Combinatorial ChIP-Seq profiles provide insights into shared or differential protein occupancies and histone marks. The preferential recruitment of bXPF to promoters prompted a comparison of the genome-wide distribution of bXPF with the ChIP-Seq profiles of protein factors known to associate with active transcription, including RNA polymerase II (RNAPII), histone 3 tri-methylation at lysine 4 (H3K4me3), and histone 3 acetylation at lysine 27 (H3K27ac) (*35*, *36*), or context-dependent transcription, i.e., histone 3 monomethylation at lysine 4 (H3K4me1) (*37*, *38*), and facultative or constitutive heterochromatin, i.e., lamin B (*39*). Our analysis also included the CTCF factor known to interact with the ERCC1-XPF complex during postnatal murine development (*32*). Observation of a representative region from chromosome 15 makes it evident that bXPF associates preferentially to gene-dense areas that closely coincide with regions bound by RNAPII, the active histone marks H3K4me3 and H3K27ac, and with CTCF. Consistently, bXPF ChIP signals were excluded from lamin B–associated heterochromatic areas reflecting low-density gene regions (Fig. 2A). Analysis conducted on a genome-wide level demonstrated that the bXPF ChIP-Seq profiles obtained from untreated or tRA-treated MEFs displayed a positive correlation with the ChIP-Seq profiles of RNAPII, the activating histone marks H3K4me3 and H3K27ac, as well as those of H3K4me1 (Fig. 2B, left). The Pearson correlation coefficient ($r$) significantly increased when bXPF was analyzed with RNAPII, H3K4me3, and H3K27ac ChIP-Seq signals on promoters alone, indicating a stronger association between bXPF and these active transcription factors in promoter regions. However, the correlation coefficient decreased or remained unchanged when bXPF was analyzed with H3K4me1 or CTCF ChIP-Seq signals, respectively (Fig. 2B, right). To gain further insight into the recruitment of XPF to promoters, we next calculated the average coverage around the TSS of genes bound by XPF in untreated and tRA-treated MEFs. Our analysis revealed that bXPF is substantially enriched (Fig. 2C, blue dotted line) with RNAPII, H3K4me3, and H3K27ac (Fig. 2C, continuous lines as indicated) around TSS. The enrichment of bXPF on TSS was further pronounced in tRA-treated MEFs (Fig. 2C, orange dotted line). The sharp dip of H3K27ac around the TSS is a common feature of the TSS centered plots that reflects the position of the nucleosome-depleted zone (*40*). Consistent with the negative correlation of bXPF with H3K4me1 on promoters (Fig. 2B, right), we found that H3K4me1 is locally depleted from TSS (Fig. 2C; as indicated). Next, we examined whether the recruitment of bXPF on promoters and gene bodies associates with productive, steady-state mRNA synthesis. The RNA-Seq analysis in untreated and tRA-treated bXPF MEFs revealed that out of 539 bXPF-bound genes [including 5′ untranslated region (5′UTR), promoter-TSS, exon, intron, transcription termination site (TTS), and 3′UTR region], 441 genes (81.8%) had ≥20 RNA-Seq counts (i.e., number of reads) in untreated MEFs (Fig. 2D, top pie chart, and fig. S2A). The number of bXPF-bound genes increases substantially to 1049 bXPF-bound genes (1049 of 1199; 87.54%) when the same analysis was carried out in tRA-treated MEFs (Fig. 2D, bottom pie chart, and fig. S2A). In addition, we observe that more than 70% of bXPF-bound genes (1014 of 1399) show a substantial increase in transcription levels upon tRA treatment [fig. S2B, fold change (FC) > 1]. Thus, upon transcription activation, bXPF is recruited with RNAPII and active histone post-translational modifications (PTMs)s to the promoters of actively transcribed genes.

## XPF is preferentially recruited to transcription-associated DNA breaks on promoters

Our finding that XPF is preferentially recruited on active promoters is consistent with recent findings indicating that transcription itself causes DNA DSBs on the promoters of actively transcribed genes (*41*, *42*). To test this, we treated primary MEFs with the potent genotoxin mitomycin C (MMC) or with tRA. As expected, we found an increase in γH2AX and 53BP1 foci, two well-established DNA damage markers, in MMC-treated MEFs (Fig. 3A). We found that transcription activation was associated with a substantial increase in γH2AX and 53BP1 foci in tRA-treated MEFs (Fig. 3A). The findings were further confirmed by the detectable increase in γH2AX protein levels in whole-cell extracts of tRA-treated cells, which was comparable to those seen in MMC-treated and UVC-irradiated cells, as well as in DNA repair–deficient *Ercc1*$^{-/-}$ MEFs (fig. S2C, as indicated). Ataxia telangiectasia mutated (ATM) and Rad3 related (ATR) kinases are central mediators of the DNA damage checkpoint with distinct DNA damage specificities. Whereas ATM is primarily activated by DSBs, ATR responds to a variety of DNA lesions that interfere with DNA replication (*43*). To test whether transcription triggers a canonical DDR signaling, tRA-treated MEFs were cultured in the presence of a selective inhibitor for ATM, i.e., KU-55933 or ATR (i.e., NU6027). As evidenced by Western blotting, in the presence of ATR and/or ATM inhibitors, there is a decrease in the phosphorylated (Ser$^{345}$) Chk1 and phosphorylated p53 protein levels, confirming the DDR impairment (fig. S2D). We find that, upon transcription induction, inhibition of ATM (tRA/ATMi cells) but not of ATR (tRA/ATRi cells) abolished the

Downloaded from https://www.science.org on February 13, 2025

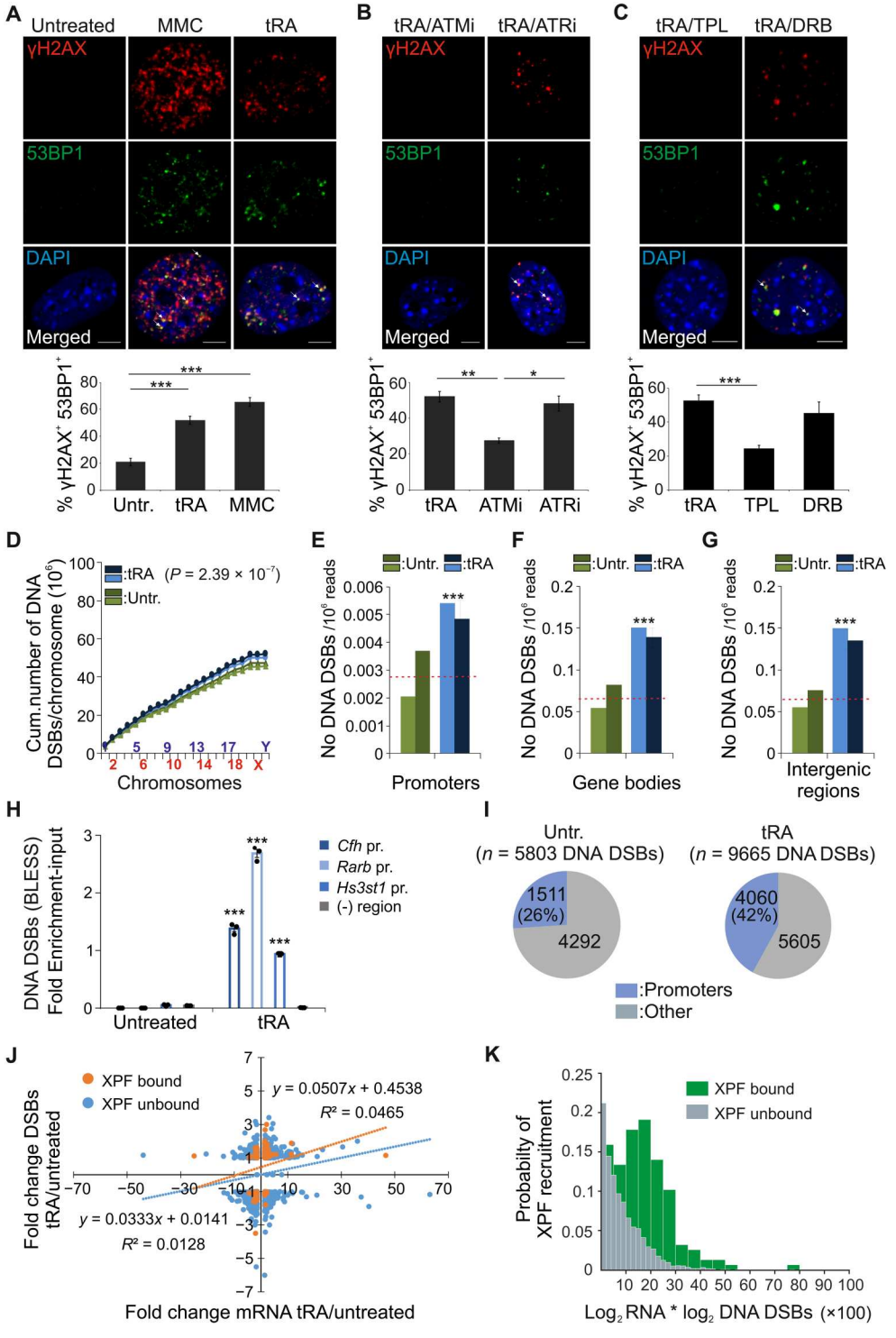


**Fig. 3. Genome-wide mapping of transcription-associated DSBs.** (**A**) Immunofluorescence detection of γH2AX and 53BP1 (white arrowheads) in wild typ (wt) mouse embryonic fibroblasts (MEFs) cultured upon basal conditions, exposure to mitomycin C (MMC), or treatment with trans retinoic acid (tRA). (**B**) Immunofluorescence detection of γH2AX and 53BP1 in tRA-treated wt MEFs precultured for 1 hour in the presence of 10 μM ataxia telangiectasia mutated (ATM) (ATMi) or ATR (ATRi) inhibitors. (**C**) Immunofluorescence detection of γH2AX and 53BP1 in tRA-treated wt MEFs in the presence of triptolide (TPL) or 5,6-dichlorobenzimidazole 1-β-d-ribofuranoside. The graphs depict the percentage of cells with ≥3 $\gamma H2AX^{+}$;$53BP1^{+}$ foci (A) in wt untreated (Untr.), MMC-treated, or tRA-treated MEFs, (B) in tRA-treated wt MEFs exposed to ATM (ATMi) or ATR (ATRi) inhibitors, and (C) in tRA-treated wt MEFs in the presence of TPL or DRB ($n$ = 3 biological replicates). (**D**) Cumulative DSBs per chromosome in untreated (Untr.) and tRA-treated MEFs. (**E** to **G**) Number of DNA DSBs per million mapped reads in untreated and tRA-treated MEFs on (E) gene promoters, (F) gene bodies, and (G) intergenic regions. Red dotted line: Average of two biological replicates (light/dark, green: untreated, blue: tRA). (**H**) BLESS signals quantified by quantitative polymerase chain reaction (qPCR) on the tRA-inducible *Cfh*, *Rarb*, and *Hs3st1* gene promoters and on a nontranscribed intergenic region (-) in untreated or tRA-treated MEFs. (**I**) Pie charts representing the number of DSBs on XPF-bound sites in untreated and tRA-treated MEFs. (**J**) Scatter plot of transcription [tRA/untreated fold change (FC)] and DSBs (tRA/untreated FC) levels for XPF-bound (orange) and XPF-nonbound (blue) genes. (**K**). Probability of XPF recruitment by means of $\log^2$RNA × $\log^2$Breaks. (E) to (H) *** indicates the significance at $P$ value of $\leq 10^{-15}$ (Mann-Whitney test).

formation of γH2AX and 53BP1 foci in tRA-treated MEFs (Fig. 3B; as indicated). Further analysis showed that γH2AX and 53BP1 foci persisted in tRA-treated MEFs cultured under serum starvation conditions or following treatment with hydroxyurea, a potent DNA replication inhibitor (fig. S2, E to G). Consistently, flow cytometry analysis revealed no detectable cell cycle differences in tRA-treated MEFs compared to untreated control cells (fig. S3A). Together, these findings indicate that the activation of transcription-associated DDR occurs independently of DNA replication in these cells.

Next, we sought to test whether the formation of γH2AX and 53BP1 foci occurs predominantly during transcription initiation and/or elongation. To do so, we treated tRA-treated MEFs with triptolide (TPL), a small-molecule XPB/TFIIH inhibitor that blocks

Downloaded from https://www.science.org on February 13, 2025

transcription initiation (*44*), or with 5,6-dichlorobenzimidazole 1-β-D-ribofuranoside (DRB), a selective inhibitor of transcription elongation by RNAPII (*45*). As evidenced by immunofluorescence staining, the incorporation of the synthetic uridine derivative BrU into newly synthesized RNA was significantly reduced in the presence of DRB or TPL in tRA-treated cells (fig. S3B). However, the percentage of $\gamma H2AX^+$ $53BP1^+$ MEFs decreased significantly only when tRA-treated MEFs were treated with TPL (Fig. 3C). Together, our findings indicate that, in the absence of exogenous genotoxic stimuli, transcription initiation, but not elongation, triggers an ATM-dependent DDR.

The accumulation of DSBs upon transcription activation in tRA-treated MEFs and the preferential recruitment of XPF to active promoters prompted us to compare the XPF ChIP-Seq profiles with the genome-wide distribution and frequency of DNA DSBs in tRA-treated MEFs. To do so, we first used breaks labeling in situ and sequencing (BLISS) (*46*) to identify processed DNA DSBs across the genome. After filtering out PCR duplications, we find that DNA DSB counts are evenly distributed across mouse chromosomes (Fig. 3D). With further analysis, we found that tRA-treated MEFs accumulate a significantly higher number of DSBs in transcription-associated regions, i.e., in promoters (Fig. 3E) and gene bodies (Fig. 3F) when compared to untreated wild-type (wt) MEFs per 100 ng of genomic DNA (gDNA) ($P = 1.14 \times 10^{-7}$ and $P = 2.39 \times 10^{-7}$, respectively). We also find that tRA treatment is associated with an increase of DNA DSBs in intergenic regions (Fig. 3G). A follow-up BLESS (breaks labeling, enrichment on streptavidin and next-generation sequencing) approach coupled with qPCR on the previously identified bXPF-bound *Rarb*, *Cfh*, and *Hs3st1* gene promoters confirmed a significant increase in DNA DSBs on promoters in tRA-treated MEFs and a pronounced decrease in DNA DSBs in UVC-irradiated tRA-treated MEFs compared to non–UVC-irradiated tRA-treated control cells (Fig. 3H and fig. S3C). There was no increase in DNA DSBs in a transcriptionally inactive genomic region (Fig. 3H) or on the promoter region of the tRA nonresponsive genes *Chordc1*, *Dcaf10*, and *Dhx16* (fig. S3D). Of the 1100 peaks identified in bXPF ChIP-Seq profiles, ~96% contained DNA DSBs (Figs. 1B and 3I; $n = 5803$), with 26% of the identified DNA DSBs detected on bXPF-bound promoters. Upon treatment with tRA, the number of DNA DSBs increased markedly to 9665, corresponding to 91.2% of the 1964 peaks identified in bXPF ChIP-Seq profiles, with 42% of the identified DNA DSBs being detected on bXPF-bound promoters (Figs. 1B and 3I; $n = 9665$). An integrative analysis combining the bXPF ChIP-Seq data with the RNA-Seq profiles and the BLISS-isolated DSBs revealed that there was an increase in the mRNA levels and the number of DSBs in XPF-bound gene targets compared to unbound genes (Fig. 3J and fig. S3, E and F). To test whether the presence of transcription-associated DSBs and/or transcription activation affects the probability of XPF recruitment to gene promoters, we developed a classification model for bXPF binding (bound/unbound) using the automated machine learning tool JAD Bio (*47*) with the logarithms of RNA-Seq, BLISS, and the tRA treatment as predictors. The importance of the interaction term $\log^2$RNA × $\log^2$Breaks is visually verified in Fig. 3K and fig. S3 (G and H), where the distributions of the bXPF-bound and bXPF-unbound sites are clearly distinguished based on this feature. In line with our previous findings, we find that recruitment of bXPF relies on transcription activation and can be predicted by the presence of DNA DSBs, independently of the tRA treatment in MEFs.

## A proteomics strategy reveals bXPF-bound protein partners involved in chromosome organization, transcription, and DNA repair

We reasoned that the selective recruitment of bXPF on promoters reflects possible interactions of ERCC1-XPF with factors associated with transcription initiation and/or transcription-associated DNA damage. To test this, we combined the in vivo biotinylation tagging approach (*32*) with a hypothesis-free, high-throughput proteomics strategy in primary bXPF MEFs. Using high-salt extraction methods, we prepared nuclear extracts from bXPF MEFs and MEFs expressing only the BirA transgene that were subsequently treated with benzonase and ribonuclease A (RNase A); the latter ensures that neither DNA nor RNA mediates the identified protein interactions (Fig. 4A). Nuclear extracts were further incubated with streptavidin-coated beads, and bound proteins were eluted and subjected to Western blot analysis, confirming that bXPF can still interact with its obligatory partner ERCC1 (Fig. 4B). The proteome was first separated into ~12 fractions using one-dimensional (1D) SDS–polyacrylamide gel electrophoresis (SDS-PAGE). The resulting gel bands were then digested, and the resulting peptides were analyzed using high-resolution liquid chromatography–tandem mass spectrometry (nLC-MS/MS) on a hybrid linear ion trap Orbitrap instrument (Fig. 4C). From three biological replicates, which comprised a total of 72 MS runs, we identified a total of 695 proteins, with 607 proteins (87.3%) shared between all three measurements under stringent selection criteria (Fig. 4D). To functionally characterize this dataset, we subjected the 607-shared bXPF-bound proteins to gene ontology (GO) classification. Biological processes (Fig. 4E) or pathways (Fig. 4F) that contained a significantly disproportionate number of proteins relative to the murine proteome were flagged as significantly overrepresented (FDR < 0.05). At this level of confidence, the overrepresented biological processes and pathways involved 77 out of the initial 607 bXPF-bound core proteins. The latter set of proteins also showed a significantly higher number of known protein interactions, with 286 interactions observed compared to an expected 76 interactions by chance (Fig. 4G). This suggests a functionally relevant and highly interconnected protein network. Using this dataset, we were able to discern four major, partially overlapping, bXPF-associated protein complexes involved in (i) chromosome organization ($P \leq 3.2 \times 10^{-37}$, e.g., CTCF, HIST1h1a-e, H1F0, SMARCA5, SMC1A, SMC3, TOP1, TOP2A, and TOP2B), (ii) transcription ($P \leq 2.8 \times 10^{-16}$, e.g., TAF6, TAF10, TAF4A, KLF13, UBTF, TOP1, TOP2A, TOP2B, RBM39, NUP107, NUP133, and NUP153), (iii) gene silencing ($P \leq 8.3 \times 10^{-14}$, e.g., BMS1, GNL3, MDN1, NOP58, UTP15, WDR36, WDR43, WDR75, and XRN2), and (iv) DNA replication ($P \leq 1.2 \times 10^{-12}$, e.g., RCF2, RCF3, RCF4, RCF5, SSRP1, and RBBP6). Pull-down experiments in nuclear extracts of bXPF and control BirA MEFs showed that the endogenous bXPF is in complex with the TATA-associated factors (TAFs) TAF4, TAF6, and TAF10 of the TFIID complex as well as with CTCF and the cohesin subunits SMC1A and SMC3 (fig. S4, A and B), thus confirming our previous findings and the proteomics data shown in this work (*32*, *33*). Likewise, a series of immunoprecipitation experiments showed that ERCC1 is in complex with CTCF and the cohesin SMC1A and SMC3 subunits in primary MEFs (fig. S4C). Together, these

Downloaded from https://www.science.org on February 13, 2025

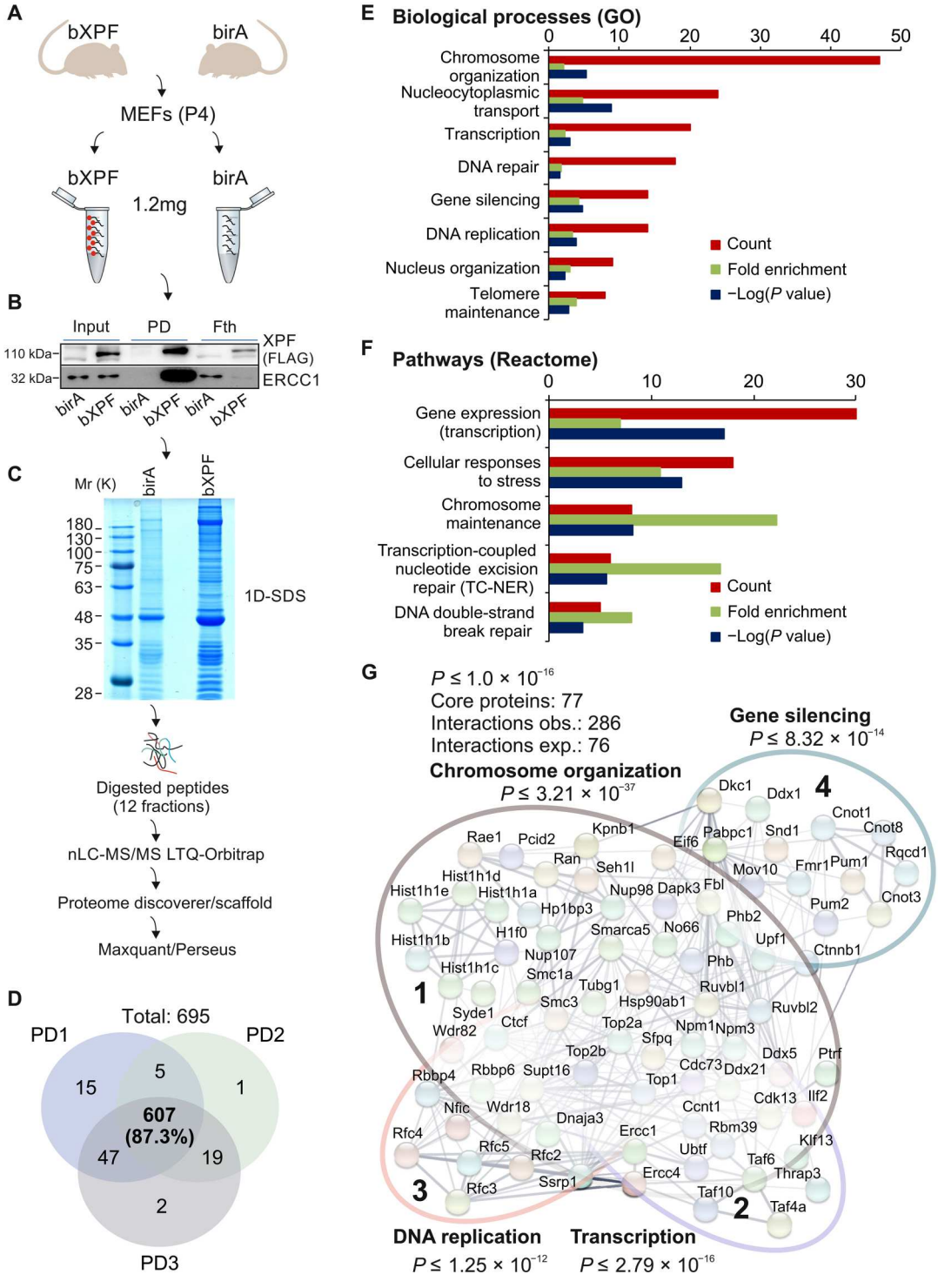


**Fig. 4. XPF interacts with chromatin remodeling and transcription factors.** (**A**) Schematic representation of the high-throughput mass spectrometry analysis performed using nuclear extracts from bXPF and BirA mouse embryonic fibroblasts (MEFs). (**B**) bXPF pulldowns (Fth, flow through; PD, pull down) and Western blot with anti-FLAG and anti-ERCC1 in nuclear extracts derived from bXPF and BirA MEFs. (**C**) Representative 1D gel of streptavidin bead protein eluates derived from bXPF and BirA MEFs. (**D**) Venn diagram of bXPF-bound protein factors from three independent pulldowns (PD) and subsequent MS analyses. (**E**) Significantly overrepresented biological processes [gene ontology (GO)] and (**F**) pathways (Reactome) of the shared 607 bXPF-bound proteins. (**G**) Number of observed (obs.) and expected (exp.) known protein interactions within the core XPF-bound protein set; highlighted circles represent the four major XPF-bound protein complexes involved in chromosome organization, gene silencing, DNA replication, and transcription.

Downloaded from https://www.science.org on February 13, 2025

findings indicate that under native conditions, the great majority of bXPF-bound protein partners are functionally associated with chromatin-associated transactions.

### XPF is in complex with TOP2B and the CTCF/cohesin complex on gene promoters

Type II DNA topoisomerase enzymes (TOP2) catalyze topological changes by strand passage reactions that involve a transient DNA break followed by TOP2-mediated ligation (*48*). Abortive catalysis of TOP2 enzymes can be a major source of spontaneous DSBs. Moreover, TOP2B causes DSBs during the transcription activation of stimulus-inducible genes (*42*, *49*–*51*). The identification of several topoisomerases among the 607 bXPF-bound core proteome (Fig. 4G) and our earlier finding that bXPF is preferentially recruited to promoters upon transcription stimulation (Fig. 1, B and C) prompted us to test whether XPF interacts with TOP2 enzymes. We found that the endogenous bXPF and its partner ERCC1 are in complex with TOP2B but not with TOP2A or TOP1 (Fig. 5, A and B). Instead, the NER structure–specific endonuclease XPG that cleaves the damaged DNA strand on the 3′ side of the lesion did not interact with TOP2B (Fig. 5C). Follow-up immunoprecipitation experiments with an antibody raised against TOP2B confirmed the reciprocity of ERCC1 and XPF interaction with TOP2B but not with TOP1 or TOP2A (Fig. 5D). We recently showed that the ERCC1-XPF complex interacts with the insulator binding protein CTCF and the cohesin subunits SMC1A and SMC3 during mammalian development (*32*). TOP2B is known to colocalize with the evolutionarily conserved CTCF/cohesin binding sites, whereas members of the cohesin complex and CTCF were recently identified as TOP2-interacting proteins in a high-throughput MS screen (*52*). Similar to ERCC1-XPF, we show that TOP2B reciprocally interacts with CTCF and SMC1A and SMC3 (Fig. 5E). The interaction of TOP2B with SMC1A or CTCF is not abolished when ERCC1-XPF is abrogated in *Ercc1*$^{-/-}$ MEFs (Fig. 5F). Confocal imaging in untreated MEFs revealed that whereas bXPF is evenly scattered in the nucleoplasm, TOP2B localizes in clear subnuclear landmarks identified as heterochromatin by 4′,6-diamidino-2-phenylindole (DAPI). However, TOP2B is redistributed throughout the nucleoplasm in tRA-treated MEFs (Fig. 5G and fig. S4D). Next, we sought to test whether TOP2B recruits to bXPF-bound promoters in MEFs. To do so, we performed a series of ChIP-qPCR assays using antibodies raised against TOP2B, TOP2A, and TOP1 on tRA-induced *Rarb*, *Cfh*, and *Hs3St1* promoters that were previously identified in the bXPF ChIP-Seq profiles. Our analysis revealed that TOP2B (Fig. 5H), but not TOP2A (fig. S4E) or TOP1 (fig. S4F), is recruited preferentially to *Rarb*, *Cfh*, and *Hs3St1* promoters. Similar to bXPF, we found that the TOP2B ChIP signals remain unchanged at the tRA nonresponsive *Chordc1* gene promoter or at a nontranscribed genomic region (Fig. 5H; as indicated). ChIP/re-ChIP analysis using antibodies against TOP2B (first ChIP) and ERCC1, FLAG-tagged XPF, or CTCF (second ChIP) showed that these proteins co-occupy the *Rarb*, *Cfh, Hs3St1*, or *Spsb3* gene promoters (Fig. 5, I and J, and fig. S4, G and H). Unlike, however, with bXPF, we find that TOP2B, CTCF, SMC1A, and SMC3 ChIP signals remain unaltered in UVC-irradiated cells. Similar data were observed for TOP1 and TOP2A (fig. S5, A to F; as indicated) as well as for CTCF, SMC1A, and SMC3 in tRA-treated, UVC-irradiated cells (fig. S5, G to I; as indicated). Together, our findings show that, in the absence of exogenous genotoxic insults, the ERCC1-XPF heterodimer is in complex with TOP2B and the CTCF/cohesin complex on gene promoters under conditions that favor transcription.

### XPF processes transcription-associated R-loops in a TOP2B-dependent manner

Naturally occurring R-loops are frequently formed during transcription, when a nascent RNA molecule hybridizes with the DNA template, and the two strands of the DNA duplex reanneal, leaving the nontemplate DNA single-stranded (*53*, *54*). R-loops expose long stretches of single-stranded DNA, which can lead to the spontaneous formation of DSBs or transcription-associated mutagenesis (*55*, *56*). We and others have recently shown that R-loops are actively processed by XPF and XPG (*57*–*62*). Moreover, TOP2 binding and activity has been documented in enhancers, promoters, and gene bodies of actively transcribed genes. This coincides with open chromatin and RNAPII occupancy (*42*, *63*, *64*). These data and our findings that ERCC1-XPF is in complex with TOP2B and the CTCF/cohesin complex on active promoters prompted us to examine their role in R-loop processing and R-loop–induced genome instability. To do so, we first assessed the genome-wide recruitment of XPF-TOP2B-CTCF complex to DNA by overlaying the XPF, TOP2B, and CTCF ChIP-Seq peaks with the RNA-DNA immunoprecipitation (DRIP)–generated, genome-wide R-loop coverage and the BLISS DSB location peaks (Fig. 6A). We found that out of the 1100 XPF ChIP-Seq peaks (column A), 987 overlapped with TOP2B recruitment (column B), and 658 overlapped with CTCF binding on DNA (column C). Almost all R-loops that were bound by XPF also recruited TOP2B and CTCF (Fig. 6B; as indicated). Upon aligning and comparing the DRIP-Seq data with the BLISS data, we found that the number of DSBs per R-loop is significantly higher on promoters compared to introns, intergenic regions, or 3′ and 5′ UTRs (Fig. 6A: columns D to G and Fig. 6C). Using the S9.6 antibody, we demonstrated that transcription activation in tRA-treated MEFs resulted in a substantial accumulation of R-loops (fig. S6A). Treatment with RNase H (RNH), which is known to digest RNA in RNA-DNA hybrids, led to a decrease in R-loops in tRA-treated MEFs, confirming the specificity of the immunostaining approach. S9.6 DRIP, followed by treatment with RNH, further confirmed that in tRA-treated MEFs, R-loops accumulate on the promoters of the tRA-inducible genes *Cfh*, *Rarb*, and *Hs3st1*, but not on the tRA-nonresponsive *Chordc1* promoter (Fig. 6D and fig. S6B). If left unresolved by endogenous RNH, R-loops can be actively processed to generate DNA DSBs, contributing to genome instability (*65*). To test this, we performed a BLESS analysis on both untreated and tRA-treated WT MEFs, in the presence or absence of transfected recombinant RNH, to investigate the role of R-loops in genetic instability and damage. This was motivated by the pronounced increase in DNA DSBs observed in tRA-treated WT MEFs (Fig. 3, E to G). Our results demonstrate that upon transcription activation, tRA-treated MEFs exhibit increased DSB generation on the Rarb promoter compared to untreated controls (Fig. 6E). Notably, the transfection of tRA-treated MEFs with recombinant RNH abolishes the formation of DSBs on the Rarb promoter (Fig. 6E; as indicated). DRIP-Western analysis revealed that TOP2B, in addition to XPF, is recruited to R-loops under native conditions. This recruitment was further enhanced upon transcription activation in tRA-treated MEFs, as demonstrated by our findings (Fig. 6F). Furthermore, both dot blot and immunofluorescence

Downloaded from https://www.science.org on February 13, 2025

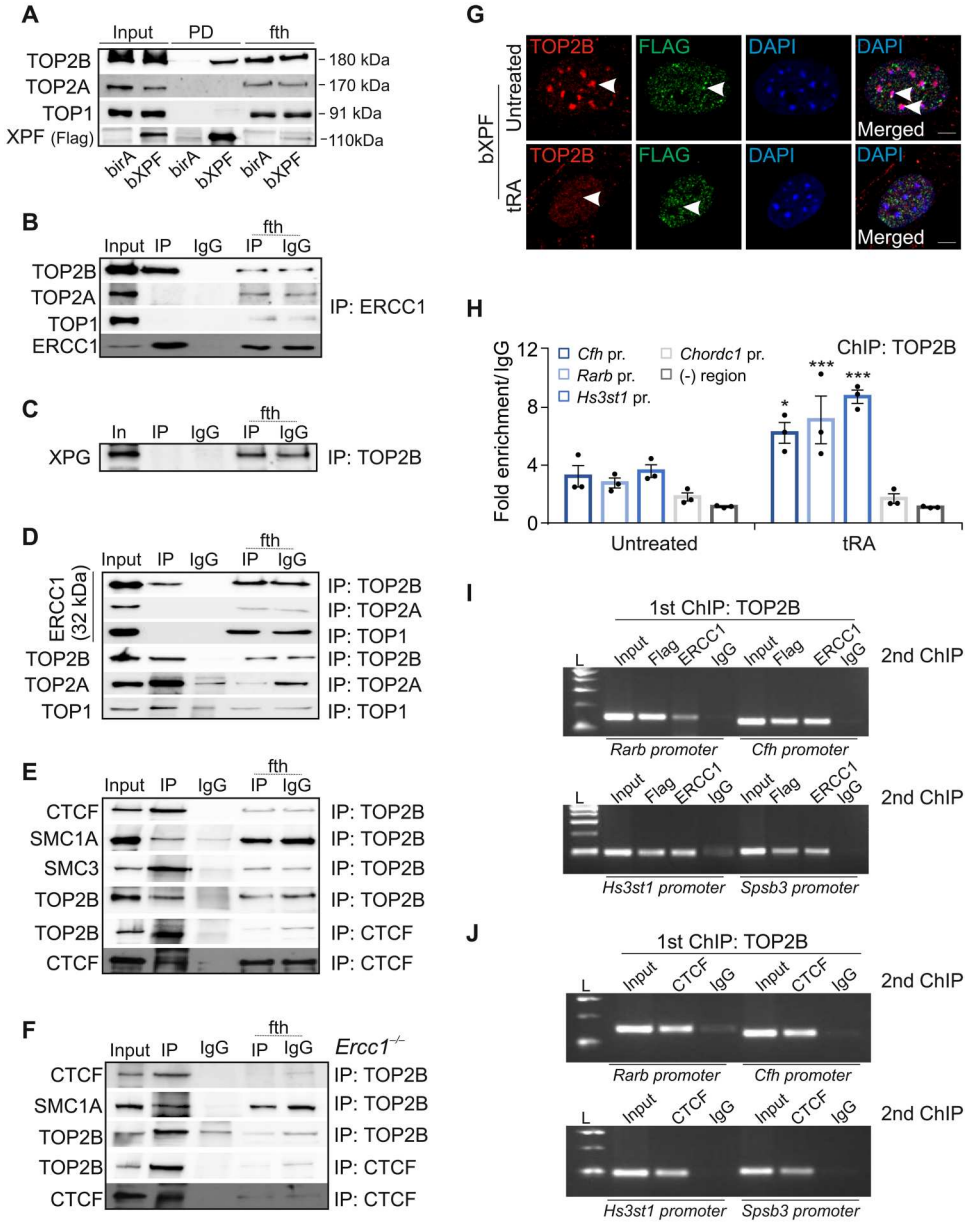


**Fig. 5. XPF is in complex with TOP2B and CTCF on promoters.** (**A**) Pull downs (PD) in bXPF/BirA or BirA mouse embryonic fibroblasts (MEFs), analyzed by Western blotting for TOP2B, TOP2A, and TOP1. (**B**) Coimmunoprecipitation of ERCC1 with TOP2B, TOP2A, and TOP1 in wild type (wt) MEFs. (**C**) Coimmunoprecipitation of TOP2B with XPG in wt MEFs. (**D**) Coimmunoprecipitation experiments of TOP2B, TOP2A, or TOP1 with ERCC1 in wt MEFs. (**E**) Coimmunoprecipitation experiments of TOP2B or CTCF with CTCF, TOP2B, SMC1A, and SMC3 in wt MEFs. (**F**) Coimmunoprecipitation experiments of TOP2B or CTCF with CTCF, TOP2B, and SMC1A in *Ercc1*$^{-/-}$ MEFs. (**G**) Immunofluorescence detection of TOP2B (red) and bXPF (FLAG, green) (arrows) in bXPF MEFs cultured upon basal conditions or upon treatment with trans retinoic acid (tRA). In untreated conditions, arrowheads point to the colocalization of TOP2B with heterochromatin. (**H**) TOP2B ChIP signals on the promoters of *Cfh*, *Rarb*, *Hs3st1*, and *Chordc1* genes and on an intergenic nontranscribed (-) region. (**I**) ChIP with anti-TOP2B and re-ChIP with anti-ERCC1 or Flag-tagged XPF on *Rarb* and *Cfh* gene promoters (top) and on *Hs3st1* and Spsb3 gene promoters (bottom). (**J**) ChIP with anti-TOP2B and re-ChIP with anti-CTCF on *Rarb* and *Cfh* gene promoters (top) and on *Hs3st1* and Spsb3 gene promoters (bottom).

analyses demonstrated that MEFs lacking TOP2B (*Top2*β$^{-/-}$) accumulated significantly higher levels of RNA-DNA hybrids compared to WT controls (Fig. 6G and fig. S6C). The accumulation of co-transcriptional R-loops is known to interfere with transcription (*53*). Consistently, the mRNA levels of tRA-responsive genes were substantially decreased in tRA-treated *Top2*β$^{-/-}$ MEFs compared to corresponding untreated controls, following an increase in R-loops (fig. S6D). The low number of bXPF-Seq peaks observed in UVC-irradiated bXPF MEFs suggests that XPF has a high affinity for UVC-induced DNA lesions, which are randomly distributed throughout the mammalian genome. Likewise, we found that XPF is released from R-loops in tRA-treated MEFs upon UVC irradiation (fig. S7A). Instead, DRIP-Western analysis revealed that XPF recruitment to R-loops was increased when tRA-treated MEFs

Downloaded from https://www.science.org on February 13, 2025

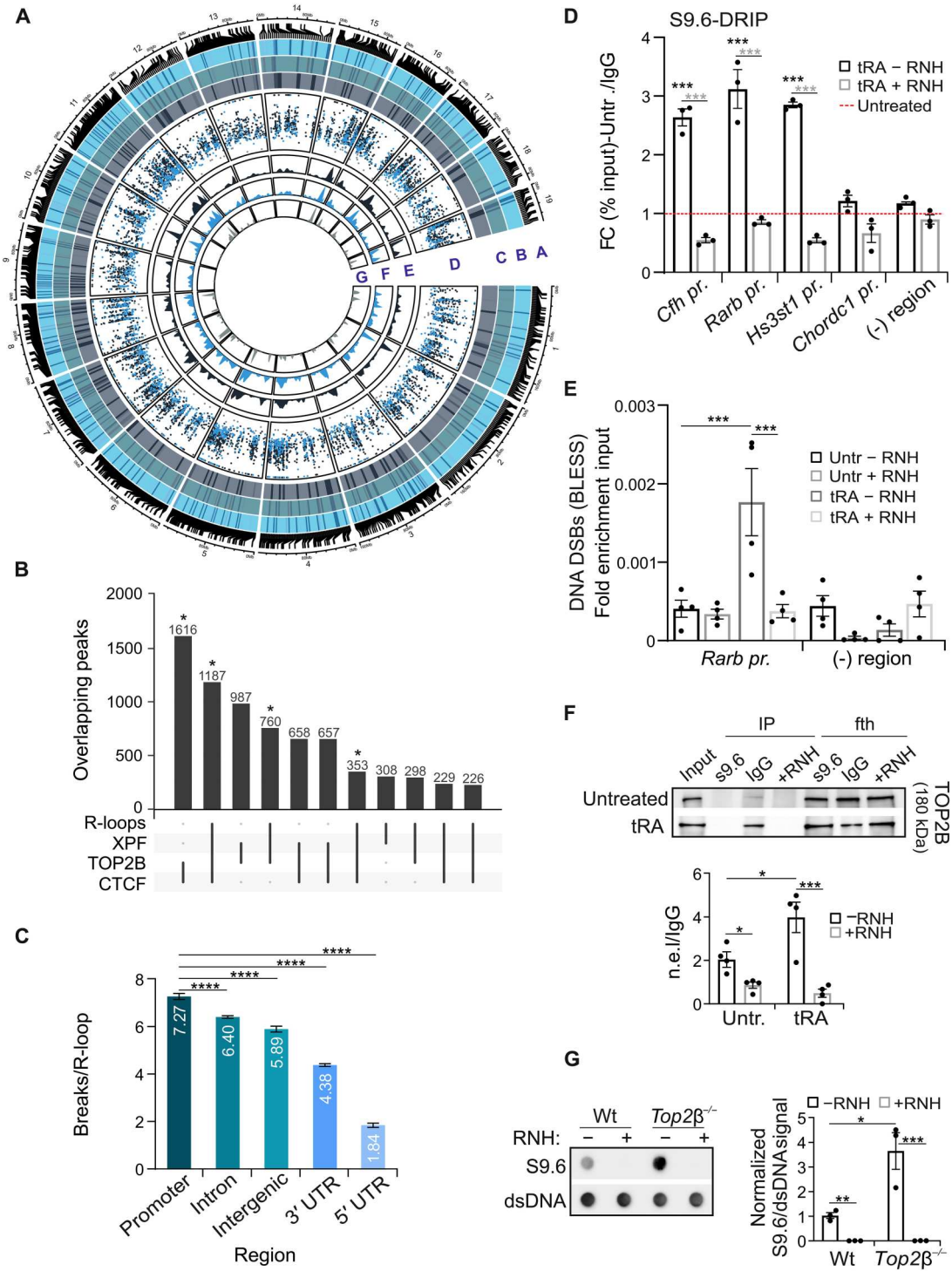


**Fig. 6. XPF is recruited with TOP2B and CTCF on transcription-induced R-loops.** (**A**) Circos plot displaying (A) XPF peaks, (B and C) overlapping peaks of XPF with TOP2B and CTCF, (D) DSBs per R-loop in promoters, 3′UTR, and 5′UTR (dark blue, light blue, and gray, respectively), and (E to G) numbers of DSBs in R-loops in promoters, 3′UTR, and 5′UTR. (**B**) Integrated comparative analysis of R-loops, XPF, TOP2B, and CTCF. Asterisk: Numbers divided by a magnitude of 1000. (**C**) Genomic distribution of DSBs in R-loops. DSBs per R-loop in a specific genomic location category (promoter, intron, intergenic, 3′UTR, and 5′UTR, respectively) are shown (unpaired two-tailed *t* test). (**D**) DRIP analysis of *Cfh*, *Rarb*, *Hs3st1*, and *Chordc1* gene promoters and of an intergenic nontranscribed (-) region with or without RNase H (RNH) in untreated and trans retinoic acid (tRA)–treated mouse embryonic fibroblasts (MEFs). The *P* values are depicted as asterisks, gray represents statistical significance between tRA − RNH and tRA + RNH conditions, and black represents statistical significance between tRA − RNH and untreated conditions. RNA-DNA immunoprecipitation (DRIP) signals are shown as fold change (FC) of percentage of (%) input of antibody over percentage of input of control antibody (IgG). (**E**) BLESS signals on the *Rarb* gene promoter and on an intergenic (-) region upon recombinant RNH transfection in untreated and tRA-treated MEFs. (**F**) DRIP followed by Western blotting for TOP2B in wild type (wt) and tRA-treated MEFs with or without RNH treatment. (**G**) S9.6 and dsDNA dot blot analysis of genomic DNA from *Top2*β$^{-/-}$ MEFs and corresponding WT control cells with or without RNH treatment.

Downloaded from https://www.science.org on February 13, 2025

were further exposed to Illudin S, a natural toxin known to cause transcription-blocking lesions and RNA-DNA hybrids (fig. S7A) (*60*, *66*). Consistent with this finding, anti-S9.6 immunofluorescence experiments demonstrated an increase in R-loop accumulation in tRA/Illudin S–treated MEFs (fig. S7B). Next, we sought to investigate the effect of TOP2B on the recruitment of XPF to transcription-associated R-loops. We found that the XPF ChIP signals were substantially reduced on the promoters of tRA-induced *Cfh*, *Rarb*, and *Hs3st1* genes in *Top2*β$^{-/-}$ cells (Fig. 7A). In addition, we observed a decrease in ERCC1 recruitment in *Top2*β$^{-/-}$ cells (fig. S7C).

DRIP–Western blotting for XPF demonstrated a consistent decrease in the recruitment of XPF to R-loops in *Top2*β$^{-/-}$ cells compared to the corresponding WT control MEFs (Fig. 7B). The

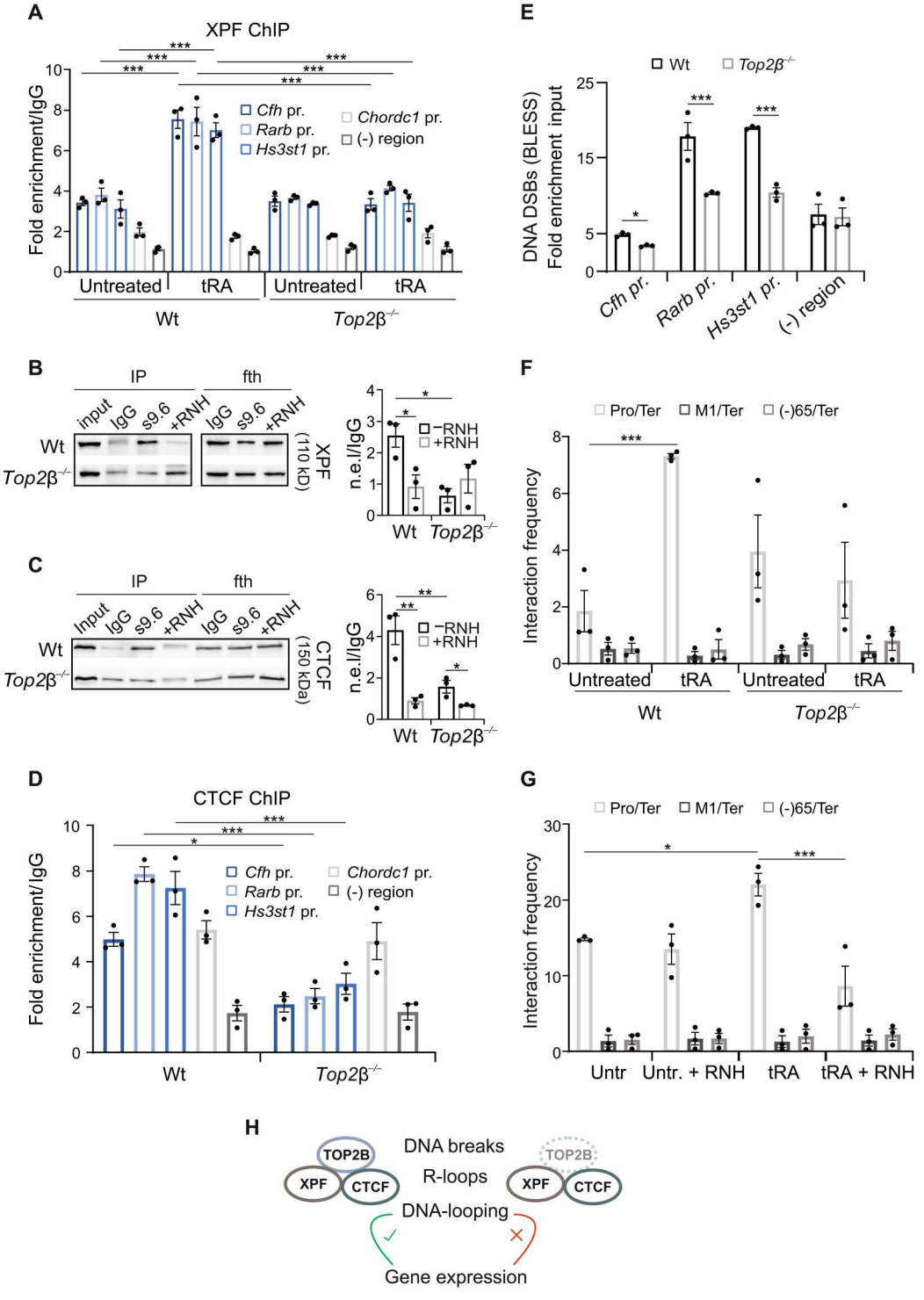


**Fig. 7. CTCF-mediated DNA looping requires the presence of TOP2B and R-loops.** (**A**) XPF ChIP signals on the promoters of *Cfh*, *Rarb*, *Hs3st1*, and *Chordc1* genes and on an intergenic nontranscribed (-) region in untreated and trans retinoic acid (tRA)–treated *Top2*β$^{-/-}$ and wild type (wt) mouse embryonic fibroblasts (MEFs). RNA-DNA immunoprecipitation (DRIP) followed by Western blotting for (**B**) XPF or (**C**) CTCF in wt and tRA-treated MEFs with or without RNase H (RNH) treatment. (**D**) CTCF ChIP signals on the promoters of *Cfh*, *Rarb*, *Hs3st1*, and *Chordc1* genes and on an intergenic nontranscribed (-) region in *Top2*β$^{-/-}$ and WT MEFs. (**E**) BLESS signals quantified by qPCR on the tRA-inducible *Cfh*, *Rarb*, and *Hs3st1* gene promoters and on an intergenic nontranscribed (-) region in *Top2*β$^{-/-}$ and WT MEFs. (**F**) Interaction frequency, quantified by 3C-qPCR, between the *Rarb* gene terminator (Ter) and the gene promoter (Pro), an intronic region (M1), or an intergenic region −65 kb upstream to the transcription start site (TSS) (-65) in untreated or tRA-treated *Top2*β$^{-/-}$ and wt MEFs. (**G**) Interaction frequency between the *Rarb* gene terminator (Ter) and the gene promoter (Pro), an intronic region (M1), or an intergenic region −65 kb upstream to the TSS (-65) in untreated or tRA-treated wt MEFs, with or without the transfection of recombinant RNH. (**H**) XPF interacts with TOP2B and the CTCF/cohesin complex, on active gene promoters, leading to DSB accumulation and R-loop processing for transcription activation. Abrogation of TOP2B leads to the diminished recruitment of XPF, CTCF, and the cohesin subunits to promoters of actively transcribed genes and R-loops and the concurrent impairment of CTCF-mediated DNA looping.

Downloaded from https://www.science.org on February 13, 2025

recruitment of XPF to R-loops was also diminished in tRA-treated MEFs when they were exposed to merbarone, a catalytic inhibitor that prevents TOP2-mediated DNA cleavage while not affecting protein-DNA binding (fig. S7D) (*67*). These findings were consistent with the reduced chromatin recruitment of XPF on the promoters of tRA-induced genes in the presence of merbarone (fig. S7E). In agreement with the observed interaction of XPF with TOP2B and the CTCF/cohesin complex, as well as with the significant overlap of XPF, TOP2B, and CTCF ChIP-Seq profiles with R-loops throughout the genome (Fig. 6, A and B), we found that CTCF, SMC1A, and SMC3 are recruited to RNA-DNA hybrids in untreated wt MEFs (Fig. 7C and fig. S7F). When TOP2B is abrogated, the recruitment of CTCF, SMC1A, and SMC3 on R-loops (Fig. 7C and fig. S7F) and gene promoters (Fig. 7D and fig. S8, A and B) is reduced. Consistent with the reduced recruitment of XPF on R-loops, we found that DSBs are significantly reduced on the promoters of tRA-induced *Cfh*, *Rarb*, and *Hs3st1* genes in *Top2*β$^{-/-}$ cells (Fig. 7E) or when TOP2B activity is abolished by merbarone in tRA-treated wt MEFs (fig. S8C), suggesting that XPF-mediated processing of R-loops into DSBs is TOP2B dependent. Next, we assessed the role of CTCF/cohesin in relation to the XPF/TOP2B complex. Recent findings revealed that the catalytic activity of XPF is required, along with XPG, for CTCF recruitment and gene looping upon transcription induction (*31*). We used the quantitative chromosome conformation capture (q3C) technique to confirm the DNA looping event between the promoter and the terminator of the *Rarb* gene, upon tRA transcription activation in wt MEFs (Fig. 7F). In agreement with the reduced recruitment of the CTCF/cohesin complex on the *Rarb* promoter upon TOP2B depletion (Fig. 7D and fig. S8, A and B), we observed that the *Rarb* promoter-terminator interaction is decreased in *Top2*β$^{-/-}$ MEFs compared to wt controls (Fig. 7F), presumably leading to the impaired mRNA expression of the *Rarb* gene seen in these cells (fig. S6D). Consistently, we find that the previously described CTCF-mediated DNA looping events that activate the major histocompatibility complex class II (MHC-II) *Aa* and *Eb1* genes in interferon-γ (IFN-γ)–treated wt MEFs (fig. S8D) (*68*), as well as the interactions between the developmentally regulated HoxC genes (fig. S8E) (*69*), are impaired in *Top2*β$^{-/-}$ MEFs. In line, the respective mRNA levels of *Aa*, *Eb1*, *HoxC10*, and *HoxC13* are decreased (fig. S8F). Moreover, it was recently shown that the *HOTTIP*-mediated induction of R-loops in CTCF binding sites regulates CTCF/cohesin binding and coordinates boundary function (*70*). These data prompted us to test whether co-transcriptional R-loops are required for the CTCF/cohesin-mediated DNA looping. q3C experiments showed that upon transfection of WT MEFs with recombinant RNH, the juxtaposition of the *Rarb* gene promoter with the gene terminator was impaired (Fig. 7G). Together, our findings indicate that TOP2B is required for the recruitment of XPF and the CTCF/cohesin complex to R-loops on gene promoters and that co-transcriptional R-loops are processed into DSBs on the promoters of actively transcribed genes. Moreover, the abrogation of the XPF/TOP2B/CTCF/cohesin complex leads to impaired R-loop processing and DNA looping, necessary for proper gene expression.

## DISCUSSION

DNA damage events are randomly or purposely generated during transcription (*2*) supporting the notion that mRNA synthesis is a potentially hazardous process. Consistently, we observed that, in the absence of exogenous genotoxic insults, transcription induces the accumulation of DSBs genome-wide, and preferentially on active promoters, leading to the formation of γH2AX and 53BP1 foci in MEFs in a TOP2B-dependent manner. This is also in line with the known role of TOP2, which binds to DNA and generates transient DSBs on promoters to alleviate the topological constraints generated by RNAPII as it moves along the DNA template (*49*). When transcription is induced, we show that XPF recruits preferentially at and upstream of the TSS of actively transcribed genes. Naturally occurring R-loops generated during transcription are actively processed by XPF and XPG into DSBs (*57*, *58*, *60*), in agreement with our findings that most of these genes contain activity-induced DNA breaks. Here, we additionally show that XPF is recruited to RNA-DNA hybrids on gene promoters, together with TOP2B and CTCF. TOP2B plays an important role in this process, as its depletion results in reduced recruitment of both XPF and CTCF, an increase in R-loop accumulation, and a decrease in transcription-associated DSBs on promoters, which leads to disrupted mRNA expression of the corresponding genes (Fig. 7H).

TOP2 cleaves and rejoins DNA ends, through the generation of a transient DSB (*71*). In some instances, the TOP2-DNA cleavage complex can become stabilized, leading to abortive catalysis and TOP2 trapping (*72*). Tyrosyl-DNA phosphodiesterase 2 or MRE11 removes 5′ TOP2 adducts to restore ligatable DNA ends for DSB repair (*73*, *74*). In this respect, the interaction of TOP2B with ERCC1-XPF could enable XPF to cleave the 3′ overhangs for successful end resection during HR, in addition to resolving R-loops (*75*). Recruitment of XPF to TOP2B sites could also trim non-complementary 3′ tails before resealing, during NHEJ repair of activity-induced DSBs (*5*, *14*, *42*, *76*, *77*). TOP2B binds and catalyzes DSBs at DNA sites that are prone to G-quadruplex secondary structures (*78*–*80*). Such G4 structures often coexist with RNA-DNA hybrids in transcribed G-rich loci (*81*) and either allow for enhanced transcription by stabilizing R-loops (*82*, *83*) or obstruct DNA replication/transcription threatening genome stability (*81*, *84*). In this respect, it is attractive to speculate that the TOP2B-XPF complex also processes G4/R-loops at actively transcribed G-rich sequences.

A central aspect of our findings is that R-loops are required for the intrachromosomal juxtaposition of promoter and terminator sequences when transcription is activated in the *Rarb2* gene. In line, we also show that in *Top2b*$^{-/-}$ MEFs, where R-loops are reduced, the CTCF-mediated DNA looping events that activate the MHC-II *Aa* and *Eb1* genes upon IFN-γ induction (*68*), as well as the interactions between the developmentally regulated HoxC genes, are impaired (*69*). This, in turn, leads to the reduction of the respective mRNA levels of *Aa*, *Eb1*, *HoxC10*, and *HoxC13* genes. These results altogether suggest that regulatory R-loops might be necessary for CTCF-mediated DNA looping and optimal gene activity. However, it remains to be seen whether R-loops are necessary in all looping events, for the formation of topologically associating domains (TADs) or chromatin accessibility. In this scenario, the proper resolution of RNA-DNA hybrids by XPF would be crucial for uninterrupted RNAPII-guided mRNA synthesis.

It has been challenging to delineate how DNA damage drives the onset of tissue-specific, developmental defects in NER progeroid syndromes. Here, we provide evidence for a functional link

Downloaded from https://www.science.org on February 13, 2025

between XPF with TOP2B and CTCF-cohesins and active transcription. The complex could promote the proximity of XPF-bound promoters with enhancers (*85*), facilitate R-loop–directed chromatin looping (*31*, *32*), and position TOP2B at TAD boundaries (*52*), allowing the selective regulation of gene expression in vivo.

## MATERIALS AND METHODS

### Animal models and primary cells

The generation and characterization of bXPF and NER-deficient mice has been previously described (*32*). Animals were kept on a regular diet and housed at the Institute of Molecular Biology and Biotechnology (IMBB) animal house, which operates in compliance with the "Animal Welfare Act" of the Greek government, using the *Guide for the Care and Use of Laboratory Animals* as its standard. As required by Greek law, formal permission to generate and use genetically modified animals was obtained from the responsible local and national authorities (6ΛΤΑ7ΛΚ-ΚΚΘ). All animal studies were approved by independent Animal Ethical Committees at Foundation for Research and Technology-Hellas (FORTH) and Biomedical Sciences Research Center (BSRC) Al. Fleming. The animals used were WT, bXPF, BirA: *Mus musculus*, strain C57Bl/6, *Ercc1*$^{-/-}$: *M. musculus*, strain FVB/nj:C57BL/6j. Cell lines used:*Top2*β$^{-/-}$, and respective control MEFs were generated and provided by C. Austin. Primary MEFs were isolated from E13.5d animals and cultured in standard medium containing Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum (FBS), streptomycin (50 μg/ml), penicillin (50 U/ml; Sigma-Aldrich), and 2 mM L-glutamine (Gibco). Cells were rinsed with phosphate-buffered saline (PBS); exposed to UVC irradiation (10 J/m$^2$), MMC (10 μg/ml, 4 hours) (AppliChem), tRA (10 μM, 16 hours) (Sigma-Aldrich), merbarone (2 μM, 16 hours) (Sigma-Aldrich), TPL (62 nM, 16 hours), DRB (6.25μμ, 16 hours), Illudin S (30 ng/ml, 3 hours), or hydroxyurea (650 μM, 16 hours); and cultured at 37°C before subsequent experiments. Preincubation with ATM inhibitor (10 μM) and ATR inhibitor (10 μM) started 1 hour before genotoxic treatments and lasted throughout the experiment. For the protein transfection experiments (Pierce Protein Transfection Reagent, Thermo Fisher Scientific), 40 U of recombinant RNH [5 U/μl; New England Biolabs (NEB)] was used according to the manufacturer's instructions.

### Immunofluorescence, antibodies, and Westerns blots

Immunofluorescence experiments were performed as previously described (*32*, *59*, *60*, *86*). Briefly, cells (primary MEFs) were fixed in 4% formaldehyde, permeabilized with 0.5% Triton X-100, and blocked with 1% bovine serum albumin (BSA). After 1-hour incubation with primary antibodies, secondary fluorescent antibodies were added and DAPI was used for nuclear counterstaining. Samples were imaged with an SP8 confocal microscope (Leica). For local DNA damage infliction, cells were UVC-irradiated (10 J/m$^2$) through isopore polycarbonate membranes containing 3-μm-diameter pores (Millipore) and experiments were performed 2 hours after UVC irradiation. Antibodies against HA (Y-11, Western blotting (wb): 1:500), ERCC1 (D-10, wb: 1:500, Immunofluorescence (IF): 1:50), TOP2A (C-15, wb: 1:200, IF: 1:50), XPG (sc-12558, wb: 1:200), and p53 (sc-6243, wb: 1:500) were from Santa Cruz Biotechnology. γH2AX (05-636, IF: 1:12000) was from Millipore. γH2AX (22551, wb: 1:1000), β-tubulin (ab6046, wb: 1:5000), and fibrillarin (ab5821, wb: 1:5000) were from Abcam. TOP1 (NBP1-30482, wb: 1:1000, IF: 1:50), TOP2B (NB100-40842, wb: 1:1000), and 53BP1 (NB100-304, IF: 1:300) were from Novus Biologicals. TOP2B (20549-I-AP, IF: 1:50) and XPG (11331-1-AP) were from Proteintech. TAF-4 (TAF2B9, wb: 1:500, IF: 1:50), TAF-6 (TAF2G7, wb: 1:500), and TAF-10 (6TA-2B11, wb: 1:500) were from ProteoGenix. Streptavidin–horseradish peroxidase (wb: 1:12,000) was from Upstate Biotechnology. pATM (wb: 1:1000, IF: 1:1000) was from Rockland. pATR (wb: 1:1000, IF: 1:500) was from Genetex. FLAGM2 (F3165, wb 1:2.000, F1804, IF: 1:1000) was from Sigma-Aldrich. Anti-BrdU (5-bromo-2′-deoxyuridine) antibody (555627) was from BD Pharmingen. Antibodies against phospho-p53 (9284, wb: 1:500), phospho-Chk1 (Ser345) (2348, wb: 1:400), and Chk1 (2G1D5) (2360, wb: 1:1000) were from Cell Signaling Technology. For S9.6 immunostainings, fixed cells were incubated with RNase T1 (4000 U; 01218429, Thermo Fisher Scientific), or RNase T1 and RNase III (3 U; AM2290, Ambion) at 37°C for 45 min with or without RNH (20 U, 5 U/μl; M0297, NEB) (*87*). In the figures, a gray line depicts the 5-μm scale bar.

### Flow cytometry and transcription assays

DNA transcription sites were labeled as previously described (*60*). Briefly, cells were washed with 20 mM tris-HCl, 25% glycerol, 5 mM $MgCl_2$, and 0.5 mM EGTA for 10 min on ice, permeabilized with 0.5% Triton X-100 in glycerol buffer on ice for 3 min, and incubated at room temperature (RT) for 30 min with 50 mM tris-HCl (pH7.4), 10 mM $MgCl_2$, 150 mM NaCl, 25% glycerol, RNase inhibitor (25 U/ml), and protease inhibitors, supplemented with 0.5 mM adenosine 5′-triphosphate (ATP), cytidine 5′-triphosphate (CTP), guanosine 5′-triphosphate (GTP), and 0.2 mM Bromo-uridine-triphosphate (BrUTP). Cells were then fixed with 4% formaldehyde in PBS on ice for 10 min. Immunofluorescence with anti-BrdU antibody was performed as described above. For cell cycle analyses, cells were fixed with 70% ethanol for 30 min, washed with PBS, treated with RNase A (1 mg/ml) at 37°C for 30 min, and stained with propidium iodide (20 mg/ml) for 1 hour at RT.

### ChIP, coimmunoprecipitation, and chromatin pull-down assays

For coimmunoprecipitation assays, nuclear protein extracts from primary MEFs were prepared as previously described (*32*) using the high-salt extraction method [10 mM Hepes-KOH (pH 7.9), 380 mM KCl, 3 mM $MgCl_2$, 0.2 mM EDTA, 20% glycerol, and protease inhibitors]. Nuclear lysates were diluted threefold by adding ice-cold HENG buffer [10 mM Hepes-KOH (pH 7.9), 1.5 mM $MgCl_2$, 0.25 mM EDTA, and 20% glycerol] and precipitated with antibodies overnight at 4°C followed by incubation for 3 hours with protein G Sepharose beads (Millipore). Normal mouse, rabbit, or goat immunoglobulin G (IgG; Santa Cruz Biotechnology) was used as a negative control. Immunoprecipitates were washed five times [10 mM Hepes-KOH (pH 7.9), 300 mM KCl, 0.3% NP-40, 1.5 mM $MgCl_2$, 0.25 mM EDTA, 20% glycerol, and protease inhibitors], eluted, and resolved on 8 to 12% SDS-PAGE. The input and flow-through are $^1/_{20}$ of the extract used. Pulldowns were performed with 1.2 mg of nuclear extracts using M-280 paramagnetic streptavidin beads (Invitrogen) as previously described (*32*). For ChIP assays, primary cells (MEFs) were cross-linked at RT for 2.5 min with 1% formaldehyde. Chromatin was prepared and sonicated

Downloaded from https://www.science.org on February 13, 2025

on ice for 15 min using Covaris S220 focused ultrasonicator. Samples were immunoprecipitated with antibodies (5 to 8 μg) overnight at 4°C followed by incubation for 3 hours with protein G–Sepharose beads (Millipore) and washed sequentially. The complexes were eluted, and the cross-linking was heat-reversed. Purified DNA fragments were analyzed by sequencing or qPCR using sets of primers targeting different regions of tRA-responsive genes. ChIP re-ChIP experiments were performed as described above with the following modifications: After the first immunoprecipitation and washing, complexes were eluted with 10 mM dithiothreitol, 1% SDS in Tris-EDTA (TE) buffer for 30 min. Eluted samples were diluted 1:20 with re-ChIP buffer [10 mM tris-HCl (pH 8), 1 mM EDTA, 150 mM NaCl, 0.01% SDS, and 1% Triton X-100] and immunoprecipitated overnight with the second antibody. The primers used were as follows: Rarb, GGGAGTTTTTAAGCGCTGTG (forward) and ACCACTTCTGTCACACGGAAT (reverse); Hs3st1, GCCTTGTTGGCTCTGGTACT (forward) and GCAGAA ATCGGGTGCTTAAC (reverse); Cfh, GCAAGGGCTGGATTTCA TAA (forward) and ATGGGTGTTGGTCCTGAAAA (reverse); neg, GAGTGCACATGTCTGTCCTCGG (forward) and CTCCCA GGGTTGAAGCTCTTGA (reverse); Chordc1, GCAGTCCGGTA GGAAATCTG (forward) and CCGGTACTGCTTCAGGAATTT (reverse); and Spsb1, CTGGGTTTCCTAGCGTTGAG (forward) and GGGCTACAGAGTTCGCAAAG (reverse). ChIP signals in the figures are shown as fold enrichment of percentage input of sample over percentage input of control.

### DRIP and DRIP-Western analysis

DRIP analysis was based on ChIP analysis with some modifications. DRIP analysis was performed without a cross-linking step. Nuclei were isolated using 0.5% NP-40 buffer. Isolated nuclei were resuspended in TE buffer supplemented with 0.5% SDS and 100 mg of proteinase K. Genomic DNA was isolated after the addition of potassium acetate (1 M) and isopropanol precipitation. DNA was sonicated on ice for 3 min using a Covaris S220 focused ultrasonicator. Samples were treated with RNH (10 U/5 μg of DNA) at 37°C overnight. Samples were immunoprecipitated with S9.6 antibodies (8 μg of antibody/5 μg of DNA) overnight at 4°C followed by incubation for 3 hours with protein G–Sepharose beads (Millipore) and washed sequentially. The complexes were eluted, and purified DNA fragments were analyzed by qPCR using sets of primers targeting different regions of related genes. DRIP signals are shown as FC of % input of S9.6 antibody over % input of control antibody (IgG). DRIP-Western analysis was performed as described previously (*60*, *88*). Briefly, non–cross-linked cells were lysed in 0.5% NP-40 buffer for 10 min on ice. Pelleted nuclei were lysed in resuspension buffer [10 mM tris-HCl (pH 7.5), 200 mM NaCl, and 2.5 mM $MgCl_2$] with 0.2% sodium deoxycholate [NaDOC, 0.1% SDS and 0.5% Triton X-100], and extracts were sonicated for 10 min (Diagenode Bioruptor). Extracts were then diluted 1:4 in RSB with 0.5% Triton X-100 (RSB + T) and subjected to immunoprecipitation with the S9.6 antibody (8 μg of antibody/5 μg of DNA), bound to protein A dynabeads (Invitrogen), and preblocked with BSA/PBS (1 mg/ml) for 1 hour. IgG antibodies were used as control. RNH (PureLink, Invitrogen) was added before immunoprecipitation as in DRIP. Beads were washed four times with RSB + T and twice with RSB and eluted in 1× Laemmli.

### Quantitative chromosome conformation capture

q3C was performed as described in (*31*). Briefly, cells were cross-linked with 2% formaldehyde for 10 min at RT. Chromatin was digested in rCutSmart (NEB, B6004) by 400 U of enzyme Hind III (NEB, R3104) for the *Rarb* and *HoxC* genes or Bgl II (NEB, R0144) for the MHC-II genes. The restriction enzyme was denatured, diluted in ligation buffer, and incubated with T4 DNA ligase (NEB, M0202) for 16 hours at RT. The cross-linking was reversed at 56°C, and DNA fragments were purified. Undigested DNA or digested, unligated DNA was used as negative control. The endogenous *Xpb* locus that has been reported to adopt the same spatial conformation in different tissues was used as an internal positive control. All q3C results were normalized by data from the *Xpb* locus, controlling for changes in nuclear size, chromatin density, and cross-linking efficiency. DNA templates (100 ng) were used for the PCRs with specific primers as anchors in combination with other oligonucleotides designed for each of the restriction fragments (*31*, *68*, *69*).

### MS studies

Proteins eluted from the beads were separated by SDS-PAGE electrophoresis on a 10% polyacrylamide gel and stained with colloidal blue silver [Thermo Fisher Scientific, USA; (*69*)]. SDS-PAGE gel lanes were cut into 2-mm slices and subjected to in-gel reduction with dithiothreitol and alkylation with iodoacetamide and digested with trypsin (sequencing grade; Promega), as described previously (*89*, *90*). Peptide mixtures were analyzed by nLC-ESI-MS/MS on an LTQ-Orbitrap XL coupled to an Easy nLC (Thermo Fisher Scientific). The sample preparation and the nLC-ESI-MS/MS analysis were performed as previously described (*91*) with minor modifications. Briefly, the dried peptides were dissolved in 0.5% formic acid aqueous solution, and the tryptic peptide mixtures were separated on a reversed-phase column (Reprosil Pur C18 AQ, Dr. Maisch GmbH), fused silica emitters 100 mm long with a 75 μm internal diameter (Thermo Fisher Scientific, USA) packed in-house using a packing bomb (Loader kit SP035, Proxeon). Tryptic peptides were separated and eluted in a linear water-acetonitrile gradient and injected into the MS.

### RNA-Seq and qPCR studies

Total RNA was isolated from cells using a Total RNA Isolation kit (Qiagen) as described by the manufacturer. For RNA-Seq studies, libraries were prepared using the Illumina TruSeq mRNA stranded sample preparation kit. Library preparation started with 1 μg of total RNA. After poly-A selection (using poly-T oligo–attached magnetic beads), mRNA was purified and fragmented using divalent cations under elevated temperature. The RNA fragments underwent reverse transcription using random primers. This is followed by second-strand complementary DNA (cDNA) synthesis with DNA polymerase I and RNH. After end repair and A-tailing, indexing adapters were ligated. The products were then purified and amplified (14 PCR cycles) to create the final cDNA libraries. After library validation and quantification (Agilent 2100 Bioanalyzer), equimolar amounts of library were pooled. The pool was quantified by using the Peqlab KAPA Library Quantification Kit and the Applied Biosystems 7900HT Sequence Detection System. The pool was sequenced by using an S2 flowcell on the Illumina NovaSeq6000 sequencer and the 2 × 100–nucleotide (nt) protocol. qPCR was performed with a Bio-Rad 1000 series thermal cycler according to the

Downloaded from https://www.science.org on February 13, 2025

instructions of the manufacturer (Bio-Rad) as previously described (*32*). The primers used are as follows: Rarb, CATGCTGCAGGAA AAGGCTC (forward) and GCTGGTACTCTGTGTCTCGA (reverse); Hs3st1, GTGAGTGCCTGTGTCCCTTC (forward) and TGCCAATTACTGAGTCGCGT (reverse); Cfh, TCCTGGGAC TACCTTCGTTG (forward) and GCAGAGTCTCCATTCTCCACA (reverse); Spsb1, TGCGCTACTTGAACGGACTT (forward) and CACTGGTAGAGGAGGTAGGCT (reverse); and Chordc1, GCA GTCCGGTAGGAAATCTG (forward) and CCGGTACTGCTTCA GGAATTT (reverse).

### sBLISS and BLESS

To map DNA DSBs genome-wide, we applied an adapted setup of the BLISS method (*46*). In suspension BLISS (sBLISS), processed, TOP2B-free, DSB ends are in situ blunted and ligated to specialized BLISS adapters that enable selective linear amplification of the genomic sequences at the DSB ends, via T7-driven in vitro transcription. Briefly, after cell treatment and before fixation, cells were washed, trypsinized, and resuspended in prewarmed PBS supplied with 10% FBS, ensuring single-cell suspensions. Then, the cells were counted and diluted to $10^6$ cells/ml and fixed with 4% paraformaldehyde aqueous solution (Electron Microscopy Sciences, #15710, formaldehyde methanol-free) for 10 min at RT. Paraformaldehyde was quenched with 2 M glycine at a final concentration of 125 mM for 5 min at RT, while gently rotating, and for an additional 5 min on ice. Fixed cells were washed with ice-cold PBS and pelleted by centrifuging at 100 to 400 g for 10 min at 4°C. For in situ DSB labeling, $10^6$ fixed cells were incubated in a lysis buffer [10 mM tris-HCl, 10 mM NaCl, 1 mM EDTA, and 0.2% Triton X-100 (pH 8)] for 60 min on ice, and the nuclei were thereafter permeabilized with a prewarmed permeabilization buffer [10 mM tris-HCl, 150 mM NaCl, 1 mM EDTA, and 0.3% SDS (pH 8)] for 60 min at 37°C. After pelleting, the nuclei were washed twice with prewarmed 1× CutSmart buffer (NEB, #B7204) supplemented with 0.1% Triton X-100 (1× CS/TX100). To prepare the DSB ends for BLISS adapter ligation, the DSB ends were blunted with the NEB's Quick Blunting Kit (NEB, #E1201) according to the manufacturer's instructions in a final volume of 100 μl for 60 min at RT. After blunting, the nuclei were washed twice with 1× CS/TX100 before proceeding with in situ ligation of BLISS adapters (see below for adapter preparation). Ligation was performed with 25 Weiss units of T4 DNA ligase (5 U/μl; Thermo Fisher Scientific, #EL0011) for 20 to 24 hours at 16°C in reaction volumes of 100 μl supplemented with BSA (Thermo Fisher Scientific, #AM2616) and ATP (Thermo Fisher Scientific, #R0441). Per preparation of $10^6$ cells, 4 μl of the selected BLISS adapter (10 μM) was ligated. Before use, BLISS double-stranded DNA (dsDNA) adapters were prepared from two complementary high-performance liquid chromatography–purified oligonucleotides ordered from Integrated DNA Technologies (IDT). Each dsDNA adapter contains a T7 promoter sequence for in vitro transcription (IVT), the RA5 Illumina RNA adapter sequence for downstream sequencing, an 8-nt unique molecular identifier (UMI) sequence generated by random incorporation of the four deoxynucleotide triphosphates (dNTPs) according to IDT's "Machine mixing" strategy, and an 8-nt sample barcode to enable multiplexing of BLISS libraries. Sense oligos diluted to 10 μM in nuclease-free water were phosphorylated with T4 PNK (NEB, #M0201) supplemented with ATP, after which an equimolar amount of antisense oligo was added. Oligos were annealed in a Thermocycler (5 min at 95°C, then ramping down to 25°C in steps of 1.5°C per min) to generate a 10 μM phosphorylated dsDNA adapter. After overnight ligation, nuclei were washed twice with 1× CS/TX100. To reverse cross-links and extract gDNA, nuclei were resuspended in 100 μl of DNA extraction buffer [10 mM tris-HCl, 100 mM NaCl, 50 mM EDTA, and 1% SDS (pH7.5)], supplemented with 10 μl of proteinase K (800 U/ml; NEB, #P8107), and incubated at 55°C for 14 to 18 hours while shaking at 800 rpm. Afterward, proteinase K was heat-inactivated for 10 min at 95°C, followed by extraction using phenol:chloroform:isoamyl alcohol 25:24:1 with 10 mM tris (pH 8.0), 1 mM EDTA (Sigma-Aldrich/Merck, #P2069), and chloroform (Merck, #1024451000), followed by ethanol precipitation. The purified gDNA was resuspended in 100 μl of TE and sonicated using BioRuptor Plus (Diagenode) with the following settings: 30 s ON, 60 s OFF, HIGH intensity, 30 cycles. Sonicated DNA was concentrated with Agencourt AMPure XP beads (Beckman Coulter), and fragment sizes were assessed using BioAnalyzer 2100 (Agilent Technologies) to range from 300 to 800 base pairs (bp), with a peak around 400 to 600 bp. To selectively and linearly amplify BLISS adapter-tagged genomic DSB ends, 100 ng of sonicated template was used for T7-mediated IVT using the MEGAscript T7 Transcription Kit [Thermo Fisher Scientific, #AMB13345, supplemented with Ribosafe RNase Inhibitor (Bioline, #BIO-65028)], according to the manufacturer's guidelines. Directly after RA3 ligation, reverse transcription was performed with Reverse Transcription Primer (RTP) (Illumina sequence, ordered via IDT) and SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific, #18090050). The manufacturer's protocol was followed extending the incubation time to 50 min at 50°C followed by 10-min heat inactivation at 80°C. Library amplification was carried out with NEBNext Ultra II Q5 Master Mix (NEB, #M0544), RP1 common primer, and a selected RPIX index primer (Illumina sequences, ordered through IDT). Libraries were amplified for eight PCR cycles, purified with a 0.8× AMPure XP bead purification, and then amplified for four additional PCR cycles. Then, the amplified libraries were cleaned up according to the two-sided AMPure XP bead purification protocol, aiming at retaining library sizes from ~300 to 850 bp. Final library profiles were assessed and quantified on a BioAnalyzer High-Sensitivity DNA chip and using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, #Q32851). Sequencing was performed at the Science for Life Laboratory, Sweden, on NextSeq 500 with NextSeq 500/550 High Output Kit v2 chemistry for SE 1 × 75 sequencing with an additional six cycles for index sequencing. Multiple indexed BLISS libraries were pooled together, aiming to retrieve at least 50 million reads per condition/library. Upon completion of the run, raw sequencing reads were demultiplexed on the basis of index sequences by Illumina's BaseSpace, after which the generated FASTQ files were downloaded. Two biological replicates were used in the analysis. The BLESS validation experiments were performed according to Crosetto *et al.* (*92*). The procedure resembles the sBLISS protocol and includes the in situ blunting of DSB ends, after mild fixation of the cells, and ligation to specialized biotinylated BLESS adapters, bearing the RA5 Illumina RNA sequence, that allow the selective affinity capture of processed DSBs. Upon ligation of the biotinylated adapter on DSBs, gDNA is purified and sonicated. Then, streptavidin beads (Dynabeads MyOne C1, #65001) are used to isolate DSB-bearing DNA fragments, followed by blunting of the other end and ligation to a second BLESS adapter containing the RA3 Illumina

Downloaded from https://www.science.org on February 13, 2025

RNA adapter sequence. PCR amplification was performed according to Illumina's guidelines, for 10 cycles using the RA5 and RA3 adapters, followed by purification and specific target qPCR amplification. The adapter sequences were previously reported for BLISS (*46*) and BLESS (*92*). For the RA3, RA5 adapters, RTP primer, and RP1 and RPIX primers, see the sequence information available for the Illumina small RNA library preparation kit.

### Data and statistical analysis

Statistically significant data were extracted by means of the IBM SPSS Statistics 19 (IBM) and R statistical package (www.r-project.org). Significant overrepresentation of pathways and gene networks was determined by GO (http://geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (https://genome.jp/kegg/pathway.html). Data analysis was performed with the PANTHER Classification System using the overrepresentation test. The *P* values were determined by the Fisher's exact test, and fold enrichment refers to the observed over the expected number within the reference list (*M. musculus*) that maps to the annotation data category. For MS, the MS/MS raw data were loaded in Proteome Discoverer 1.3.0.339 (Thermo Fisher Scientific) and run using the Mascot 2.3.02 (Matrix Science) search algorithm against the *M. musculus* theoretical proteome (last modified 6 July 2015) containing 46,470 entries in UniProt. A list of common contaminants was included in the database. For protein identification, the following search parameters were used: precursor error tolerance of 10 parts per million, fragment ion tolerance of 0.8 Da, trypsin full specificity, maximum number of missed cleavages of 3, and cysteine alkylation as a fixed modification. The resulting .dat and .msf files were subsequently loaded and merged in Scaffold (version 3.04.05, Proteome Software) for further processing and validation of the assigned MS/MS spectra. Thresholds for protein and peptide identification were set to 99 and 95% accordingly for proteins with minimum 1 different peptides identified, resulting in a protein FDR of <0.1%. For single peptide identifications, we applied the same criteria in addition to manual validation of MS/MS spectra. Protein lists were constructed from the respective peptide lists through extensive manual curation based on previous knowledge. For label-free relative quantitation of proteins, we applied a label-free relative quantitation method between the different samples (control versus bait) to determine unspecific binders during the affinity purification. All .dat and .msf files created by Proteome Discoverer were merged in Scaffold, where label-free relative quantification was performed using the total ion current (TIC) from each identified MS/MS spectrum. The TIC is the sum of the areas under all the peaks contained in an MS/MS spectrum, and total TIC value results by summing the intensity of the peaks contained in the peak list associated to an MS/MS sample. Protein lists containing the Scaffold-calculated total TIC quantitative value for each protein were exported to Microsoft Excel for further manual processing including categorization and additional curation based on previous knowledge. The FC of protein levels was calculated by dividing the mean total TIC quantitative value in bait samples with the mean value of the control samples for each of the proteins. Proteins having ≥60% protein coverage, ≥1 peptide in each sample, and an FC of ≥1.2 in all three measurements were selected as being significantly enriched in bXPF compared with BirA MEF samples. Proteins that were significantly enriched in bait samples were considered with a *P* value of ≤0.05 and an FC of ≥2. Significant overrepresentation of pathways, protein-protein interactions, and protein complexes were derived by STRING68 (https://string-db.org/cgi/input.pl). The quality of ChIP-Seq raw reads was checked using FastQC software (https://bioinformatics.babraham.ac.uk/projects/fastqc/). For both transcription factors (https://encodeproject.org/chip-seq/transcription_factor/) and histones (https://encodeproject.org/chip-seq/histone/), the appropriate pipelines proposed by ENCODE were adopted. All analyses were performed using as a reference, the mm10 mouse genome from University of California Santa Cruz (UCSC) using Kundaje's laboratory ChIP-Seq pipeline, and selecting the conservative set of peaks at the end. Peak annotation was performed using the HOMER Analysis package (*93*). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Peak visualization around TSS was performed using ChIPSeeker R package (*94*). ChIPseeker was also used for genomic peak annotation (annotatePeak function, annotation package TxDb.Mmusculus.UCSC.mm10.knownGene). For sBLISS, the generated amplified RNA is sequenced using next-generation sequencing, after which the obtained reads are mapped to the reference genome to identify the genomic locations of the DSBs. As described previously (*46*), a custom-built pipeline was used to keep only those reads that contain the expected prefix of 8-nt UMI and 8-nt sample barcode, using SAM tools and scan for matches, allowing at most one mismatch in the barcode sequence. The prefixes were then clipped off and stored, and the trimmed reads per condition were aligned to the GRCm38/mm10 reference genome with BWA-MEM. Only those reads with mapping quality scores of ≥30 were retained. Next, PCR duplicates were identified and removed, by searching for proximal reads (at most 30 bp apart in the reference genome) with at most two mismatches in the UMI sequence. Last, we generated BED files for downstream analyses, comprising a list of DSB end locations and a number of unique UMIs identified at these locations, which we refer to as "UMI-DSB ends" or unique DSB ends. DSBs from all samples and all replicates have been annotated using HOMER software, and a generic genome distribution (intergenic, 3′UTR, microRNA, noncoding RNA, TTS, pseudo, exon, intron, promoter, 5′UTR, small nucleolar RNA, and ribosomal RNA) was created. To analyze the cumulative distribution of DNA DSBs ± 2 kb around the TSS, we used ComputeMatrix (deepTools suite) to calculate the scores per genome region, i.e., 2 kb around TSS of cumulative DSB reads [normalized using reads per kilobase per million mapped reads (RPKM)] and plotProfile (deepTools suite) for data representation (*95*). The BLISS-ChIP-Seq as well as the DRIP-Seq, CTCF ChIP-Seq, and TOP2B ChIP-Seq comparisons were performed using bedtools. Drip-Seq data genome coordinates were converted from mm9 to mm10 with the liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver).The significance of difference between correlations was tested by using the tool (https://psychometrica.de/correlation.html) as previously described (*96*). Last, circular visualization was performed with circlize (version 0.4.15) R package, and intersections were visualized with the use of UpSetR (version 1.4.0) R package.

Error bars in the figures indicate SEM among $n > 3$ biological replicates. Asterisk indicates the significance set at *P* value: $^{*}P \leq 0.05$, $^{**}P \leq 0.01$, and $^{***}P \leq 0.001$ [two-way analysis of variance (ANOVA) with post hoc testing].

Downloaded from https://www.science.org on February 13, 2025

### Multivariate classification analysis

We performed a multivariate classification analysis with a binary outcome of bXPF binding to DNA (bound/unbound). As possible predictors, we used the log values of RNA-Seq and the BLISS measurements. The tRA treatment (yes/no) was also included as a potential predictor. The analysis determines whether the predictors correlate with the bXPF binding status in a multivariate way and included feature selection by filtering out features that are either irrelevant or redundant in predicting the outcome. Because the same biological sample was measured twice, e.g., one treated with tRA and one without, these measurements are not independently and identically distributed (repeated measurements). To perform the analysis, we used the "Just Add Data Bio (JAD Bio)" tool (www.jadbio.com). JAD Bio provides conservative estimates of predictive performance and corresponding confidence intervals and included the following user preferences: enforcing feature selection, not-enforcing interpretable models, using sample ID to indicate the repeated measurements, and the extensive analysis setting, the most exhaustive in terms of models it tries. The winning model did not contain the tRA treatment in the predictors as it was thrown by the feature selection step. Out of all models tested, the winning model was a Support Vector Machine model, using the full polynomial kernel of degree 2. This is equivalent to a linear model with an intercept term and predictors logRNA-Seq, logRNA-Seq$^2$, logBLISS, logBLISS$^2$, and the interaction term logRNA-Seq × logBLISS. The predictive performance of the model, adjusted for trying several algorithms, is 0.726 as measured by the area under the receiver operating characteristic curve (AUC), with confidence interval (0.671 to 0.781). The internal workings of JAD Bio and the methods it uses were previously described (*97*).

## Supplementary Materials

**This PDF file includes:**

Figs. S1 to S8

## REFERENCES AND NOTES


1. U. Ohler, D. A. Wassarman, Promoting developmental transcription. *Development* **137**, 15–26 (2010).
2. H. Gaillard, A. Aguilera, Transcription as a threat to genome integrity. *Annu. Rev. Biochem.* **85**, 291–317 (2016).
3. V. H. Oestergaard, M. Lisby, Transcription-replication conflicts at chromosomal fragile sites—Consequences in M phase and beyond. *Chromosoma* **126**, 213–222 (2017).
4. J. H. Hoeijmakers, Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
5. M. Manandhar, K. S. Boulware, R. D. Wood, The ERCC1 and ERCC4 (XPF) genes and gene products. *Gene* **569**, 153–161 (2015).
6. S. Q. Gregg, A. R. Robinson, L. J. Niedernhofer, Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. *DNA Repair* **10**, 781–791 (2011).
7. A. M. Sijbers, W. L. de Laat, R. R. Ariza, M. Biggerstaff, Y. F. Wei, J. G. Moggs, K. C. Carter, B. K. Shell, E. Evans, M. C. de Jong, S. Rademakers, J. de Rooij, N. G. J. Jaspers, J. H. J. Hoeijmakers, R. D. Wood, Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* **86**, 811–822 (1996).
8. P. C. Hanawalt, Subpathways of nucleotide excision repair and their regulation. *Oncogene* **21**, 8949–8956 (2002).
9. J. A. Marteijn, H. Lans, W. Vermeulen, J. H. Hoeijmakers, Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **15**, 465–481 (2014).
10. M. van Duin, J. de Wit, H. Odijk, A. Westerveld, A. Yasui, M. H. Koken, J. H. Hoeijmakers, D. Bootsma, Molecular characterization of the human excision repair gene ERCC-1: cDNA cloning and amino acid homology with the yeast DNA repair gene RAD10. *Cell* **44**, 913–923 (1986).
11. D. Klein Douwel, R. A. C. M. Boonen, D. T. Long, A. A. Szypowska, M. Räschle, J. C. Walter, P. Knipscheer, XPF-ERCC1 acts in unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4. *Mol. Cell* **54**, 460–471 (2014).
12. M. R. G. Hodskinson, J. Silhan, G. P. Crossan, J. I. Garaycoechea, S. Mukherjee, C. M. Johnson, O. D. Schärer, K. J. Patel, Mouse SLX4 is a tumor suppressor that stimulates the activity of the nuclease XPF-ERCC1 in DNA crosslink repair. *Mol. Cell* **54**, 472–484 (2014).
13. G. M. Adair, R. L. Rolig, D. Moore-Faver, M. Zabelshansky, J. H. Wilson, R. S. Nairn, Role of *ERCC1* in removal of long non-homologous tails during targeted homologous recombination. *EMBO J.* **19**, 5552–5561 (2000).
14. A. Ahmad, A. R. Robinson, A. Duensing, E. van Drunen, H. B. Beverloo, D. B. Weisberg, P. Hasty, J. H. J. Hoeijmakers, L. J. Niedernhofer, ERCC1-XPF endonuclease facilitates DNA double-strand break repair. *Mol. Cell. Biol.* **28**, 5082–5092 (2008).
15. R. G. Sargent, R. L. Rolig, A. E. Kilburn, G. M. Adair, J. H. Wilson, R. S. Nairn, Recombination-dependent deletion formation in mammalian cells deficient in the nucleotide excision repair gene ERCC1. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13122–13127 (1997).
16. A. Z. Al-Minawi, N. Saleh-Gohari, T. Helleday, The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. *Nucleic Acids Res.* **36**, 1–9 (2008).
17. J. L. Ma, E. M. Kim, J. E. Haber, S. E. Lee, Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol. Cell. Biol.* **23**, 8820–8828 (2003).
18. P. Munoz, R. Blanco, J. M. Flores, M. A. Blasco, XPF nuclease-dependent telomere loss and increased DNA damage in mice overexpressing TRF2 result in premature aging and cancer. *Nat. Genet.* **37**, 1063–1071 (2005).
19. X. D. Zhu, L. Niedernhofer, B. Kuster, M. Mann, J. H. J. Hoeijmakers, T. de Lange, ERCC1/XPF removes the 3′ overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. *Mol. Cell* **12**, 1489–1498 (2003).
20. J. Woodrick, S. Gupta, S. Camacho, S. Parvathaneni, S. Choudhury, A. Cheema, Y. Bai, P. Khatkar, H. V. Erkizan, F. Sami, Y. Su, O. D. Schärer, S. Sharma, R. Roy, A new sub-pathway of long-patch base excision repair involving 5′ gap formation. *EMBO J.* **36**, 1605–1622 (2017).
21. D. Bootsma, K. H. Kraemer, J. E. Cleaver, J. H. J. Hoeijmakers, in *The Genetic Basis of Human Cancer*, B. Vogelstein, K. W. Kinzler, Eds. (McGraw-Hill, 1998), pp. 245–274.
22. D. K. Bootsma, K. H. Cleaver, J. E. Hoeijmakers, JHJ., in*The Metabolic and Molecular Basis of Inherited Disease*, C. R. Scriver, Ed. (McGraw-Hill, 2001), pp. 677–703.
23. T. Mori, M. J. Yousefzadeh, M. Faridounnia, J. X. Chong, F. M. Hisama, L. Hudgins, G. Mercado, E. A. Wade, A. S. Barghouthy, L. Lee, G. M. Martin, D. A. Nickerson, M. J. Bamshad; University of Washington Center for Mendelian Genomics, L. J. Niedernhofer, J. Oshima, *ERCC4* variants identified in a cohort of patients with segmental progeroid syndromes. *Hum. Mutat.* **39**, 255–265 (2018).
24. K. Kashiyama, Y. Nakazawa, D. T. Pilz, C. Guo, M. Shimada, K. Sasaki, H. Fawcett, J. F. Wing, S. O. Lewin, L. Carr, T. S. Li, K. I. Yoshiura, A. Utani, A. Hirano, S. Yamashita, D. Greenblatt, T. Nardo, M. Stefanini, D. McGibbon, R. Sarkany, H. Fassihi, Y. Takahashi, Y. Nagayama, N. Mitsutake, A. R. Lehmann, T. Ogi, Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, Xeroderma pigmentosum, and Fanconi anemia. *Am. J. Hum. Genet.* **92**, 807–819 (2013).
25. M. Bogliolo, B. Schuster, C. Stoepker, B. Derkunt, Y. Su, A. Raams, J. P. Trujillo, J. Minguillón, M. J. Ramírez, R. Pujol, J. A. Casado, R. Baños, P. Rio, K. Knies, S. Zúñiga, J. Benítez, J. A. Bueren, N. G. J. Jaspers, O. D. Schärer, J. P. de Winter, D. Schindler, J. Surrallés, Mutations in *ERCC4*, encoding the DNA-repair endonuclease XPF, cause Fanconi anemia. *Am. J. Hum. Genet.* **92**, 800–806 (2013).
26. N. G. J. Jaspers, A. Raams, M. C. Silengo, N. Wijgers, L. J. Niedernhofer, A. R. Robinson, G. Giglia-Mari, D. Hoogstraten, W. J. Kleijer, J. H. J. Hoeijmakers, W. Vermeulen, First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure. *Am. J. Hum. Genet.* **80**, 457–466 (2007).
27. M. Tian, R. Shinkura, N. Shinkura, F. W. Alt, Growth retardation, early death, and DNA repair defects in mice deficient for the nucleotide excision repair enzyme XPF. *Mol. Cell. Biol.* **24**, 1200–1205 (2004).
28. L. J. Niedernhofer, G. A. Garinis, A. Raams, A. S. Lalai, A. R. Robinson, E. Appeldoorn, H. Odijk, R. Oostendorp, A. Ahmad, W. van Leeuwen, A. F. Theil, W. Vermeulen, G. T. J. van der Horst, P. Meinecke, W. J. Kleijer, J. Vijg, N. G. J. Jaspers, J. H. J. Hoeijmakers, A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature* **444**, 1038–1043 (2006).
29. N. L. May, D. Mota-Fernandes, R. Vélez-Cruz, I. Iltis, D. Biard, J. M. Egly, 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 (2010).
30. N. Le May, J. M. Egly, F. Coin, True lies: The double life of the nucleotide excision repair factors in transcription and DNA repair. *J. Nucleic Acids* **2010**, 616342 (2010).


Downloaded from https://www.science.org on February 13, 2025

31. N. Le May, D. Fradin, I. Iltis, P. Bougneres, J. M. Egly, XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol. Cell* **47**, 622–632 (2012).
32. G. Chatzinikolaou, Z. Apostolou, T. Aid-Pavlidis, A. Ioannidou, I. Karakasilioti, G. L. Papadopoulos, M. Aivaliotis, M. Tsekrekou, J. Strouboulis, T. Kosteas, G. A. Garinis, ERCC1-XPF cooperates with CTCF and cohesin to facilitate the developmental silencing of imprinted genes. *Nat. Cell Biol.* **19**, 421–432 (2017).
33. I. Kamileri, I. Karakasilioti, A. Sideri, T. Kosteas, A. Tatarakis, I. Talianidis, G. A. Garinis, Defective transcription initiation causes postnatal growth failure in a mouse model of nucleotide excision repair (NER) progeria. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2995–3000 (2012).
34. J. Bastien, C. Rochette-Egly, Nuclear retinoid receptors and the transcription of retinoid-target genes. *Gene* **328**, 1–16 (2004).
35. X. Huang, X. Gao, W. Li, S. Jiang, R. Li, H. Hong, C. Zhao, P. Zhou, H. Chen, X. Bo, H. Li, Stable H3K4me3 is associated with transcription initiation during early embryo development. *Bioinformatics* **35**, 3931–3936 (2019).
36. F. Tie, R. Banerjee, C. A. Stratton, J. Prasad-Sinha, V. Stepanik, A. Zlobin, M. O. Diaz, P. C. Scacheri, P. J. Harte, CBP-mediated acetylation of histone H3 lysine 27 antagonizes *Drosophila* Polycomb silencing. *Development* **136**, 3131–3141 (2009).
37. J. Cheng, R. Blum, C. Bowman, D. Hu, A. Shilatifard, S. Shen, B. D. Dynlacht, A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. *Mol. Cell* **53**, 979–992 (2014).
38. B. G. Hoffman, G. Robertson, B. Zavaglia, M. Beach, R. Cullum, S. Lee, G. Soukhatcheva, L. Li, E. D. Wederell, N. Thiessen, M. Bilenky, T. Cezard, A. Tam, B. Kamoh, I. Birol, D. Dai, Y. J. Zhao, M. Hirst, C. B. Verchere, C. D. Helgason, M. A. Marra, S. J. M. Jones, P. A. Hoodless, Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and liver. *Genome Res.* **20**, 1037–1051 (2010).
39. X. Zheng, Y. Kim, Y. Zheng, Identification of lamin B-regulated chromatin regions based on chromatin landscapes. *Mol. Biol. Cell* **26**, 2685–2697 (2015).
40. C. Jiang, B. F. Pugh, A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. *Genome Biol.* **10**, R109 (2009).
41. B. Schwer, P. C. Wei, A. N. Chang, J. Kao, Z. Du, R. M. Meyers, F. W. Alt, Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2258–2263 (2016).
42. R. Madabhushi, F. Gao, A. R. Pfenning, L. Pan, S. Yamakawa, J. Seo, R. Rueda, T. X. Phan, H. Yamakawa, P. C. Pao, R. T. Stott, E. Gjoneska, A. Nott, S. Cho, M. Kellis, L. H. Tsai, Activity-induced DNA breaks govern the expression of neuronal early-response genes. *Cell* **161**, 1592–1605 (2015).
43. A. Marechal, L. Zou, DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb. Perspect. Biol.* **5**, a012716 (2013).
44. F. Chen, X. Gao, A. Shilatifard, Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide. *Genes Dev.* **29**, 39–47 (2015).
45. K. Yankulov, K. Yamashita, R. Roy, J. M. Egly, D. L. Bentley, The transcriptional elongation inhibitor 5,6-dichloro-L-β-D-ribofuranosylbenzimidazole inhibits transcription factor IIH-associated protein kinase. *J. Biol. Chem.* **270**, 23922–23925 (1995).
46. W. X. Yan, R. Mirzazadeh, S. Garnerone, D. Scott, M. W. Schneider, T. Kallas, J. Custodio, E. Wernersson, Y. Li, L. Gao, Y. Federova, B. Zetsche, F. Zhang, M. Bienko, N. Crosetto, BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, 15058 (2017).
47. K. Lakiotaki, G. Georgakopoulos, E. Castanas, O. D. Røe, G. Borboudakis, I. Tsamardinos, A data driven approach reveals disease similarity on a molecular level. *NPJ Syst. Biol. Appl.* **5**, 39 (2019).
48. S. J. McKie, K. C. Neuman, A. Maxwell, DNA topoisomerases: Advances in understanding of cellular roles and multi-protein complexes via structure-function analysis. *Bioessays* **43**, e2000286 (2021).
49. Y. Pommier, Y. Sun, S. N. Huang, J. L. Nitiss, Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat. Rev. Mol. Cell Biol.* **17**, 703–721 (2016).
50. A. Canela, Y. Maman, S. Y. N. Huang, G. Wutz, W. Tang, G. Zagnoli-Vieira, E. Callen, N. Wong, A. Day, J. M. Peters, K. W. Caldecott, Y. Pommier, A. Nussenzweig, Topoisomerase II-induced chromosome breakage and translocation is determined by chromosome architecture and transcriptional activity. *Mol. Cell* **75**, 252–266.e8 (2019).
51. S. K. Calderwood, A critical role for topoisomerase IIb and DNA double strand breaks in transcription. *Transcription* **7**, 75–83 (2016).
52. L. Uusküla-Reimand, H. Hou, P. Samavarchi-Tehrani, M. V. Rudan, M. Liang, A. Medina-Rivera, H. Mohammed, D. Schmidt, P. Schwalie, E. J. Young, J. Reimand, S. Hadjur, A.-C. Gingras, M. D. Wilson, Topoisomerase II β interacts with cohesin and CTCF at topological domain borders. *Genome Biol.* **17**, 182 (2016).
53. K. Skourti-Stathaki, N. J. Proudfoot, A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* **28**, 1384–1396 (2014).
54. K. Skourti-Stathaki, N. J. Proudfoot, N. Gromak, Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805 (2011).
55. M. Muers, Mutation: The perils of transcription. *Nat. Rev. Genet.* **12**, 156 (2011).
56. H. Wimberly, C. Shee, P. C. Thornton, P. Sivaramakrishnan, S. M. Rosenberg, P. J. Hastings, R-loops and nicks initiate DNA breakage and genome instability in non-growing *Escherichia coli*. *Nat. Commun.* **4**, 2115 (2013).
57. J. Brustel, Z. Kozik, N. Gromak, V. Savic, S. M. M. Sweet, Large XPF-dependent deletions following misrepair of a DNA double strand break are prevented by the RNA:DNA helicase Senataxin. *Sci. Rep.* **8**, 3850 (2018).
58. J. Sollier, C. T. Stork, M. L. García-Rubio, R. D. Paulsen, A. Aguilera, K. A. Cimprich, Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol. Cell* **56**, 777–785 (2014).
59. O. Chatzidoukaki, K. Stratigi, E. Goulielmaki, G. Niotis, A. Akalestou-Clocher, K. Gkirtzimanaki, A. Zafeiropoulos, J. Altmüller, P. Topalis, G. A. Garinis, R-loops trigger the release of cytoplasmic ssDNAs leading to chronic inflammation upon DNA damage. *Sci. Adv.* **7**, eabj5769 (2021).
60. E. Goulielmaki, M. Tsekrekou, N. Batsiotos, M. Ascensão-Ferreira, E. Ledaki, K. Stratigi, G. Chatzinikolaou, P. Topalis, T. Kosteas, J. Altmüller, J. A. Demmers, N. L. Barbosa-Morais, G. A. Garinis, The splicing factor XAB2 interacts with ERCC1-XPF and XPG for R-loop processing. *Nat. Commun.* **12**, 3153 (2021).
61. Y. L. Lin, P. Pasero, Caught in the Act: R-loops are cleaved by structure-specific endonucleases to generate DSBs. *Mol. Cell* **56**, 721–722 (2014).
62. C. Rinaldi, P. Pizzul, M. P. Longhese, D. Bonetti, Sensing R-loop-associated DNA damage to safeguard genome stability. *Front. Cell Dev. Biol.* **8**, 618157 (2021).
63. L. Muniz, E. Nicolas, D. Trouche, RNA polymerase II speed: A key player in controlling and adapting transcriptome composition. *EMBO J.* **40**, e105740 (2021).
64. V. K. Tiwari, L. Burger, V. Nikoletopoulou, R. Deogracias, S. Thakurela, C. Wirbelauer, J. Kaut, R. Terranova, L. Hoerner, C. Mielke, F. Boege, R. Murr, A. H. Peters, Y. A. Barde, D. Schübeler, Target genes of topoisomerase IIβ regulate neuronal survival and are defined by their chromatin state. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E934–E943 (2012).
65. M. P. Crossley, M. Bocek, K. A. Cimprich, R-loops as cellular regulators and genomic threats. *Mol. Cell* **73**, 398–411 (2019).
66. G. A. Garinis, L. M. Uittenboogaard, H. Stachelscheid, M. Fousteri, W. van Ijcken, T. M. Breit, H. van Steeg, L. H. F. Mullenders, G. T. J. van der Horst, J. C. Brüning, C. M. Niessen, J. H. J. Hoeijmakers, B. Schumacher, Persistent transcription-blocking DNA lesions trigger somatic growth attenuation associated with longevity. *Nat. Cell Biol.* **11**, 604–615 (2009).
67. J. M. Fortune, N. Osheroff, Merbarone inhibits the catalytic activity of human topoisomerase IIalpha by blocking DNA cleavage. *J. Biol. Chem.* **273**, 17643–17650 (1998).
68. P. Majumder, J. T. Lee, B. G. Barwick, D. G. Patterson, A. P. R. Bally, C. D. Scharer, J. M. Boss, The murine MHC class II super enhancer IA/IE-SE contains a functionally redundant CTCF-binding component and a novel element critical for maximal expression. *J. Immunol.* **206**, 2221–2232 (2021).
69. H. Min, K. A. Kong, J. Y. Lee, C. P. Hong, S. H. Seo, T. Y. Roh, S. S. Bae, M. H. Kim, CTCF-mediated chromatin loop for the posterior Hoxc gene expression in MEF cells. *IUBMB Life* **68**, 436–444 (2016).
70. H. Luo, G. Zhu, M. A. Eshelman, T. K. Fung, Q. Lai, F. Wang, B. B. Zeisig, J. Lesperance, X. Ma, S. Chen, N. Cesari, C. Cogle, B. Chen, B. Xu, F. C. Yang, C. W. E. So, Y. Qiu, M. Xu, S. Huang, HOTTIP-dependent R-loop formation regulates CTCF boundary activity and TAD integrity in leukemia. *Mol. Cell* **82**, 833–851.e11 (2022).
71. B. G. Ju, V. V. Lunyak, V. Perissi, I. Garcia-Bassets, D. W. Rose, C. K. Glass, M. G. Rosenfeld, A topoisomerase IIβ-mediated dsDNA break required for regulated transcription. *Science* **312**, 1798–1802 (2006).
72. S. Morimoto, M. Tsuda, H. Bunch, H. Sasanuma, C. Austin, S. Takeda, Type II DNA topoisomerases cause spontaneous double-strand breaks in genomic DNA. *Genes (Basel)* **10**, 868 (2019).
73. F. Cortes Ledesma, S. F. El Khamisy, M. C. Zuma, K. Osborn, K. W. Caldecott, A human 5′-tyrosyl DNA phosphodiesterase that repairs topoisomerase-mediated DNA damage. *Nature* **461**, 674–678 (2009).
74. N. N. Hoa, T. Shimizu, Z. W. Zhou, Z. Q. Wang, R. A. Deshpande, T. T. Paull, S. Akter, M. Tsuda, R. Furuta, K. Tsutsui, S. Takeda, H. Sasanuma, Mre11 is essential for the removal of lethal topoisomerase 2 covalent cleavage complexes. *Mol. Cell* **64**, 580–592 (2016).
75. F. Aymard, B. Bugler, C. K. Schmidt, E. Guillou, P. Caron, S. Briois, J. S. Iacovoni, V. Daburon, K. M. Miller, S. P. Jackson, G. Legube, Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat. Struct. Mol. Biol.* **21**, 366–374 (2014).
76. L. D. McDaniel, R. A. Schultz, XPF/ERCC4 and ERCC1: Their products and biological roles. *Adv. Exp. Med. Biol.* **637**, 65–82 (2008).


Downloaded from https://www.science.org on February 13, 2025

77. R. Biehs, M. Steinlage, O. Barton, S. Juhász, J. Künzel, J. Spies, A. Shibata, P. A. Jeggo, M. Löbrich, DNA double-strand break resection occurs during non-homologous end joining in G1 but is distinct from resection during homologous recombination. *Mol. Cell* **65**, 671–684.e5 (2017).
78. M. Pan, W. C. Wright, R. H. Chapple, A. Zubair, M. Sandhu, J. E. Batchelder, B. C. Huddle, J. Low, K. B. Blankenship, Y. Wang, B. Gordon, P. Archer, S. W. Brady, S. Natarajan, M. J. Posgai, J. Schuetz, D. Miller, R. Kalathur, S. Chen, J. P. Connelly, M. M. Babu, M. A. Dyer, S. M. Pruett-Miller, B. B. Freeman III, T. Chen, L. A. Godley, S. C. Blanchard, E. Stewart, J. Easton, P. Geeleher, The chemotherapeutic CX-5461 primarily targets TOP2B and exhibits selective activity in high-risk neuroblastoma. *Nat. Commun.* **12**, 6468 (2021).
79. L. Uuskula-Reimand, M. D. Wilson, Untangling the roles of TOP2A and TOP2B in transcription and cancer. *Sci. Adv.* **8**, eadd4920 (2022).
80. K. Szlachta, A. Manukyan, H. M. Raimer, S. Singh, A. Salamon, W. Guo, K. S. Lobachev, Y. H. Wang, Topoisomerase II contributes to DNA secondary structure-mediated double-stranded breaks. *Nucleic Acids Res.* **48**, 6654–6671 (2020).
81. A. De Magis, S. G. Manzo, M. Russo, J. Marinello, R. Morigi, O. Sordet, G. Capranico, DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 816–825 (2019).
82. J. Tan, X. Wang, L. Phoon, H. Yang, L. Lan, Resolution of ROS-induced G-quadruplexes and R-loops at transcriptionally active sites is dependent on BLM helicase. *FEBS Lett.* **594**, 1359–1367 (2020).
83. C. Y. Lee, C. McNerney, K. Ma, W. Zhao, A. Wang, S. Myong, R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation. *Nat. Commun.* **11**, 3392 (2020).
84. P. Kotsantis, S. Segura-Bayona, P. Margalef, P. Marzec, P. Ruis, G. Hewitt, R. Bellelli, H. Patel, R. Goldstone, A. R. Poetsch, S. J. Boulton, RTEL1 regulates G4/R-loops to avert replication-transcription collisions. *Cell Rep.* **33**, 108546 (2020).
85. G. Ren, W. Jin, K. Cui, J. Rodrigez, G. Hu, Z. Zhang, D. R. Larson, K. Zhao, CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. *Mol. Cell* **67**, 1049–1058.e6 (2017).
86. E. Goulielmaki, A. Ioannidou, M. Tsekrekou, K. Stratigi, I. K. Poutakidou, K. Gkirtzimanaki, M. Aivaliotis, K. Evangelou, P. Topalis, J. Altmüller, V. G. Gorgoulis, G. Chatzinikolaou, G. A. Garinis, Tissue-infiltrating macrophages mediate an exosome-based metabolic reprogramming upon DNA damage. *Nat. Commun.* **11**, 42 (2020).
87. F. Chedin, S. R. Hartono, L. A. Sanz, V. Vanoosthuyse, Best practices for the visualization, mapping, and manipulation of R-loops. *EMBO J.* **40**, e106394 (2021).
88. A. Cristini, M. Groh, M. S. Kristiansen, N. Gromak, RNA/DNA hybrid interactome IDEntifies DXH9 as a molecular player in transcriptional termination and R-loop-associated DNA damage. *Cell Rep.* **23**, 1891–1905 (2018).
89. P. Schwertman, A. Lagarou, D. H. W. Dekkers, A. Raams, A. C. van der Hoek, C. Laffeber, J. H. J. Hoeijmakers, J. A. A. Demmers, M. Fousteri, W. Vermeulen, J. A. Marteijn, UV-sensitive syndrome protein UVSSA recruits USP7 to regulate transcription-coupled repair. *Nat. Genet.* **44**, 598–602 (2012).
90. M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis, M. Mann, Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466–469 (1996).
91. J. Rappsilber, U. Ryder, A. I. Lamond, M. Mann, Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231–1245 (2002).
92. N. Crosetto, A. Mitra, M. J. Silva, M. Bienko, N. Dojer, Q. Wang, E. Karaca, R. Chiarle, M. Skrzypczak, K. Ginalski, P. Pasero, M. Rowicka, I. Dikic, Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **10**, 361–365 (2013).
93. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
94. G. Yu, L. G. Wang, Q. Y. He, ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
95. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
96. M. Eid, M. Gollwitzer, M. Schmitt, *Hypothesis Tests for Comparing Correlations* (Psychometrica, 2014).
97. V. Lagani, G. Athineou, A. Farcomeni, M. Tsagris, I. Tsamardinos, Feature selection with the R package MXM: Discovering statistically equivalent feature subsets. *J. Stat. Softw.* **80**, 1–25 (2017).

**Acknowledgments**

**Funding:** This work was supported by the Horizon 2020 ERC Consolidator grant "DeFiNER" (GA 64663); the ERC PoC "Inflacare" (GA 874456); the Horizon 2020 Marie Curie ITN "aDDRess" (GA 812829), and "HealthAge" (GA 812830), ELIDEK grants 631, 196, and 1059; the "Research-Create-Innovate" actions (MIA-RTDI) "Panther"-00852 and "Liquid Pancreas"-00940; Uni-Pharma Kleon Tsetis Pharmaceutical laboratories S.A (PAR00838) and Pharmathen S.A. (PAR00863) funds; and Greece 2.0, National Recovery and Resilience Plan Flagship program TAEDR-0535850. A.A.-C. was supported by the ELIDEK Fellowship 6204. **Author contributions:** Conceptualization: G.C., K.S., E.G., C.A., N.C., and G.A.G. Methodology: G.C., K.S., A.S., E.G., and B. A.M.B. Data analysis: G.C., K.S., E.G., A.A.-C., I.T., P.T., B.A.M.B., N.C., and J.A. Writing: G.C., K.S., and G.A.G. **Competing interests:** I.T. is affiliated with Gnosis Data Analysis that owns JAD Bio. N.C. is a coinventor on an international patent (patent application serial no. PCT/US2015/067138 filed on 21 December 2015 and published as PCT publication no. WO2016/100974 on 23 June 2016) describing, among other things, applications of the BLISS method for CRISPR nuclease off-target detection, filed by The Broad Institute, Cambridge MA, USA. All other authors declare that they have no competing interests. **Data and materials availability:** The MS proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD017063. The ChIP-Seq data (E-MTAB-8154) and RNA-Seq data (E-MTAB-8156) are deposited in ArrayExpress (https://ebi.ac.uk/arrayexpress/). The BLISS data are deposited in SRA (BioProject PRJNA555448) (https://ncbi.nlm.nih.gov/bioproject/?term=PRJNA555448). Data for the comparative analysis were obtained from the following sources: ChIP-seq for H3K4me1 (GSM723005), CTCF (GSM2635593), H3K4me3 (GSM723006), H3K27ac (GSM851277), RNAPII (GSM723007), TOP2B (GSM2635608), and LaminB DamID (GSE17051). DRIP-seq data were acquired from the following study (GSE70189). All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 11 April 2023
Accepted 5 October 2023
Published 8 November 2023
10.1126/sciadv.adi2095

Downloaded from https://www.science.org on February 13, 2025