3292–3306 *Nucleic Acids Research, 2022, Vol. 50, No. 6* *Published online 2 March 2022*
https://doi.org/10.1093/nar/gkac143

8-oxodG accumulation within super-enhancers marks fragile CTCF-mediated chromatin loops

Giovanni Scala[1], Francesca Gorini[2], Susanna Ambrosio[1], Andrea M. Chiariello[3], Mario Nicodemi[3], Luigi Lania[2], Barbara Majello[1,*] and Stefano Amente[2,*]

[1]Department of Biology, University of Naples 'Federico II', Naples, Italy, [2]Department of Molecular Medicine and Medical Biotechnologies, University of Naples 'Federico II', Naples, Italy and [3]Department of Physics, University of Naples Federico II, and INFN, Naples, Italy

Received November 02, 2021; Revised February 12, 2022; Editorial Decision February 14, 2022; Accepted February 15, 2022

ABSTRACT

8-Oxo-7,8-dihydro-2′-deoxyguanosine (8-oxodG), a major product of the DNA oxidization process, has been proposed to have an epigenetic function in gene regulation and has been associated with genome instability. NGS-based methodologies are contributing to the characterization of the 8-oxodG function in the genome. However, the 8-oxodG epigenetic role at a genomic level and the mechanisms controlling the genomic 8-oxodG accumulation/maintenance have not yet been fully characterized. In this study, we report the identification and characterization of a set of enhancer regions accumulating 8-oxodG in human epithelial cells. We found that these oxidized enhancers are mainly super-enhancers and are associated with bidirectional-transcribed enhancer RNAs and DNA Damage Response activation. Moreover, using ChIA-PET and HiC data, we identified specific CTCF-mediated chromatin loops in which the oxidized enhancer and promoter regions physically associate. Oxidized enhancers and their associated chromatin loops accumulate endogenous double-strand breaks which are in turn repaired by NHEJ pathway through a transcription-dependent mechanism. Our work suggests that 8-oxodG accumulation in enhancers–promoters pairs occurs in a transcription-dependent manner and provides novel mechanistic insights on the intrinsic fragility of chromatin loops containing oxidized enhancers-promoters interactions.

INTRODUCTION

8-Oxo-7,8-dihydro-2′-deoxyguanosine (8-oxodG) is a major product of the DNA oxidization process (1–3). Besides the ability to form a canonical Watson–Crick base pair with 2′-deoxycytidine, 8-oxodG is also capable to form a Hoogsteen base pair with 2′deoxyadenosine (dA) during replication. Hence, 8-oxodG causes dC:dG to dA:dT premutagenic transversion (1,4). To cope with this phenomenon, human cells evolved a multi-layer defense system named base excision repair (BER) (1,5,6). The 8-oxodG DNA glycosylase (OGG1) constitutes in human cells the front line of the BER defense system for the 8-oxodG repair. OGG1 recognizes and excises the 8-oxodG when paired with dC, thus generating a 3′-α, β-unsaturated aldehyde terminus (3′-dRP) and 5′-phosphate (AP) site. The AP site is subsequently processed by the apurinic/apyrimidinic site nuclease (APE1) which cleaves the DNA phosphate backbone, creates an SSB, and forms a polymerase-ready 3′-OH residue. Thus, APE1 enables downstream BER reactions, which generally engage DNA polymerase beta and a DNA ligase (1,5,6).

NGS-based methodologies have recently been used to provide a high-resolution mapping of 8-oxodG in human and mouse genomes (7–18). These studies highlighted the non-stochastic distribution of 8-oxodG and its enrichment in specific genomic regions. Moreover, in association with the occurrence of single/double-strand breaks (SSBs and DSBs) and the activation of DNA damage response (DDR), 8-oxodG displays a pattern of accumulation at specific promoter regions that are linked to chromatin organization and DNA transcription and replication (14,15).

The genome-wide co-occurrence of 8-oxodG and OGG1 at promoters (15,19), together with their dynamic accumulation at specific regions in response to transcriptional stimuli (20–23), corroborates the epigenetic role of 8-oxodG and the role of the BER pathway as an epigenetic modulator (1,24,25). Interestingly, 8-oxodG leads to SSB formation which in turn determines chromatin loop formation and crosstalk between protein complexes involved in transcription regulation (22,23,26). However, unrepaired and/or incomplete-repaired 8-oxodG is a source of SSBs accumulation that in turn determines DSBs formation (1). It is plausible that 8-oxodG and its associated BER activi-

*To whom correspondence should be addressed. Tel: +39 081 7463044; Email: stamente@unina.it
Correspondence may also be addressed to Barbara Majello. Tel: +39 081 679062; Email: majello@unina.it

ties might serve as drivers for transcription and might cause genome instability under deregulated conditions or within intrinsic fragile genomic regions.

To better understand the 8-oxodG function in the genome, it is important to identify the genomic elements in which 8-oxodG accumulate and to investigate the mechanism(s) that might control its accumulation/maintenance in the human genome.

Here, we report the identification of a subset of enhancer regions that accumulate 8-oxodG (oxidized enhancers) at the steady-state condition in non-tumorigenic human epithelial MCF10A cells. We found that the oxidized enhancers are mainly super-enhancers with bidirectional-transcribed enhancer RNA (eRNA) associated with DDR activation. Using ChIA-PET and HiC data analysis, we found that, within specific CTCF-mediated chromatin loops, the oxidized enhancer and promoter regions physically associate. Finally, the oxidized enhancers and the associated CTCF-mediated chromatin loops accumulate endogenous SSBs and DSBs that are repaired by the NHEJ pathway through a transcription-dependent mechanism.

MATERIALS AND METHODS

Cell culture and treatments

MCF10A cells were cultured in 1:1 mixture DMEM-F12 supplemented with 5% horse serum, 10 μg/ml insulin, 0.5 μg/ml hydrocortisone, 20 ng/ml epidermal growth factor, 100 ng/ml cholera enterotoxin and incubated at 37°C in a humidified atmosphere with 5% CO_2 (15). For UV treatment, exponentially growing MCF10A cells were exposed to 40 J/m^2 UV light (254 nm). Medium was refreshed after irradiation and cells were incubated for 30 min at 37°C. Then, cells were washed twice with ice-cold phosphate buffered saline (PBS) and collected (14,15). For NAC treatment, 1 mM *N*-acetyl cysteine (A7250, Sigma) was added to the medium for 1 h and then MCF10A cells were washed twice with ice-cold phosphate buffered saline (PBS) and collected.

OxiDIP assays in UV-treated and NAC-treated MCF10A cells

The 8-oxodG-enriched genomic fragments from UV- and NAC-treated MCF10A cells were obtained as previously described (15). Briefly, genomic DNA from UV- or NAC-treated MCF10A cells was extracted by using Dneasy Blood&Tissue kit (Qiagen). 10 μg of genomic DNA per immuno-precipitation were sonicated in 100 μl TE buffer (100 mM Tris–HCl pH 8.0, 0.5 M EDTA pH 8.0) using Bioruptor. 4 μg of fragmented DNA was denatured and immuno-precipitated overnight at 4°C with 4 μl of polyclonal antibodies against 8-oxodG (AB5830, Millipore) in a final volume of 500 μl IP buffer (110 mM NaH_2PO_4, 110 mM NaH_2PO_4 pH 7.4, 0.15 M NaCl, 0.05% Triton X-100, 100 mM Tris–HCl pH 8.0, 0.5 M EDTA pH 8.0). Then, 50 μl Dynabeads Protein G (ThermoFisher Scientific) was added for 3 h at 4°C, under constant rotation, and washed three times with 1 ml washing buffer (110 mM NaH_2PO_4, 110 mM Na_2HPO_4 pH 7.4, 0.15 M NaCl, 0.05% Triton X-100). The immune-complexes were then disrupted by incubation in 200 μl lysis buffer (50 mM Tris–HCl pH 8, 10 mM EDTA pH 8, 1% SDS, 0.5 mg/ml Proteinase K) for 4 h at 37°C, and 1 h at 52°C following addition of 100 μl lysis buffer. MinElute PCR purification kit (Qiagen) was used to purify the immune-precipitated ssDNA. All the steps of OxiDIP-seq protocol, including the washes of the immune-complexes, were carried out in low-light conditions, and 50 μM *N-tert*-butyl-α-phenylnitrone (B7263, Sigma) was added to each Dneasy Blood&Tissue buffer, IP and washing buffers, to preserve the oxidized DNA. The recovered ssDNA was converted to dsDNA using Random Primers DNA Labeling System (ThermoFisher Scientific).

Preparation of OxiDIP sequencing libraries

Library preparation was performed as described in (15) using 2 ng of DIP or Input DNA. Prior to sequencing, libraries were quantified using Qubit dsDNA HS Assay Kit (Invitrogen) and quality-controlled using Agilent Bioanalyzer. Samples from UV-treated MCF10A cells were sequenced generating 50 bp single-end reads using Illumina HiSeq 2000 platform according to standard operating procedures. Samples from NAC-treated MCF10A cells were sequenced generating 150 bp paired-end reads using Illumina NOVASEQ platform according to standard operating procedures.

OxiDIP analysis

OxiDIP-Seq was analyzed as indicated in (14,15). Briefly, reads were quality checked, filtered, and aligned to reference genome using NGS-QC Toolkit (27) and BWA (28), respectively, with default parameters. Subsequent, filtering and format conversion steps were performed with SAMtools (29) and Bedtools (30). UCSC genome browser was employed for data visualization. For the signal OxiDIP-Seq analysis, uniquely mapped reads were normalized over genomic input (log_2 8-oxodG/Input ratio) using the bamCompare tool from Deeptools suite (31), with SES method (32) as scaling factor. This should account for GC content sequencing bias, which would affect the pull-down samples and the inputs alike, as well as for the bias linked to different amounts of DNA.

Linear correlation of normalized 8-oxodG signals (expressed as log_2[8-oxodG/Input]) between UV/NAC-treated and untreated MCF10A cells was tested by means of Pearson's correlation test using the multi-BigwigSummary and the plotCorrelation tools from the Deeptools suite (31) with default parameters.

The comparison of 8-oxodG signals in UV or NAC-treated MCF10A cells was performed using the bamCompare tool from Deeptools suite (31) with exactScaling method as scaling factor, considering for each locus the read depth from OxiDIP-Seq of treated cells normalized over the same quantity from untreated cells (8-oxodG level as log_2[UV/Untr] and log_2[NAC/Untr]).

Metagene analysis and heatmaps were generated using the computeMatrix and plotHeatmap tools from the Deeptools suite with default parameters. Signal profile plots were derived using R starting from the matrices generated by the computeMatrix tool.

ChIP-Seq analysis

ChIP-Seq of PARP1, Pol II-S5P, Pol II-S2P, TOP2B, TOP1, gH2AX, NBS1, XRCC4 and RAD51 were downloaded from GSE93040; H3K27ac, H3K4me3 and H3K4me1 ChIP-Seq data were retrieved from GSE85158; SMC1, SA1 and SA2 were downloaded from GSE101921 and CTCF from E-MTAB-740, (Supplementary Table S4). ChIP-Seq reads (from GSE93040, GSE85158 and GSE101921) were quality checked and filtered with Trimmomatic (33). Alignments to the reference genome were performed with BWA using default parameters. SAMtools and Bedtools were used to perform filtering steps and format conversions. Uniquely mapped reads of each signal were normalized over genomic input ($\log_2$ ChIP/Input ratio) using the bamCompare tool from Deeptools suite, using SES method as scaling factor. For the CLAPS-Seq samples analyses (34), pre-processed data (bigwig files) were downloaded from GEO (GSE181312), and each signal was normalized over genomic input ($\log_2$ [ChIP/Input ratio]) using bigwigCompare tool from Deeptools suite. Pre-processed data (bigwig files) of acAPE1, acOGG1, AP-site and BG4 data (18) were downloaded from GEO (GSE142284), and each signal was normalized over IgG signal ($\log_2$ [ChIP/IgG]) using bigwigCompare tool from Deeptools suite. Metagene analysis and heatmaps were generated using computeMatrix and plotHeatmap tools from the Deeptools suite with default parameters, signal profile plots were derived using a custom script in R starting from the matrices generated by the computeMatrix tool.

Detection of active enhancers and super-enhancers and control set.

Active enhancers in MCF10A are detected by analyzing H3K27ac, H3K4me3 and H3K4me1 ChIP-Seq data as follows: H3K27ac peaks were called using MACS in the broad region calling mode and all the other parameters with default values. Then, we calculated the H3K4me3 signal (average $\log_2$ H3K4me3/Input ratio) in H3K27ac peaks mapping within and outside promoter regions ($n = 21\,074$, defined in (15) and determined the threshold of H3K4me3 enrichment as the equality point of the two density distributions. The H3K27ac peaks mapping outside promoters and showing H3K4me3 signal above the threshold were removed from the following analyses as they represent non-annotated TSSs/promoters. The H3K27ac peaks mapping outside promoters and showing H3K4me3 signal below the threshold were intersected with H3K4me1 peaks (detected using the same MACS settings as for H3K27ac peaks). Finally, H3K4me1-positive H3K27ac peaks are considered active enhancers ($n = 27\,256$).

Super-enhancers are then identified using the H3K27Ac ChIP-Seq data (35). In particular, the active enhancers are ranked and plotted based on their average H3K27Ac ChIP-Seq signal (average H3K27Ac/Input ratio). The plot (Supplementary Figure S3A) revealed a clear point in the rank where the H3K27Ac signal starts increasing rapidly. A geometric measurement (y-axis point for which a line with a slope of 1 was tangent to the curve) was used to identify a threshold value of H3K27ac marking this change in slope. We defined respectively super-enhancers the class of enhancers showing H3K27ac level higher than this threshold, and typical enhancers the ones showing H3K27ac levels lower than the threshold (Supplementary Figure S3B; (35,36). The overlap between the sets of super-enhancers and oxidized or control enhancer classes were determined using Bedtools. Statistical enrichment was determined by using Fisher's exact test.

Control (non-oxidized) enhancers ($n = 4520$) were defined as enhancers whose levels of 8-oxodG were lower than the bottom 20% of 8-oxodG values distribution. In particular, the 8-oxodG signal was binned over the active enhancers using a 100 bp bin size. Then, the level of the 8-oxodG signal in the enhancer was defined as the maximum 8-oxodG value computed overall the bins by using the computeMatrix tool from Deeptools suite with default parameters and custom R scripts.

To exclude potential confounding effects related to differences in the H3K27ac levels between oxidized and control enhancers (Supplementary Figure S5A), 150 enhancers were selected from each set, having uniformly distributed H3K27ac values ($\log_2$ [H3K27ac/Input]) ranging between 2.2 and 2.5 (respectively the control and the oxidized H3K27ac values medians) (Supplementary Figure S5B). The analyses of Pol2 S5P, Pol2 S2P, TOP1, TOP2B, PARP1, γH2AX, XRCC4 were repeated over the above defined subset of enhancers from each one of the two groups (Supplementary Figure S5C).

GC content and G4 enrichment analysis at the enhancer regions

For the quantification of GC content, the hg18.gc5Base track was retrieved from UCSC and the average GC content was assessed at oxidized and control enhancers using the computeMatrix tool from the Deeptools suite with default parameters.

The GC content and 8-oxodG levels relationship was investigated as follows: the oxidized enhancers were analyzed by dividing the corresponding regions in bins of 50 bp (Figure 1C) and then calculating the average GC content or 8-oxodG level per bin using the computeMatrix tool from the Deeptools suite with default parameters.

Bedtools suite was used to determine the overlap between the data sets of potential G-quadruplex sequences (PQSs) (37) and G4-containing regions (38) with oxidized or control enhancers. Statistical enrichment was determined by using Fisher's exact test. ComputeMatrix tool from Deeptools was used to analyze the normalized BG4 signal (18) at oxidized or control enhancers.

Motif Enrichment pathway analysis

Motif enrichment analysis of oxidized enhancers was performed by using HOMER suite tools (39). We provided the coordinates of the oxidized enhancers as input to the *findMotifs.pl* tool and obtained a list of enriched motifs (along with associated transcription factors) ranked by their enrichment *P*-values. We considered as statistically significant only the motifs with a reported *P*-value <1e–8. Enrichment pathway analysis was then performed by providing the list of transcription factors associated with the statistically sig-

nificant motifs to the Enrichr tool (40,41) and considering the MSigDB Hallmarks 2020 pathways.

GRO-seq analysis

GRO-seq data was obtained from Array-Express (E-MTAB-742). FASTQ files were aligned to the reference genome using Bowtie (42) with default parameters and allowing a maximum of two mismatches for the identification of uniquely mapping regions. Bidirectional transcription at oxidized or control enhancer regions was determined by analyzing the GRO-Seq signal (sense/antisense signal ratio) with the computeMatrix tool from the Deeptools suite with default parameters and *R* was used for plot generation.

Fusion breakpoints association analysis

We referred to 2822 fusion transcripts identified in patients with breast cancer containing 5147 unique fusion breakpoints (43). Bedtools suite was used to measure the number of fusion breakpoints in the oxidized or control CTCF-mCLs and normalized per loop length. The t-test was used for statistical significance.

ChIA-PET analysis

CTCF-ChIA-PET and RNAPII-ChIA-PET prefiltered and processed data were downloaded from the ENCODE project dataset, ENCSR403ZYJ and ENCSR499JGQ, respectively. Diffloop R/Bioconductor package (44) was employed to filter and analyze ChIA-PET processed dataset as follows: bedpe files of two available CTCF ChIA-PET replicates data were pooled and loaded by using the *loopsMake* tool. By using the *filterLoops* tool, we then filtered out loops whose width was less than 5kb or whose PET count was <3 in both replicates. Moreover, anchors that were separated by <5000 bp were merged using the *mergeAnchors* tool. CTCF loops were then annotated and filtered for the presence of a CTCF peak at the anchor regions by using *annotateLoops* tool. CTCF peaks were retrieved from the bed files of the dataset and only the intersections of overlapping CTCF peaks in both replicates were considered. CTCF loops were finally annotated and filtered for the presence, within the loop region, of (oxidized and non-oxidized) enhancers and promoters by using *loopGenes* tool. CTCF CCDs were determined by merging overlapping loops using Bedtools *merge*. Bedtools suite was used to measure the number of oxidized and non-oxidized SEs or promoters contained in oxidized CTCF-mCLs. Fisher's exact test was used for statistical significance.

A similar analysis was performed for the RNAPII ChIA-PET dataset. In particular, the bedpe files of the two available RNAPII ChIA-PET replicates data were pooled and loaded by using the *loopsMake* tool. We then filtered out loops whose width was less than 10 kb or whose PET count was <3 in both replicates. Moreover, anchors that were separated by <5000 bp were merged using the *mergeAnchors* tool. RNAPII loops were then annotated and filtered, for the presence at the anchor regions of RNAPII peaks and of oxidized and non-oxidized enhancers/promoters by using annotateLoops tool, in oxidized and non-oxidized RNAPII loops, respectively. RNAPII peaks were retrieved from the bed files of the above dataset considering only intersections of overlapping RNAPII peaks in both replicates. Bedtools suite was used to measure (i) the number of oxidized or non-oxidized SEs and promoters contained in oxidized RNAPII-mCLs and (ii) the number of oxidized or non-oxidized RNAPII loops that were entirely contained in oxidized CTCF-mCLs. Fisher's exact test was used for statistical significance.

HiC analysis

We used HiC data in MCF10A cell line binned at 40 kb resolution, taken from (45). For each chromosome, we first classify all the oxidized promoters-oxidized enhancers pairs (briefly oxi-oxi pairs) according to their genomic distance. Then we calculate the average HiC score for each genomic distance: where is the HiC score, is the number of pairs at genomic distance and the sum runs over the lines containing the detected oxi-oxi pairs (Figure 5C, left panel, dark blue squares). Analogously, we calculate the control score (briefly, ctrl-oxi pairs) by considering pairs with control promoters and oxidized enhancers (Figure 5C, left panel, light-blue squares). We then evaluate how many times and build the cumulative curve every 400 kb (i.e. every 10 diagonals) covering a distance range up to 40 Mb (Figure 5C, right panel, blue line). Obviously, in the above analysis, we include only distances where both and are well defined, that is diagonals where at least one oxi–oxi and ctrl–oxi pair have been detected, otherwise the distance is discarded. A standard gaussian filter is applied to HiC maps to avoid sparsity at long genomic distances. To exclude numerical effects due to the different number of oxidized and control promoters (we have typically less ctrl promoters in our datasets), we considered an average curve over 10 independent realizations with equal number of events, from which we estimated the standard deviation shown in Figure 5C. Random control curve (Figure 5C, right panel, orange line) is produced by applying the above-described analysis to oxi-oxi pairs and ctrl-oxi pairs where the position of the oxidized and control promoters is bootstrapped along the chromosomes. For each chromosome were considered overall 100 permutations. For each genomic distance, to evaluate the statistical significance we performed Mann-Whitney U test, the maximum p-value among the considered distances was used as reference.

Statistical analysis

Statistical significance of the observed differences in loop width and frequency of contact was evaluated by means of two-tailed *t*-test with heteroskedasticity assumption. Fisher's exact test was used to test the statistical enrichment of enhancers and promoters within oxidized and control sets. Mean value and standard deviation of each genomic signal at the enhancers, promoters or chromatin loops were calculated using the computeMatrix tool from Deeptools suite with default parameters, while the standard error was

calculated in R. Statistical differences of genomic signals between oxidized and control groups were assessed for each bin using Welch modified two sample *t*-test, implemented in the tsum.test function from the BSDA R package. This study was conducted using 0.05 as the significance threshold; all statistical analyses were performed using R version 4.0.5.

RESULTS

Identification of oxidized Enhancers in the human genome.

We recently described the genome-wide distribution of 8-oxodG in MCF10A cells using the OxiDIP-Seq (14) showing that OxiDIP-Seq peaks were enriched in intragenic regions and within a subset of promoter sequences (15). Because recent studies addressed the role of DNA damage at enhancers (35,46,47), we searched for specific enrichment of OxiDIP-Seq peaks at enhancer regions. We first analyzed the H3K4me3, H3K4me1 and H3K27ac ChIP-Seq signals and identified 27 256 active enhancers in MCF10A cells (Materials and Methods and Supplementary Figure S1A, B). Then, we intersected these active enhancers with the dataset of OxiDIP-Seq peaks (52,298) and found 841 active enhancers containing 839 OxiDIP-Seq peaks (Figure 1A, Supplementary Figures S1C, S2 and Supplementary Table S1). Analysis of a recently published 8-oxodG signal, derived by CLAPS-Seq method, confirms the presence of 8-oxodG accumulation at the oxidized enhancers (Supplementary Figure S1D).

To investigate the role of GC content on 8-oxodG accumulation at enhancers, we measured the GC composition of both oxidized and control enhancers and found a slight increase in the oxidized class (median oxidized = 45.3%, median control = 42.6%, $P < 2.2e{-}16$, Bonferroni adjusted pairwise *t*-test; Figure 1B). However, the analysis of the oxidized enhancers at a higher resolution level revealed a strong correlation between 8-oxodG and GC content signals with high GC content regions showing increasing levels of 8-oxodG (Figure 1C). We also found that oxidized enhancers are enriched for the presence of PQSs (37) (Figure 1D, $P < 2.2e{-}16$, Fisher's exact test) and G-quadruplets as determined by the analysis of the G4 dataset (19% of oxidized enhancers versus 7% of control enhancers, $P < 2.2e{-}16$ Fisher's exact test) and of the BG4 signal (Supplementary Figure S3A). Overall, these findings suggest that the accumulation of oxidatively-modified nucleobases within enhancers depends on GC content.

Next, to determine whether oxidized enhancers are associated with specific transcription factors and biological cellular processes, we used Hypergeometric Optimization of Motif Enrichment (HOMER tool) and found an enrichment of DNA sequence motifs ($P < 1e{-}8$) associated with the factors JUNB, CEBP proteins, GRE, TBR1, NFAT5, NKX6-3, EHF and PRDM1 (Figure 1E), these latter involved in several inflammatory response pathways. This is also confirmed by a pathway enrichment analysis based on the above-indicated transcription factors (Figure 1F).

Altogether, these data show that 8-oxodG accumulates in MCF10 cells at a subset of enhancers characterized by distinct genomic features and associated with inflammatory response activity.

Oxidized enhancers are enriched for super-enhancers.

In 2013, the term 'super-enhancer' (SE) has been coined to describe large, hyper-active regulatory regions containing multiple enhancers clustered in close genomic proximity, densely occupied by transcription factors and coactivators (35,36,48,49).

Hence, we questioned whether oxidized enhancers are SEs. To this end, we first selected 2405 SEs among the above described 27 256 enhancers in MCF10A based on the distribution of the H3K27ac ChIP-Seq values (Materials and Methods and Supplementary Figure S4). Then, we computed the SE/Enhancer ratio respectively overall and for the oxidized and control enhancers classes (Figure 2A; 'all', 'oxidized' and 'control') and found that the oxidized enhancers are enriched for SEs ($P < 2.2e{-}16$, Fisher's exact test).

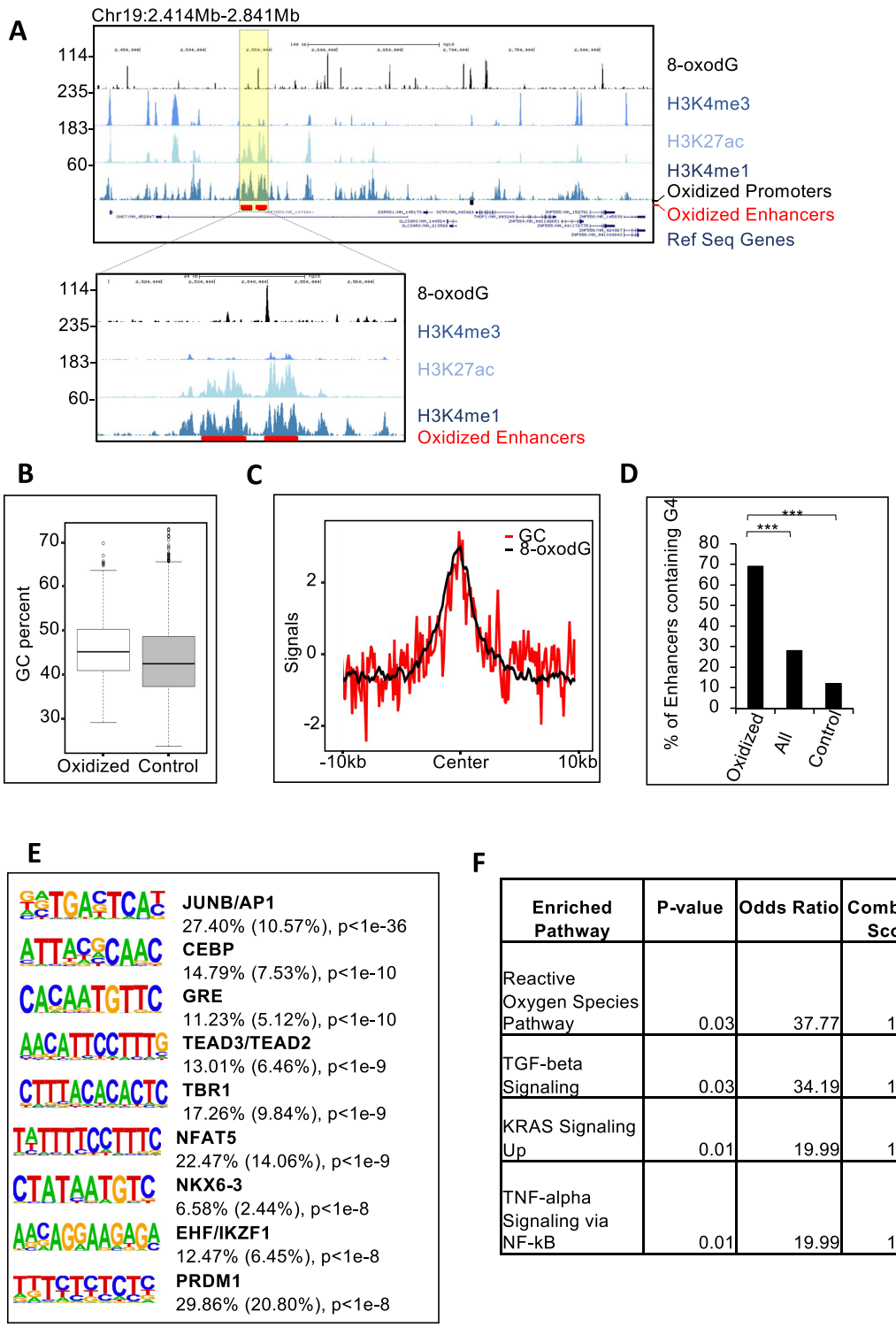
SEs are marked by RNA Polymerase II (RNAPII) loading which, in turn, generates enhancer RNAs (eRNAs) also through active bidirectional transcription (35,50). Therefore, we measured, in both oxidized and control enhancers, the levels of Ser5- and Ser2-phosphorylated isoforms of RNAPII (Pol II-S5P and Pol II-S2P) and two RNAPII-associated topoisomerases (TOP2B and TOP1). We found that oxidized enhancers show a marked presence of both RNAPII isoforms and of both topoisomerases (Figure 2B–E). To exclude potential confounding effects due to differences in H3K27ac levels between oxidized and control enhancers (Supplementary Figure S5A), we selected a subset of enhancers from the oxidized and the control sets having comparable H3K27ac distributions (Supplementary Figure S5B) and reanalyzed the levels of the two RNAPII isoforms and of the two topoisomerases. For all the analyzed signals, we confirmed the presence of higher levels of these latter at oxidized enhancers (Supplementary Figure S5C). Finally, by measuring the levels of eRNAs by GRO-Seq analysis in the regions surrounding control and oxidized enhancers, we found a marked presence of bidirectional transcription activity only in the loci marked by oxidized enhancers (Figure 2F).

Overall, these data show that oxidized enhancers are enriched for SEs associated with the bi-directionally transcribed eRNA.

Identification of oxidized CTCF and RNAPII-mediated chromatin loops.

SEs regulatory functions on target genes are typically restricted to discrete spatial units within CCCTC-binding factor (CTCF)-insulated chromatin domains and within topologically associating domains (TADs) (50–52). Therefore, we asked whether oxidized enhancers localize within specific CTCF and RNAPII-mediated chromatin loops.

We analyzed the public available CTCF ChIA-PET data from MCF10A (ENCODE 4 project (53)) and, using Diffloop tools, we identified 72 012 high-quality CTCF-mediated interactions, with 14 059 of them hosting enhancers and promoters within their loop regions. Next, we split these 14 059 CTCF-mediated interactions based on the presence of at least one oxidized enhancer within the loop region and we found 855 oxidized CTCF-mediated chromatin loops (oxidized CTCF-mCL, Supplementary Ta-



Enriched Pathway	P-value	Odds Ratio	Combined Score
Reactive Oxygen Species Pathway	0.03	37.77	133.69
TGF-beta Signaling	0.03	34.19	117.77
KRAS Signaling Up	0.01	19.99	101.78
TNF-alpha Signaling via NF-kB	0.01	19.99	101.78

Figure 1. (**A**) Genome browser screenshots showing, at the indicated genomic region, 8-oxodG, H3K4me3, H3K4me1, H3K27ac signals from MCF10A cells. Red bars indicsate the position of the oxidized enhancers and black bars indicate promoters. RefSeq Genes within the indicated genomic region are shown in the bottom panel. The expanded view of the yellow highlighted region in A shows overlapping peaks. (**B**) Box plot showing the GC content (%) distribution measured at oxidized (white) and control (gray) enhancers ($P < 2.2e-16$; Bonferroni adjusted pairwise t-test). (**C**) Mean-density profile of the normalized 8-oxodG signal (black) and average GC content (red) within a distance of ±10 kb from the center of the oxidized enhancers. (**D**) Bar plot reporting the percentage of oxidized enhancers containing G4 structures (left bar) compared to all the active enhancers (middle bar) and the control set (right bar) (*** $P < 2.2e{-}16$; Fisher's exact test). (**E**) Table showing the motif analysis of oxidized enhancers with the percentages of transcription factors motif enrichment, the background percentages (in brackets) and relative P values. (**F**) Table showing the pathway enrichment analysis by Enrichr based on the transcription factors indicated in (E).

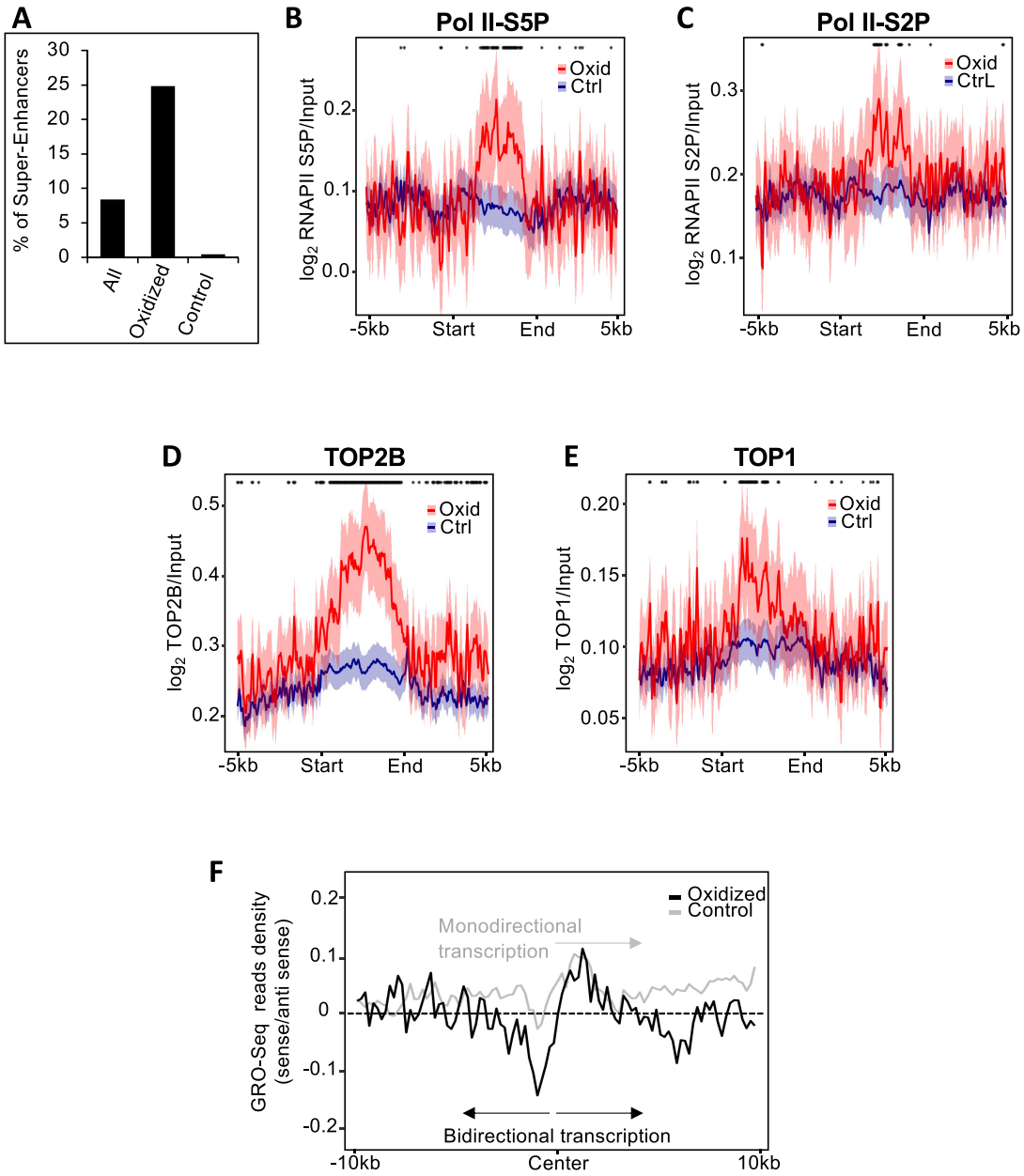


Figure 2. (**A**) Bar plot reporting the percentage of oxidized enhancers that are SEs compared to all active enhancers or the control set (*** $P < 2.2e{-}16$; Fisher's exact test). (**B–E**) Distribution of mean Pol II-S5P (B), Pol II-S2P (C), TOP2B (D) and TOP1 (E) signal profiles, normalized over the input DNA, at oxidized (red) and control (blue) enhancers (−5 kb from the Start to +5 kb from end) identified in MCF10A cells. The 95% confidence interval (mean ± 2 standard error) of the mean is indicated by the light-red or blue shaded areas respectively. In the top line, bins whose associated signal is significantly different between oxidized and control enhancers (Welch *t*-test, $P < 0.05$) are marked with an asterisk. (**F**) Plot showing the GRO-Seq read density, as sense/antisense transcripts ratio, at ±10 kb from the center of oxidized (black) and control (gray) enhancers.

ble S2). When compared to all non-oxidized chromatin loops (control CTCF-mCL), oxidized CTCF-mCL showed higher loop widths and similar contact frequency (Supplementary Figure S6A). In addition, we found that oxidized CTCF-mCL are localized in 199 chromatin contact domains (CCDs). Since the cohesin proteins complex is highly associated with CTCF in chromatin biology, we analyzed the CTCF-anchor reg ions of the oxidized CTCF-mCLs and looked at the co-occupancy with cohesin subunits SMC1, SA1 and SA2 (54–56). In line with the previous reports (54–56), we observed high co-occupancy levels of CTCF and cohesin proteins in both anchors (Figure 3A).

To test whether oxidized enhancers physically interact with oxidized promoters identified in our previous study (15), we analyzed the RNAPII ChIA-PET data from MCF10A (ENCODE 4 project, (53)). We found 55 753 high-quality RNAPII-mediated interactions and, based on the presence of anchor-promoters and/or anchor-enhancers, 23 447 interactions were identified as: enhancer–promoter (E–P = 15 224/23 447), promoter–promoter (P–P = 4663/23 447) and enhancer–enhancer

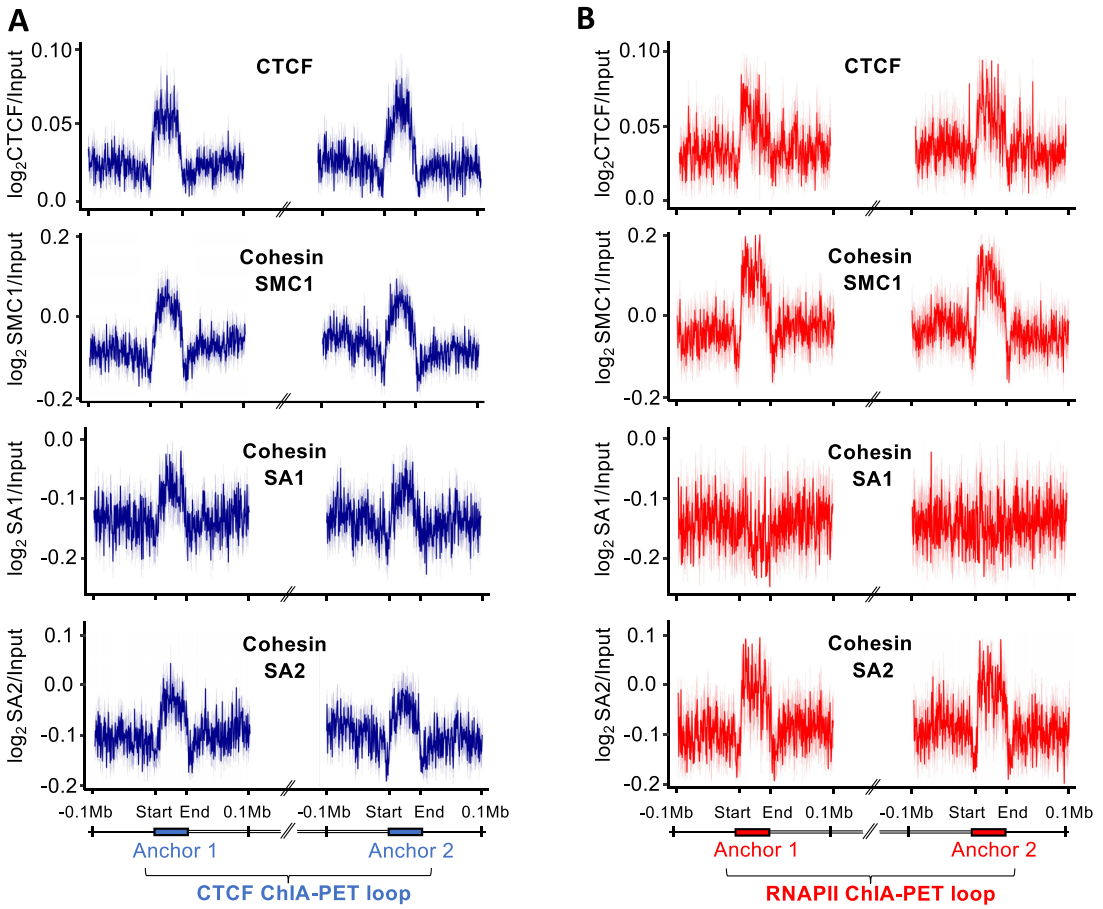


Figure 3. (**A**) Metaprofiles showing the distribution of mean CTCF and cohesins (SMC1, SA1 and SA2 subunits) signal profiles, normalized over the input DNA, at the anchor regions (−0.1 Mb from the Start to +0.1 Mb from end) of oxidized CTCF-mCLs identified in MCF10A cells. The 95% confidence interval (mean ± 2 standard error) of the mean is indicated by the light-blue shaded areas. (**B**) Metaprofiles showing the distribution of mean CTCF and cohesins (SMC1, SA1 and SA2 subunits) signal profile, normalized over the input DNA, at the anchor regions (−0.1 Mb from the Start to +0.1 Mb from End) of oxidized RNAPII-mCLs in MCF10A cells. The 95% confidence interval (mean ± 2 standard error) of the mean is indicated by the light-red shaded areas.

(E–E = 3560/23 447) interactions. Hence, we filtered E–P, P–P and E–E interactions for the presence of oxidized enhancers and/or oxidized promoters and identified a set of 408 oxidized RNAPII-mediated chromatin loops (oxidized RNAPII-mCL, Supplementary Table S3) characterized by anchor-oxidized-promoters and/or anchor-oxidized-enhancers. Intriguingly, the analysis of the oxidized RNAPII-mCL showed higher loop widths and contacts frequency when compared to their non-oxidized counterpart (control RNAPII-mCL) (Supplementary Figure S6B). Since it is known that the SA2 cohesin promotes contacts between enhancers and promoters independently of CTCF while SA1 is mainly involved in the stabilization of chromatin domain boundaries (54–56), we analyzed the CTCF, SMC1, SA1 and SA2 ChIP-Seq signals over the anchor regions of the oxidized RNAPII-mCL and found high co-occupancy levels of CTCF, SMC1 and SA2 and absence of SA1 (Figure 3B).

Thus, we identified oxidized CTCF-mediated chromatin loops (oxidized CCDs) and oxidized RNAPII-mediated loops in MCF10A cells.

Oxidized enhancers and promoters interact within oxidized CTCF-mediated chromatin loops

SEs are known to mediate long-range interactions with their target genes, within specific chromatin domains (36,48,51). We, therefore, asked whether oxidized SEs tend to cluster within oxidized CTCF-mCLs and whether they can physically interact with the oxidized promoters.

We counted the number of oxidized and control SEs and promoters harbored within the oxidized CTCF-mCL and found that 44% of the oxidized SEs versus 27% of control SEs (P = 1.4e–6, Fisher's exact test), and 31% of oxidized promoters versus 20% of the control class (P < 2.2e–16, Fisher's exact test) are contained in at least one oxidized CTCF-mCL (Figure 4A). A similar analysis was performed on oxidized RNAPII-mCLs and we found that they are enriched for oxidized SEs and promoters compared to the respective control set (37% of oxidized SEs versus 11% of control, P = 1.4e–13, Fisher's exact test, and 34% of oxidized promoters versus 11% of control, P < 2.2e–16, Fisher's exact test; Figure 4B).

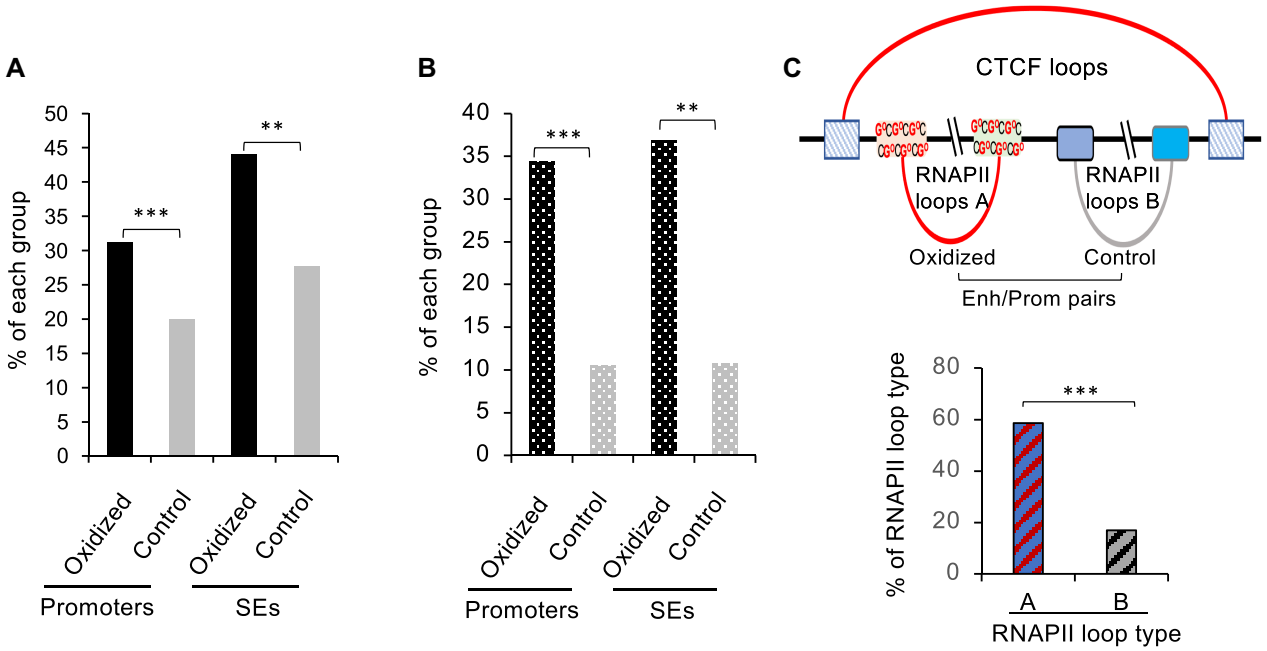


Figure 4. (**A**) Bar plot reporting the percentage of oxidized (black) vs control (gray) promoters or SEs (as indicated) contained within oxidized CTCF-mCLs (**P = 1.4e–6 for SEs and ***P < 2.2e–16 for promoters; Fisher's exact test). (**B**) Bar plot reporting the percentage of oxidized (dotted black bars) versus control (dotted gray bars) promoters and SEs contained within oxidized RNAPII-mCLs (**P = 1.4e–13 for SEs and ***P < 2.2e–16 for promoters; Fisher's exact test). (**C**) Top: a schematic representation of oxidized RNAPII-mCLs (RNAPII-loop A) and control set (RNAPII-loop B) within a CTCF-mCls, as indicated (Prom = promoter; Enh = enhancer). Bottom: Bar plot showing the percentage of RNAPII loop A versus control RNAPII loop B contained within oxidized CTCF-mCLs (***P < 2.2e–16; Fisher's exact test).

Then, we asked whether the interactions involving oxidized enhancers and promoters tend to occur more frequently within oxidized CTCF-mCLs. We found 59% of the oxidized RNAPII-mCLs versus 19% of control RNAPII-mCLs occur within oxidized CTCF-mCLs (Figure 4C, RNAPII Loop A and B, respectively; P < 2.2e–16, Fisher's exact test).

Genome inspection of oxidized CTCF-mCLs, oxidized RNAPII-mCLs, oxidized enhancers/promoters positioning and HiC data, confirmed the presence of anchor regions containing oxidized enhancers/promoters and the presence of oxidized E–P, P–P and E–E interaction within CTCF-mediated chromatin loops (Figure 5A–C and Supplementary Figure S7).

To further corroborate this finding, we analyzed HiC data, which gives information about the contact frequency between any pairs of genomic loci. Here, we used published available HiC data in MCF10A cells, available at 40 kb resolution (45). In analogy with the previously described analysis, we first measured the HiC score between loci containing oxidized promoters and oxidized enhancers and compared it with the HiC score between control promoters (15) and oxidized enhancers (see Materials and Methods). Importantly, the comparison was made by classifying promoter-enhancer pairs according to their genomic distance. Then, for each genomic distance, we calculated the average score of all the pairs (Figure 5D, left panel). Consistently with the above-described results, we found that the average score associated with oxidized pairs was higher than the average score associated with control pairs in ∼70% of the explored distance range (Figure 5D, left panel, blue curve). Such a fraction resulted significantly higher (P < 2e–3, Mann–Whitney U test) against a random control level which was, as expected, around 50% (Figure 5D, left panel, orange curve, Materials and Methods) for the entire distance range (up to 40Mb), suggesting that oxidized pairs tend to exhibit a higher interaction frequency than normal enhancer-promoter pairs.

Collectively, these data show that the oxidized enhancers tend to cluster within specific chromatin domains and physically interact with oxidized promoters.

Transcription-mediated DDR accumulation at oxidized CTCF-insulated chromatin loops.

Accumulation of 8-oxodG in MCF10A has been associated with SSBs and/or DSBs formation together with DDR activation (14,15).

Considering these findings, we analyzed the ChIP-Seq signals of PARP1 and γH2AX proteins, markers of SSBs and DSBs respectively, and we found that the oxidized enhancers show accumulation of PARP1 in their central region together with higher γH2AX levels compared to control enhancers (Figure 6A,B and Supplementary Figure S5C). In addition, we found accumulation of AP-sites, acOGG1 and acAPE1 at oxidized enhancers compared to the control set (Supplementary Figure S3B–D). This suggests that the accumulation of 8-oxodG at enhancer regions is associated with increased SSBs and/or DSBs formation and DDR activation.

DDR activation favors the recruitment of DSB-repair proteins such as RAD51 (involved in homologous recom-

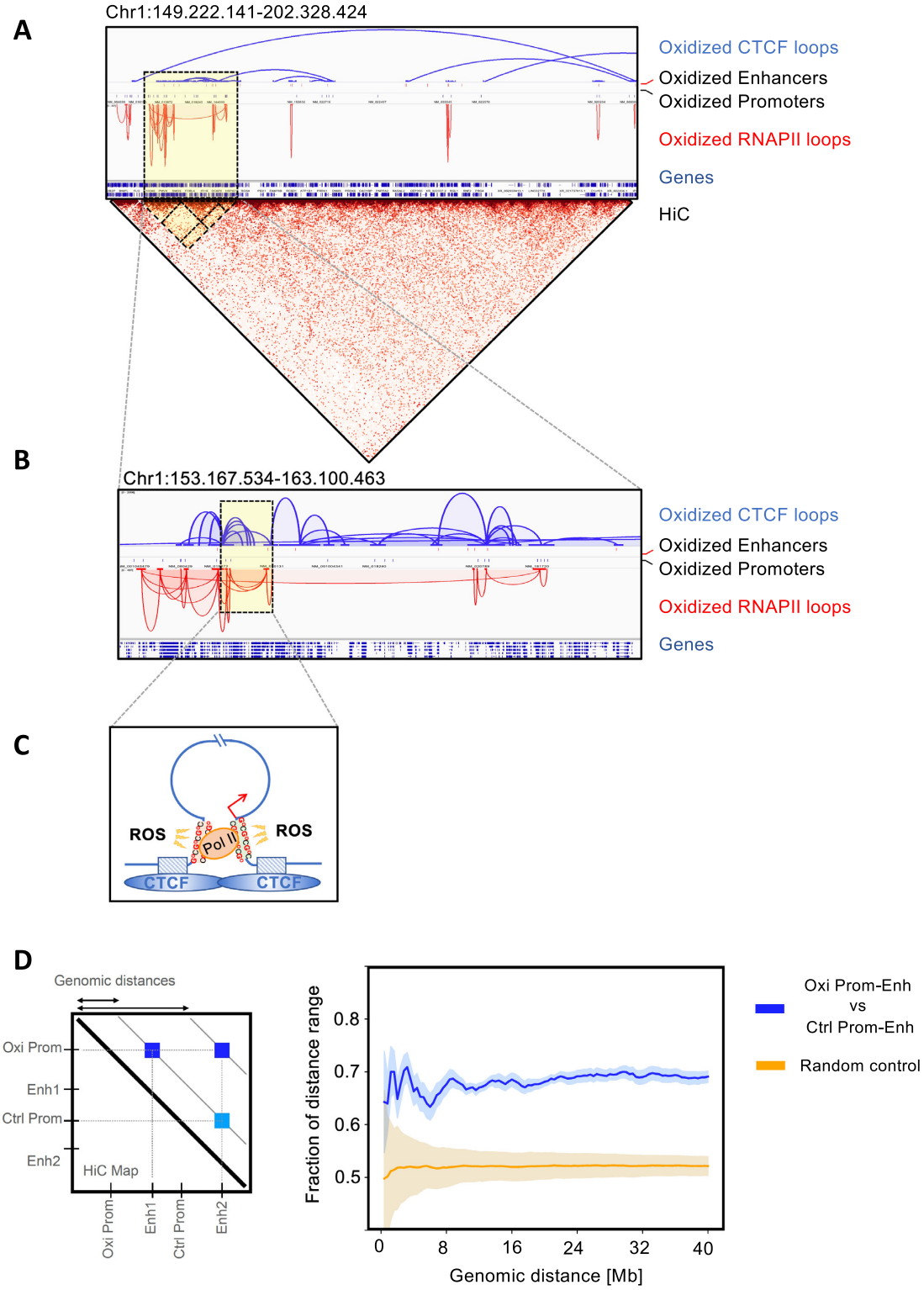


Figure 5. (**A**) Genome browser screenshot showing, at the indicated genomic region, the oxidized CTCF-mCLs and RNAPII-mCLs derived from ChIA-PET data. Oxidized enhancers, promoters, genes and the Hi-C contact heatmap are shown. (**B**) Zoom-in region highlighting oxidized enhancers and promoters (vertical red or blue bars, respectively) within anchor regions of CTCF- or RNAPII-mCLs (horizontal red or blue bars, respectively). (**C**) a schematic representation of a model of the CTCF-mCL highlighted in (B). (**D**) Left: for each genomic distance represented in the upper diagonals and each enhancer and promoter pair and consider the HiC score is reported in a chromatic scale. Right: fraction of genomic distances where the average HiC score associated with oxidized enhancer-promoter pairs is higher than average HiC score of control promoters-enhancer pairs (blue curve). Oxidized and control promoters randomly permuted along the genome (Random control) give an average fraction around 50% (light-orange or blue shaded areas indicate standard deviations).

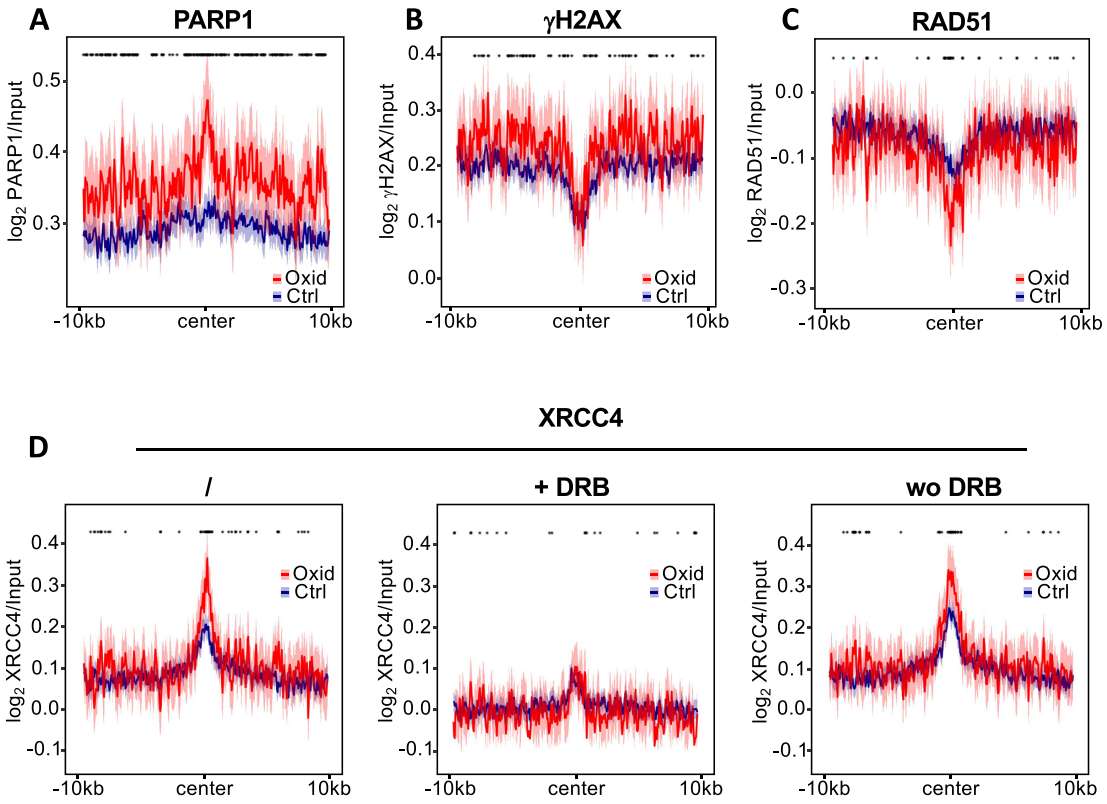


Figure 6. (**A–C**) Distribution of mean PARP1 (A), γH2AX (B) and RAD51 (C) signal profiles, normalized over the input DNA, at ± 10 kb region from the center of the oxidized (red line) and control (blue line) enhancers. The 95% confidence interval (mean ± 2 standard error) of the mean is indicated by the light-red or blue shaded areas. In the top line, bins whose associated signal is significantly different between oxidized and control enhancers (Welch *t*-test, $P < 0.05$) are marked with an asterisk. (**D**) Distribution of mean XRCC4 signal profile, normalized over the input DNA, at ±10 kb region from the center of the oxidized (red) and control (blue) enhancers in untreated (/), DRB treated (+DRB) and DRB washed-out (wo DRB) MCF10A cells. The 95% confidence interval (mean ± 2 standard error) of the mean is indicated by the light-red or blue shaded areas. In the top line, bins whose associated signal is significantly different between oxidized and control enhancers (Welch *t*-test, $P < 0.05$) are marked with an asterisk.

bination, HR) or XRCC4 (non-homologous end-joining, NHEJ) (57,58). Following analyses of RAD51 and XRCC4 ChIP-Seq signals (46), we found that XRCC4 was recruited with higher levels at the oxidized enhancers (Figure 6C, D, / = untreated cells and Supplementary Figure S5C). Intriguingly, we also found a sharp reduction of XRCC4 levels in DRB-treated cells (Figure 6D, +DRB) followed by full recovery after DRB removal (Figure 6D, woDRB). Moreover, similar results were obtained for oxidized promoters (Supplementary Figure S8A), suggesting that transcription inhibition decrease the XRCC4 levels at the oxidized promoter and enhancer regions.

Since the oxidized enhancers and promoters accumulate DSBs and form clusters within specific chromatin domains, we asked whether the oxidized CTCF-mCLs host DSBs and DSB-repair proteins more than the control ones. We analyzed ChIP-Seq signals of PARP1 and γH2AX and RAD51 or XRCC4 at the oxidized and control CTCF-mCLs including a +/− 2Mb flanking region. Strikingly, we found higher accumulation levels of PARP1, γH2AX and NBS1 at both anchor and loop regions of the oxidized CTCF-mCLs compared with the control ones (Figure 7A–C, and Supplementary Figure S8B). In addition, ChIP-Seq signal analysis of the DSBs-repair proteins RAD51 and XRCC4 revealed lower levels of RAD51 in the oxidized compared to control CTCT-mCLs (Figure 7D), with the RAD51 signal reaching its highest levels within the loop region and its lowest at the anchors of both oxidized and control CTCF-mCLs. Conversely, the XRCC4 signal showed a mirroring distribution compared to that of RAD51 (Figure 7E, / = untreated cells) with the oxidized CTCF-mCLs showing the highest levels at the anchor regions. Moreover, similarly to oxidized enhancers and promoters, also the oxidized CTCF-mCLs showed an XRCC4 occupancy that is dependent on the RNAPII elongation activity, with an almost absent XRCC4 signal in DRB-treated cells which is completely recovered after DRB removal (Figure 7E, +DRB and woDRB).

Since NHEJ, an error-prone pathway, is associated with oxidized CTCF-mCLs, we asked whether oxidized CTCF-mCLs were associated with chromosomal instability. To this end, we analyzed a dataset containing 2822 fusion transcripts from breast cancer patients (43) and derived the number of fusion breakpoints localized in the oxidized and control CTCF loops respectively. We found that, compared with control CTCF loops, the oxidized CTCF loops contain, on average, a higher number of breakpoints per loop (median = 0.33 versus 0.12, $P = 1.0e{-}2$, pairwise *t*-test, Figure 7F).

Taken together, these data indicate that endogenous oxidized CTCF-mCLs mark intrinsic fragile chromatin do-

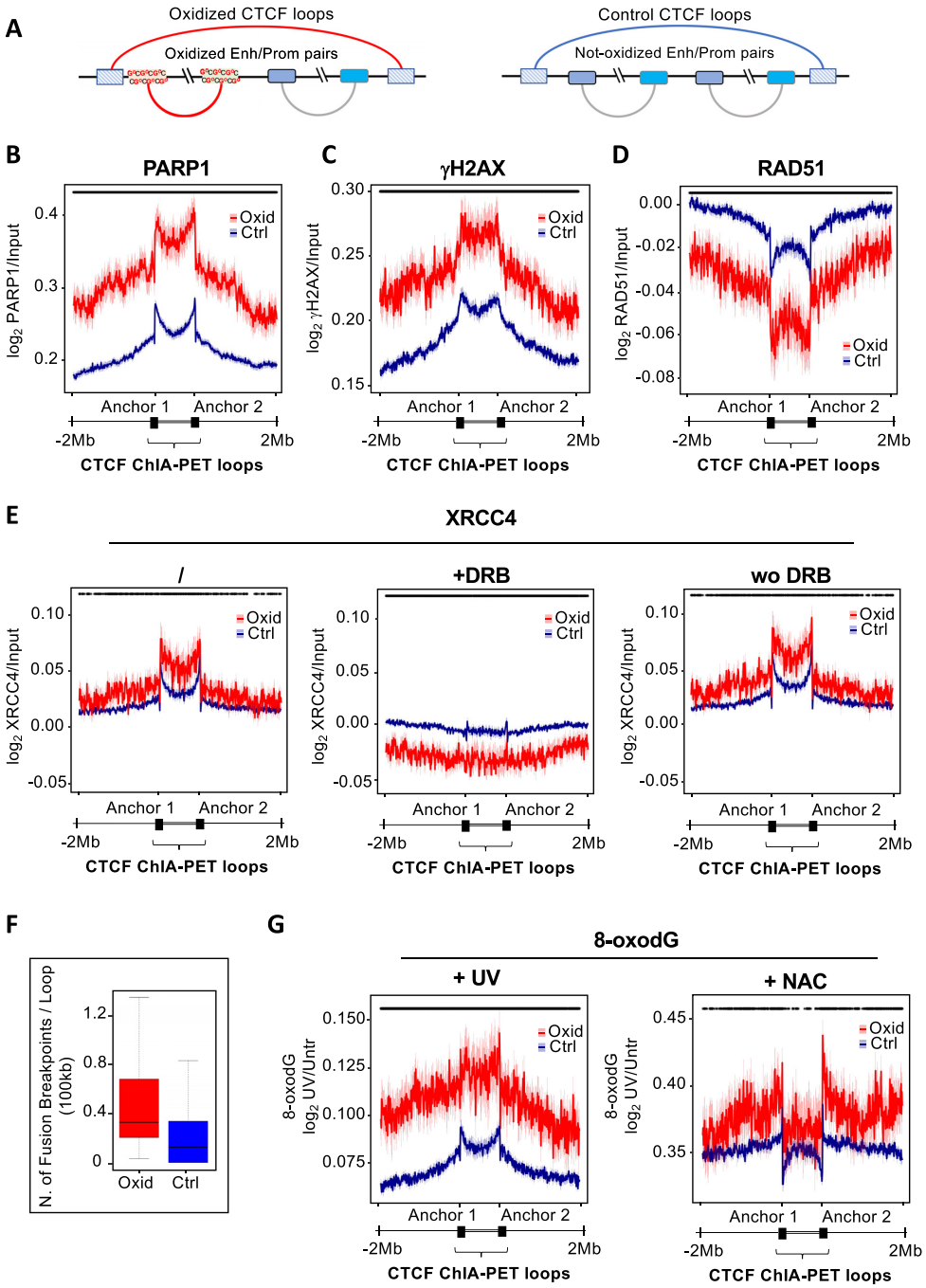


Figure 7. (**A**) a schematic representation of a model of oxidized (left) and control (right) CTCF-mCLs. (**B**, **C**) Distribution of mean PARP1 (B), γH2AX (C) and RAD51 (D) signal profiles, normalized over the input DNA, at the oxidized (red line) and control (blue line) CTCF-mCLs (−2 Mb region from the loop start site, Start, to +2 Mb region from loop end site, End). The 95% confidence interval (mean ± 2 standard error) of the mean is indicated by the light-red or blue shaded areas. In the top line, bins whose associated signal is significantly different between oxidized and control enhancers (Welch *t*-test, $P < 0.05$) are marked with an asterisk. (**E**) Distribution of mean XRCC4 signal profile, normalized over the input DNA, at the oxidized (red) and control (blue) CTCF-mCLs (−2 Mb from the loop start site, Start, to +2 Mb from loop end site, end) in untreated (/), DRB treated (+DRB) and DRB washed-out (wo DRB) MCF10A cells. The 95% confidence interval (mean ± 2 standard error) of the mean is indicated by the light-red or blue shaded areas. In the top line, bins whose associated signal is significantly different between oxidized and control enhancers (Welch *t*-test, $P < 0.05$) are marked with an asterisk. (**F**) Bar plot showing the number of Fusion Breakpoints identified in breast cancers, normalized per loop width (100 kb), in oxidized and control CTCF-mCLs ($P = 1.0e{-}2$; pairwise *t*-test). (**G**) Left panel (+UV), distribution of mean 8-oxodG signal as ratio of UV-treated over the untreated MCF10A cells ($\log_2$[UV/Untr]) at the oxidized (red line) and control (blue line) CTCF-mCLs (−2 Mb region from the loop start site, Start, to +2 Mb region from loop end site, end). The 95% confidence interval (mean ± 2 standard error) of the mean is indicated by the light-red or blue shaded areas. Right panel (+NAC), distribution of mean 8-oxodG signal as ratio of NAC-treated over untreated MCF10A cells ($\log_2$[NAC/Untr]) at the oxidized (red line) and control (blue line) CTCF-mCLs (−2 Mb region from the loop start site, Start, to +2 Mb region from loop end site, end). The 95% confidence interval (mean ± 2 standard error) of the mean is indicated by the light-red or blue shaded areas. In the top line, bins whose associated signal is significantly different between oxidized and control enhancers (Welch *t*-test, $P < 0.05$) are marked with an asterisk.

mains. To further corroborate this finding, since UV irradiation is known to increase genomic 8-oxodG levels (14,15), we assessed whether 8-oxodG levels increase at enhancers and at endogenous oxidized CTCF-mCLs in UV irradiated cells. Hence, we performed OxiDIP-Seq using genomic DNA from UV-treated MCF10A cells. Firstly, we found only a moderate correlation between the 8-oxodG levels measured in UV-treated and in untreated cells (Pearson's coefficient, r = 0.70; Supplementary Figure S9A), indicating the existence of genomic regions with different 8-oxodG levels that are linked to the UV the treatment. In particular, we found an increase of 8-oxodG levels at enhancers/SEs (Supplementary Figures S9B and C), gene loci (Supplementary Figure S9D) and at endogenous oxidized CTCF-mCLs (Figure 7G, +UV). Notably, the analyses of CLAPS-Seq experiments, performed in untreated and potassium bromate-treated cells (where the potassium bromate is used as a strong oxidizing agent), shows again accumulation of 8-oxodG at endogenous oxidized CTCF-mCLs (Supplementary Figure S10). Finally, we treated MCF10A cells with the N-acetyl-cysteine (NAC), used as ROS scavenger to reduce the genome 8-oxodG levels (14,15), and performed OxiDIP-Seq. Analyses of 8-oxodG levels in NAC-treated vs untreated cells revealed a weak correlation between oxidations signals in these two conditions (Pearson's coefficient, $r = 0.47$; Supplementary Figure S11A). In particular, we observed a depletion of endogenous 8-oxodG levels at enhancers/SEs (Supplementary Figure S11B, C), gene loci (Supplementary Figure S11D) and CTCF-mCLs (Figure 7G, +NAC).

DISCUSSION

Here, we report data suggesting the existence of human 8-oxodG-enriched enhancer regions in human epithelial mammary MCF10A cells. 8-oxodG accumulation at these enhancers is positively correlated with the GC content. Moreover, in agreement with previous findings (1,14,15,24,25,59–61), oxidized enhancers are enriched for PQSs and G4 DNA structures. We found that the 8-oxodG-enriched enhancers are associated with specific transcription factors involved in the inflammatory response pathways, as observed for OGG1-associated promoters (62,63).

We found that oxidized enhancers are bidirectionally transcribed and enriched in super-enhancers. SEs are a class of regulatory elements that participate in the transcriptional regulation of key master regulators of cellular identity. SEs are characterized by bidirectionally transcribed eRNA and are known to mediate long-range interactions with their target genes within CTCF-insulated chromatin domains and within TADs (35,36,48–52).

Based on the above-described model, we identified 855 CTCF-mCLs (which define 199 CTCF CCDs) and 408 RNAPII-mCLs. These loops are enriched for the presence of both oxidized SEs and promoters, thus suggesting that 8-oxodG-enriched regulatory regions tend to cluster within specific chromatin domains. The CTCF-mCLs anchors show co-occurrence of CTCF with both SA1 and SA2 cohesins and this is in agreement with previous data showing that SA1 and SA2 together with CTCF have a structural role in a sub-TAD chromatin loop (54–56). On the other hand, only SA2 is present in the RNAPII-mCLs anchors, thus supporting the critical role of SA2 in the establishment of functional SE-promoter intra-TAD contacts (54–56). Finally, we found that oxidized enhancers/SEs are physically associated with oxidized promoters within the oxidized CTCF-mCLs, suggesting that the oxidation of the SEs and target gene promoters simultaneously occur during the transcription process. 8-oxodG, together with its repair intermediates, plays a temporary epigenetic role promoting up- or downregulation of gene expression (1,24,25,64). Our findings reinforce this model and suggest that the 8-oxodG accumulation at enhancer-promoter pair contacts is likely the result of crosstalk occurring between transcription protein complexes recruited at enhancer and promoter regions.

We also demonstrate that 8-oxodG accumulation at the enhancers is associated with the presence of AP-sites, OGG1/APE1 binding and with both SSBs (as marked by the presence of nick sensor PARP1) and DSBs (as marked by γH2AX accumulation) formation along with DDR activation. This is consistent with data reported for specific promoter regions (15) and supports the model where an incomplete 8-oxodG repair contributes to the DSBs formation through the accumulation of SSBs intermediates and by triggering DDR activation. Intriguingly, the oxidized CTCF-mediated chromatin loops display SSBs and DSBs accumulation. These findings suggest that 8-oxodG-associated DNA damage accumulation and DDR activation is not circumscribed at oxidized enhancer-promoter pairs only, but it spreads along the CTCF-mCLs while being insulated by CTCFs. Notably, these fragile chromatin domains, as well as the associated oxidized enhancers and promoters, accumulate specifically XRCC4 and not RAD51, thus suggesting that the DSBs are repaired by the NHEJ pathway. Finally, the recruitment of XRCC4 at the damaged regions is sensitive to transcription inhibition, thus suggesting that the fragility of the identified chromatin domains is linked to a transcription-dependent mechanism.

Comparing OxiDIP-Seq experiments performed in untreated, UV-irradiated and NAC-treated cells, we found that enhancers, gene loci and their containing CTCF-mCLs constitute hot-spots of endogenous- and/or exogenous-generated 8-oxodG. This suggests that the identified 8-oxodG-enriched regions may represent intrinsically fragile loci or, alternatively, that transcription-associated mechanisms or structures (i.e. G4 and R-loops) make these regions more sensitive to damage events. Notably, similar results were obtained by analyzing the 8-oxodG signals derived from CLAPS-Seq.

Collectively, we propose a model in which 8-oxodG drives the transcription process by promoting, at CTCF-insulated domains, chromatin looping and crosstalk between protein complexes recruited to enhancers and promoter regions. However, 8-oxodG also makes these regions more prone to accumulate DSBs, that are in turn repaired via NHEJ pathway. Under oxidative or oncogenic stress, the rate of 8-oxodG incomplete repair events may increase within the identified intrinsically fragile chromatin regions, thus generating accumulation of NHEJ-repaired DSBs, genome instability and eventually tumorigenesis. This model might be particularly relevant in breast cancer whereby we found that 8-oxodG associates with enrichment of fusion breakpoints

and chromosomal translocation events (known byproduct of NHEJ) at oxidized CTCF-mCLs. Intriguingly, a recent study demonstrated that the removal of 5mC (a well-known epigenetic marker), via an oxidative demethylation process, within neuronal enhancer regions causes SSBs formation leading to the hypothesis that failure of the BER-mediated SSB repair at enhancers sites could lead to aberrant phenotypes (47).

Overall, our work identified oxidized enhancers marking specific fragile chromatin domains and provides mechanistic insights on the fragility of these genomic regions and the functional implications of 8-oxodG in genome instability.

DATA AVAILABILITY

Sequencing data have been submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession no. GSE194053.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Author contributions: F.G., S.A. and S.A. (Amente) carried out OxiDIP-Seq in UV/NAC-treated MCF10A cells. G.S., F.G. and S.A. (Amente) carried out bioinformatics and statistical analyses. A.M.C. performed HiC analysis under the supervision of M.N. S.A. (Amente), G.S., B.M., S.A. and L.L. interpreted data. S.A. (Amente) conceived and coordinated the work and wrote the manuscript. All authors read and commented on the manuscript.

FUNDING

POR Campania [FESR 2014–2020 'SATIN' to S.A.]; AIRC [IG13173 to B.M.]; G.S. acknowledges support from PON AIM 2014–2020 [E69F19000070001]. Funding for open access charge: AIRC.
Conflict of interest statement. None declared.

REFERENCES

1. Gorini,F., Scala,G., Cooke,M.S., Majello,B. and Amente,S. (2021) Towards a comprehensive view of 8-oxo-7,8-dihydro-2′-deoxyguanosine: highlighting the intertwined roles of DNA damage and epigenetics in genomic instability. *DNA Repair (Amst).*, **97**, 103027.
2. Aust,A.E. and Eveleigh,J.F. (1999) Mechanisms of DNA oxidation. In: *Proceedings of the Society for Experimental Biology and Medicine*.
3. Fleming,A.M. and Burrows,C.J. (2020) On the irrelevancy of hydroxyl radical to DNA damage from oxidative stress and implications for epigenetics. *Chem. Soc. Rev.*, **49**, 6524–6528.
4. Batra,V.K., Beard,W.A., Hou,E.W., Pedersen,L.C., Prasad,R. and Wilson,S.H. (2010) Mutagenic conformation of 8-oxo-7,8-dihydro-2′-dGTP in the confines of a DNA polymerase active site. *Nat. Struct. Mol. Biol.*, **17**, 889–890.
5. Lindahl,T. (1990) Repair of intrinsic DNA lesions. *Mutat. Res. Genet. Toxicol.*, **858–860**, 503273.
6. Markkanen,E. (2017) Not breathing is not an option: how to deal with oxidative DNA damage. *DNA Repair (Amst).*, **59**, 82–105.
7. Ohno,M., Miura,T., Furuichi,M., Tominaga,Y., Tsuchimoto,D., Sakumi,K. and Nakabeppu,Y. (2006) A genome-wide distribution of 8-oxoguanine correlates with the preferred regions for recombination and single nucleotide polymorphism in the human genome. *Genome Res.*, **16**, 567–475.
8. Yoshihara,M., Jiang,L., Akatsuka,S., Suyama,M. and Toyokuni,S. (2014) Genome-wide profiling of 8-oxoguanine reveals its association with spatial positioning in nucleus. *DNA Res.*, **21**, 603–612.
9. Cao,B., Wu,X., Zhou,J., Wu,H., Liu,L., Zhang,Q., DeMott,M.S., Gu,C., Wang,L., You,D. *et al.* (2020) Nick-seq for single-nucleotide resolution genomic maps of DNA modifications and damage. *Nucleic Acids Res.*, **48**, 6715–6725.
10. Amente,S., Scala,G., Majello,B., Azmoun,S., Tempest,H.G., Premi,S. and Cooke,M.S. (2021) Genome-wide mapping of genomic DNA damage: methods and implications. *Cell. Mol. Life Sci.*, **78**, 6745–6762.
11. Ding,Y., Fleming,A.M. and Burrows,C.J. (2017) Sequencing the mouse genome for the oxidatively modified base 8-Oxo-7,8-dihydroguanine by OG-Seq. *J. Am. Chem. Soc.*, **139**, 2569–2572.
12. Wu,J., McKeague,M. and Sturla,S.J. (2018) Nucleotide-resolution genome-wide mapping of oxidative DNA damage by click-code-seq. *J. Am. Chem. Soc.*, **140**, 9783–9787.
13. Poetsch,A.R., Boulton,S.J. and Luscombe,N.M. (2018) Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis 06 biological sciences 0604 genetics. *Genome Biol.*, **19**, 215.
14. Amente,S., Di Palo,G., Scala,G., Castrignanò,T., Gorini,F., Cocozza,S., Moresano,A., Pucci,P., Ma,B., Stepanov,I. *et al.* (2019) Genome-wide mapping of 8-oxo-7,8-dihydro-2′-deoxyguanosine reveals accumulation of oxidatively-generated damage at DNA replication origins within transcribed long genes of mammalian cells. *Nucleic Acids Res.*, **47**, 221–236.
15. Gorini,F., Scala,G., Di Palo,G., Dellino,G.I., Cocozza,S., Pelicci,P.G., Lania,L., Majello,B. and Amente,S. (2020) The genomic landscape of 8-oxodG reveals enrichment at specific inherently fragile promoters. *Nucleic Acids Res.*, **48**, 4309–4324.
16. Liu,Z.J., Martínez Cuesta,S., van Delft,P. and Balasubramanian,S. (2019) Sequencing abasic sites in DNA at single-nucleotide resolution. *Nat. Chem.*, **11**, 629–637.
17. Fang,Y. and Zou,P. (2020) Genome-Wide mapping of oxidative DNA damage via engineering of 8-Oxoguanine DNA glycosylase. *Biochemistry*, **59**, 85–89.
18. Roychoudhury,S., Pramanik,S., Harris,H.L., Tarpley,M., Sarkar,A., Spagnol,G., Sorgen,P.L., Chowdhury,D., Band,V., Klinkebiel,D. *et al.* (2020) Endogenous oxidized DNA bases and APE1 regulate the formation of G-quadruplex structures in the genome. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 11409–11420.
19. Pan,L., Zhu,B., Hao,W., Zeng,X., Vlahopoulos,S.A., Hazra,T.K., Hegde,M.L., Radak,Z., Bacsi,A., Brasier,A.R. *et al.* (2016) Oxidized guanine base lesions function in 8-oxoguanine DNA glycosylase-1-mediated epigenetic regulation of nuclear factor κB-driven gene expression. *J. Biol. Chem.*, **291**, 25553–25566.
20. Perillo,B., Ombra,M.N., Bertoni,A., Cuozzo,C., Sacchetti,S., Sasso,A., Chiariotti,L., Malorni,A., Abbondanza,C. and Avvedimento,E.V. (2008) DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. *Science*, **319**, 202–206.
21. Amente,S., Bertoni,A., Morano,A., Lania,L., Avvedimento,E.V. and Majello,B. (2010) LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc-induced transcription. *Oncogene*, **29**, 3691–702.
22. Zuchegna,C., Aceto,F., Bertoni,A., Romano,A., Perillo,B., Laccetti,P., Gottesman,M.E., Avvedimento,E.V. and Porcellini,A. (2014) Mechanism of retinoic acid-induced transcription: histone code, DNA oxidation and formation of chromatin loops. *Nucleic Acids Res.*, **42**, 11040–11055.
23. Pezone,A., Taddei,M.L., Tramontano,A., Dolcini,J., Boffo,F.L., De Rosa,M., Parri,M., Stinziani,S., Comito,G., Porcellini,A. *et al.* (2020) Targeted DNA oxidation by LSD1-SMAD2/3 primes TGF-β1/EMT genes for activation or repression. *Nucleic Acids Res.*, **48**, 8943–8958.
24. Fleming,A.M., Ding,Y. and Burrows,C.J. (2017) Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 2604–2609.
25. Fleming,A.M. and Burrows,C.J. (2021) Oxidative stress-mediated epigenetic regulation by G-quadruplexes. *NAR Cancer*, **3**, zcab038.
26. Perillo,B., Ombra,M.N., Bertoni,A., Cuozzo,C., Sacchetti,S., Sasso,A., Chiariotti,L., Malorni,A., Abbondanza,C. and Avvedimento,E.V. (2008) DNA oxidation as triggered by H3K9me2

demethylation drives estrogen-induced gene expression. *Science*, **319**, 202–206.
27. Patel,R.K. and Jain,M. (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*, **7**, e30619.
28. Li,H. and Durbin,R. (2010) Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics*, **26**, 589–595.
29. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
30. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
31. Ramírez,F., Dündar,F., Diehl,S., Grüning,B.A. and Manke,T. (2014) DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.*, **42**, W187–W91.
32. Diaz,A., Park,K., Lim,D.A. and Song,J.S. (2012) Normalization, bias correction, and peak calling for chip-seq. *Stat. Appl. Genet. Mol. Biol.*, **11**, 9.
33. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
34. An,J., Yin,M., Yin,J., Wu,S., Selby,C.P., Yang,Y., Sancar,A., Xu,G.-L., Qian,M. and Hu,J. (2021) Genome-wide analysis of 8-oxo-7,8-dihydro-2′-deoxyguanosine at single-nucleotide resolution unveils reduced occurrence of oxidative damage at G-quadruplex sites. *Nucleic Acids Res.*, **49**, 12252–12267.
35. Meng,F.L., Du,Z., Federation,A., Hu,J., Wang,Q., Kieffer-Kwon,K.R., Meyers,R.M., Amor,C., Wasserman,C.R., Neuberg,D. *et al.* (2014) Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell*, **159**, 1538–1548.
36. Whyte,W.A., Orlando,D.A., Hnisz,D., Abraham,B.J., Lin,C.Y., Kagey,M.H., Rahl,P.B., Lee,T.I. and Young,R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
37. Chambers,V.S., Marsico,G., Boutell,J.M., Di Antonio,M., Smith,G.P. and Balasubramanian,S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.*, **33**, 877–881.
38. Hänsel-Hertsch,R., Beraldi,D., Lensing,S.V., Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M. *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.
39. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-Regulatory elements required for macrophage and b cell identities. *Mol. Cell*, **38**, 576–589.
40. Chen,E.Y., Tan,C.M., Kou,Y., Duan,Q., Wang,Z., Meirelles,G.V., Clark,N.R. and Ma'ayan,A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, **14**, 128.
41. Kuleshov,M.V., Jones,M.R., Rouillard,A.D., Fernandez,N.F., Duan,Q., Wang,Z., Koplev,S., Jenkins,S.L., Jagodnik,K.M., Lachmann,A. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
42. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S. (2009) 2C-Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
43. Yoshihara,K., Wang,Q., Torres-Garcia,W., Zheng,S., Vegesna,R., Kim,H. and Verhaak,R.G.W. (2015) The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*, **34**, 4845–4854.
44. Lareau,C.A. and Aryee,M.J. (2018) Diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data. *Bioinformatics*, **34**, 672–674.
45. Barutcu,A.R., Lajoie,B.R., McCord,R.P., Tye,C.E., Hong,D., Messier,T.L., Browne,G., van Wijnen,A.J., Lian,J.B., Stein,J.L. *et al.* (2015) Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol.*, **16**, 214.
46. Dellino,G.I., Palluzzi,F., Chiariello,A.M., Piccioni,R., Bianco,S., Furia,L., De Conti,G., Bouwman,B.A.M., Melloni,G., Guido,D. *et al.* (2019) Release of paused RNA polymerase II at specific loci favors DNA double-strand-break formation and promotes cancer translocations. *Nat. Genet.*, **51**, 1011–1023.
47. Wu,W., Hill,S.E., Nathan,W.J., Paiano,J., Callen,E., Wang,D., Shinoda,K., van Wietmarschen,N., Colón-Mercado,J.M., Zong,D. *et al.* (2021) Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature*, **593**, 440–444.
48. Wang,X., Cairns,M.J. and Yan,J. (2019) Super-enhancers in transcriptional regulation and genome organization. *Nucleic Acids Res.*, **47**, 11481–11496.
49. Cheng,M., Zhang,Z.W., Ji,X.H., Xu,Y., Bian,E. and Zhao,B. (2020) Super-enhancers: a new frontier for glioma treatment. *Biochim. Biophys. Acta - Rev. Cancer*, **1873**, 188353.
50. Yamagata,K., Nakayamada,S. and Tanaka,Y. (2020) Critical roles of super-enhancers in the pathogenesis of autoimmune diseases. *Inflamm. Regen.*, **40**, 16.
51. Li,G.H., Qu,Q., Qi,T.T., Teng,X.Q., Zhu,H.H., Wang,J.J., Lu,Q. and Qu,J. (2021) Super-enhancers: a new frontier for epigenetic modifiers in cancer chemoresistance. *J. Exp. Clin. Cancer Res.*, **40**, 174.
52. Huang,H., Zhu,Q., Jussila,A., Han,Y., Bintu,B., Kern,C., Conte,M., Zhang,Y., Bianco,S., Chiariello,A.M. *et al.* (2021) CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat. Genet.*, **53**, 1064–1074.
53. Tang,Z., Luo,O.J., Li,X., Zheng,M., Zhu,J.J., Szalaj,P., Trzaskoma,P., Magalska,A., Wlodarczyk,J., Ruszczycki,B. *et al.* (2015) CTCF-Mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell*, **163**, 1611–1627.
54. Kojic,A., Cuadrado,A., De Koninck,M., Giménez-Llorente,D., Rodríguez-Corsino,M., Gómez-López,G., Le Dily,F., Marti-Renom,M.A. and Losada,A. (2018) Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat. Struct. Mol. Biol.*, **25**, 496–504.
55. Cuadrado,A. and Losada,A. (2020) Specialized functions of cohesins STAG1 and STAG2 in 3D genome architecture. *Curr. Opin. Genet. Dev.*, **61**, 9–16.
56. Viny,A.D., Bowman,R.L., Liu,Y., Lavallée,V.P., Eisman,S.E., Xiao,W., Durham,B.H., Navitski,A., Park,J., Braunstein,S. *et al.* (2019) Cohesin members stag1 and stag2 display distinct roles in chromatin accessibility and topological control of HSC self-renewal and differentiation. *Cell Stem Cell*, **25**, 682–696.
57. Rogakou,E.P., Boon,C., Redon,C. and Bonner,W.M. (1999) Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J. Cell Biol.*, **146**, 905–916.
58. Ceccaldi,R., Rondinelli,B. and D'Andrea,A.D. (2016) Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.*, **26**, 52–64.
59. Fleming,A.M., Zhu,J., Ding,Y., Esders,S. and Burrows,C.J. (2019) Oxidative modification of guanine in a potential Z-DNA-forming sequence of a gene promoter impacts gene expression. *Chem. Res. Toxicol.*, **32**, 899–909.
60. Fleming,A.M. and Burrows,C.J. (2020) Interplay of guanine oxidation and G-quadruplex folding in gene promoters. *J. Am. Chem. Soc.*, **142**, 1115–1136.
61. Fleming,A.M., Zhu,J., Ding,Y. and Burrows,C.J. (2017) 8-Oxo-7,8-dihydroguanine in the context of a gene promoter G-quadruplex is an on-off switch for transcription. *ACS Chem. Biol.*, **12**, 2417–2426.
62. Ba,X., Bacsi,A., Luo,J., Aguilera-Aguirre,L., Zeng,X., Radak,Z., Brasier,A.R. and Boldogh,I. (2014) 8-oxoguanine DNA glycosylase-1 augments proinflammatory gene expression by facilitating the recruitment of site-specific transcription factors. *J. Immunol.*, **192**, 2384–2394.
63. Visnes,T., Cázares-Körner,A., Hao,W., Wallner,O., Masuyer,G., Loseva,O., Mortusewicz,O., Wiita,E., Sarno,A., Manoilov,A. *et al.* (2018) Small-molecule inhibitor of OGG1 suppresses proinflammatory gene expression and inflammation. *Science (80-.).*, **362**, 834–839.
64. Giorgio,M., Dellino,I.G., Gambino,V., Roda,N. and Pelicci,P.G. (2020) On the epigenetic role of guanosine oxidation. *Redox Biol.*, **29**, 101398.