UNIVERSITY OF NAPOLI "FEDERICO II"

Doctorate School in Molecular Medicine

Doctorate Program in
Genetics and Molecular Medicine
Coordinator: Prof. Lucio Nitsch
XXVIII Cycle

"The Intertwined Roles of DNA Damage and Transcription "

GIACOMO DI PALO

Napoli 2016

INDEX

1. INTRODUCTION

1.1 The Intertwined Role of DNA Damage and Transcription

Chromosomal DNA is the molecule containing genetic information that defines all the features of an organism. For this reason, it plays a central role within the cellular environment, and numerous processes depend on it.

Transcription, replication, DNA damage response, chromatin remodelling: DNA is continuously overwhelmed by a tremendous number of factors involved in these processes that must be perfectly coordinated in order to maintain the fidelity of genetic information.

These processes often overlap, spatially and temporally: chromatin remodelling enzymes are recruited together with transcription factor on gene promoter to allow gene transcription, or DNA repair proteins act during DNA replication if a mismatch occurs.

Transcription, in particular, is a vital process for development and survival of the whole organism, and therefore it must be perfectly accurate. The different executions of cell-type gene expression programs require profound protein-DNA and protein-protein transactions to recruit sequence-specific DNA-binding factors on gene promoters, in order to assemble the preinitiation complex (PIC), and many other interactions will occur before transcription starts (Lemon and Tjian 2000).

The complexity of the events leading to the execution of transcription is necessary and fundamental to ensure the correct accomplishment of gene expression program and the transcriptional fidelity. Proteins involved prior the transcription guarantee that the right gene is activated in a specific moment and context; factors involved during the transcription assure the correct chromatin architecture and that genetic information is transferred to RNA without mistakes.

Thus, because of the highly coordinated series of these events, transcription is particularly sensitive to any perturbation occurring in the cell environment. These perturbations affect the integrity of the genome, hence compromising also the correct execution of transcription.

Chromosomal DNA, in fact, is continuously threatened by both endogenous and exogenous sources of stress. Byproducts produced by cellular metabolism or radiations coming from external environment can in any moment interact with DNA and produce lesions that, if left unresolved, can be harmful for DNA-related processes like transcription and replication and, eventually, cell function and survival (Hoeijmakers 2001).

There is a wide spectrum of types of damage that affect DNA at high frequency, but cells have evolved several mechanisms to sense, recognize, mark and repair these DNA lesions, reducing to minimum their detrimental effects on cell functions (Lindahl and Barnes, 2000).

Mammalian cells possess four major DNA repair pathways to fight the numerous threats to genome: Base Excision Repair (BER), Nucleotide Excision Repair (NER), Mismatch Repair (MMR) and Recombinational repair (Figure 1) (Fong et al 2013).

Each of these pathway is specialized to correct a specific type of damage. For example, BER and NER pathways are involved in removing base adducts produced by reactive oxygen species (ROS) and ultra-violet light, respectively; MMR acts to repair erroneous misincorporation of bases, deletion and insertion occurring during DNA replication; recombinational repair removes DNA double-strand breaks (DSB) carrying out a recombination with a homologous DNA molecule (Iyama and Wilson 2013; Sancar 2004).

Factors involved in both transcription and repair pathways contemplate intimate transactions with DNA. Furthermore, transcription and repair often overlap, acting together on a specific gene region, and often sharing same features. Many repair factors, in fact, were demonstrated to show some characteristics of transcription activators, such as
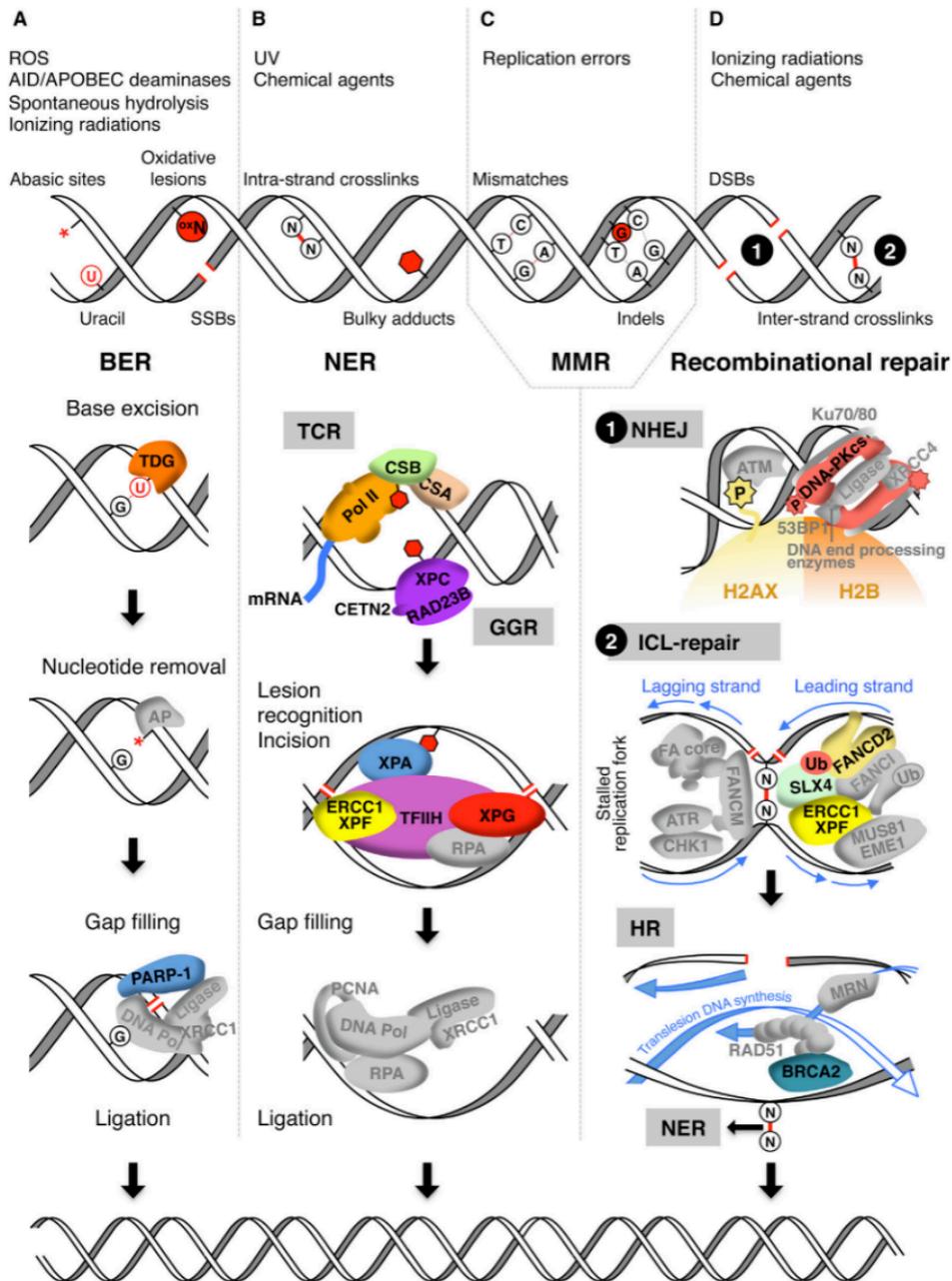
Figure 1. Exogenous and endogenous damaging events generate a variety of DNA damage, such as single- and double-strand breaks (SSB, DSB), insertions, deletions. These lesions are detected and repaired by four major DNA repair pathways: base excision repair (BER, A), nucleotide excision repair (NER, B), mismatch repair (MMR, C) and recombinational repair (D).

sequence-specific and damage-independent binding to DNA (Venema et al 1990; Bradsher et al 2002), or the ability to recruit transcriptional co-activators on gene promoters (Table 1) (Cortellino 2011). For example, Xeroderma pigmentosum complementation group C (XPC) protein, is a member of NER pathway involved in the recognition of base adducts deriving from ultraviolet irradiation or action of chemical agents, and it was shown to specifically bind hormone-inducible gene promoters sequences of human fibroblasts in a damage-independent manner. Strikingly, the depletion of XPC significantly attenuated the expression of hormone-inducible genes, thus suggesting a role of XPC in transcription activation besides the well known role in DNA repair (LeMay et al 2010). XPC is only one of the several DNA repair factors that play a role in transcription as well. As reported in Table 1, many other factors can facilitate transcription in different ways, e.g.: by actively remodelling chromatin or by recruiting specific remodelling ennzymes, or by changing the topological conformation of DNA double-helix, or again directly functioning as coactivator or repressor of gene expression, or by stabilizing transcription activators (Fong et al 2013).

Indeed, in recent years a growing list of evidences has been revealing a new and unexpected tight connection between transcription and DNA repair. Transcription is a potential source of DNA damage leading to mutagenic events that can compromise cell functions. For example, deaminases or cellular byproducts can damage the non-template DNA single-stranded forming upon the advancing transcription fork and, moreover, it is prone to recombination potentially leading to mutagenic events (Rahmouni and Wells 1992); for this reason, transcription is constantly monitored by DNA repair factors in order to assure that DNA strands, both template and non-template one, remains undamaged after gene has been transcribed and chromatin has been modified. Moreover, as previously said, proteins originally characterized exclusively as repair factors, were observed to be involved in transcription as well. Furthermore, in some cases transcription is required to initiate a repair process, such as

transcription-coupled repair (TCR), a subpathway of NER, where DNA damage is not recognized until RNA Polymerase stalls in correspondence of the DNA lesion (Hanawalt and Spivak 2008; Mellon et al 1987).

Protein	Pathway	Function in DNA Repair	Function in Transcription
DNA-PKcs	RR, BER	Facilitates DNA end processing and resealing Stimulates BER for oxidative damage repair	Promotes gene activation by chromatin remodeling Modulates activity of TF
FANCD2	RR	Inititates RR	
TFIIH	NER	Unwinds DNA at damaged sites	Unwinds DNA at gene promoters Phosphorylates CTD of RNA Pol II
PARP-1	BER, NER	Interacts with components of RR, BER and NER	Functions as activator/coactivator or repressor Modulates chromatin structure
XPC	NER, RR	Stimulates components of BER	Activates transcription of NR target genes Functions as coactivator in ESCs
XPG	NER	Incises damaged strand 3' to the DNA lesion	Stabilizes TFIIH Promotes chromatin remodeling at NR target genes Facilitates DNA looping at NR target genes

Table 1. Table reports some DNA repair factors involved both in repair and transcription. Definitions are as follows: RR, recombinational repair; BER, base excision repair; TF, transcription factor; NER, nucleotide excision repair; CTD, carboxy terminal domain; ESC, embryonic stem cell.

But there is another aspect that correlates more strictly these two processes. In fact, not only transcription can induce damage and repair protein can be involved in transcription, but DNA damage itself can be required for transcription.

Although global damage on DNA is generally associated with gene silencing, there are well documented cases in which some types of DNA damage, and their processing by repair factors, are essential events in order to initiate and support gene transcription (Ju et al 2006;

Lin et al 2009). For example, localized damage of a specific nucleotide and DSB formation can be implicated in transcription activation of select genes. A common obstacle to the processivity of transcription fork and the accuracy of transcription process is represented by DNA supercoiling accumulating behind and ahead of elongating RNA polymerase. DNA breaks are thought to eliminate these supercoiling, thus relaxing DNA strands and inducing a permissive chromatin architecture necessary for transcription activation. Moreover, relaxation of DNA strands via DNA breaks could also allow chromosome bending that facilitates spatial interaction between factors bound to enhancer and to promoter (Perillo et al 2008).

More detailed examples will be described below, but definitely it is becoming increasingly clearer that DNA damage/repair and transcription are two processes much more intertwined than one could expect few years ago. This innovative point of view provides a new approach to the study of these two important processes in order to better understand the mechanisms regulating them.

1.2   Harmful and Beneficial Effects of R Loops

R loops represent a typical example where DNA damage and transcription converge, because of its intrinsic capability to lead to genome instability on one side, and because of recent evidence reporting them as regulators of transcription on the other side.
R loop is a transcription-induced three stranded structure composed by the nascent transcribed RNA hybridized with its template DNA strand, and the displaced non-template strand (Figure 2) (Reaban et al 1994). They naturally occur when RNA Polymerase transcribes a C-rich DNA template, generating a G-rich transcript. R loops are particularly stable, as RNA/DNA hybrid is thermodynamically more stable than DNA/DNA duplex, and their persistence can lead to DNA damage in different ways. The displaced single-strand, in fact, being no longer

protected in the double-helix structure, is highly susceptible to chemical modifications; for example, cytosine residues can spontaneously deaminate to uracil or can be deaminate by specific activation-induced cytidine deaminase (AID) enzymes; if U:G mismatch is then replicated, a C:G → T:A transition mutation is fixed in the genome, with potential deleterious effects (Skourti-Stathaki and Proudfoot 2014). However, uracil residues can also be repaired by BER components, but this process can create DNA nicks or abasic sites possibly resulting in further mutations.
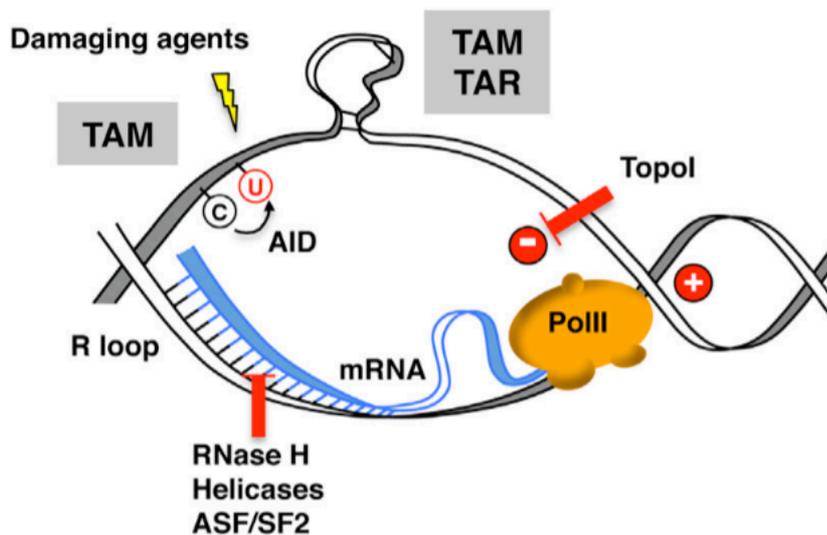
Figure 2. R loop forms when nascent mRNA hybridizes back to its template. Negative (-) and positive (+) supercoiling accumulate behind and ahead RNA Pol II, stabilizing R loop. The displaced single-stranded DNA is exposed to damaging agents, AID-induced deamination, formation of secondary structures that are pronte to transcription-associated mutagenesis (TAM) and recombination (TAR)

Moreover, since exposed non-template strand of R loop is a G-rich DNA, G-quadruplex structures can form, that represent the substrate for nuclease enzymes, again potentially generating DNA breaks.

Another interesting way for stable R loop to generate damage is its capability to interfere with DNA replication; if a RNA Polymerase is blocked in correspondence of a R loop, a collision can occur between transcription and replication fork, thus generating transcription-

induced recombination or even DNA breaks. Regardless of how DNA breaks induced by R loop are generated, they can be converted to DSB and ultimately resulting in DNA translocation and genome instability (Aguilera and Garcia-Muse 2012).

R loop is an evolutionarily conserved mechanism, as it was observed in bacteria and mammalian cells, and therefore organisms developed different mechanism to prevent their formation, or remove them once formed. For example, endonucleases topoisomerase act to relax negative DNA supercoiling, allowing the non-template strand to anneal more easily with its homologous strand, thus preventing R loop formation; RNase H enzymes, instead, can act upon R loop formation to cleave the RNA of RNA/DNA hybrids and restore the double-stranded DNA structure.

As anticipated, R loop have recently been shown to have a role in regulation of gene expression. In human protein-coding genes, they preferentially form on CpG promoters with positive GC skew (where template strand has an excess of C vs. G residues) and on G-rich termination regions (Figure 3) (Ginno et al 2012). The persistence of R loop seems to suggest that these structures can play a role in recruiting histone methyltransferases or DNA demethylases in order to provide the epigenetic landscape required for activation of genes.
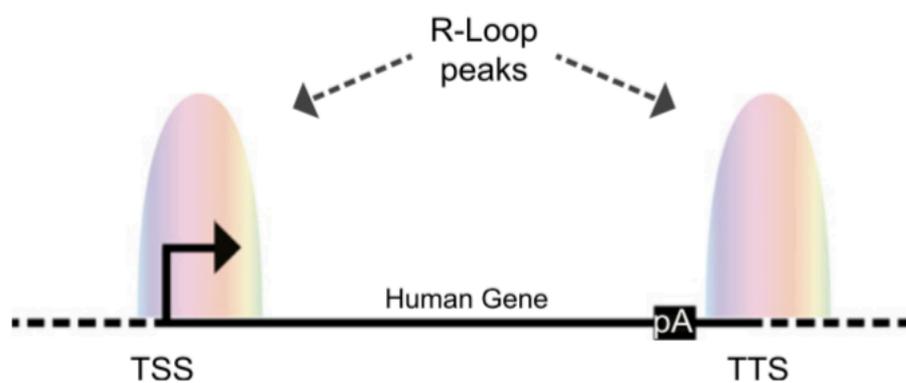
Figure 3. R loops are enriched at both gene ends. In human protein-coding genes, R loops form over unmethylated CpG island promoters with positive GC skew and G-rich termination regions. Promoter-enriched R loops could activate gene expression, whereas terminator-enriched R loops promote transcrip- tional termination by facilitating Pol II pausing downstream from the poly(A) signal.

On the opposite side of the gene, accumulation of R loop on G-rich transcription termination site can facilitate RNA Polymerase termination downstream from the poly(A) signal; in human genome, many terminator sites of protein-coding genes have a positive GC skew that, as said, is where R loop preferentially forms. Here, R loop can block the elongation of RNA Polymerase providing a stoichiometric obstacle for its processivity (Skourti-Stathaki and Proudfoot 2014).

Taken together, these information indicate R loops to be transcription-induced structures that could potentially induce DNA damage, but also be involved in transcription regulation.

1.3 "Scheduled" DNA Damage Triggers Transcription Activation

Cells developed several and precise mechanisms in order to prevent and repair several types of damage that can affect DNA molecule.

But in recent years it is becoming evident that cells evolved a "scheduled" and localized DNA damage that is required to facilitate transcription, adding new evidence to support the interconnection between these two only apparently distant events.

DNA breaks are an example of this "scheduled" damage that cell uses to relax DNA strands thus facilitating transcription.

For example, Madabhushi et al. reported that a subset of mouse genes, the early-response genes, whose expression was normally regulated by external stimuli, was activated upon double-strand break (DSB) induction even in absence of any stimulus. DSBs randomly generated in mouse genome by using etoposide (an inhibitor of topoisomerase II that binds the enzyme to DNA leading to a potentially toxic DSB) or site-specific generated by using CRISPR-Cas9 system to localize DSB in the

promoters of these genes, were alone sufficient to trigger activation of transcription that, in physiological condition, would require an external stimulus. Formation of DSB required for activation of early-response genes, was mediated by topoisomerase IIβ (TopoIIβ), that is recruited to relieve torsional stress during transcription by transiently breaking the two strands of DNA, thus forming a DSB. Indeed, knockdown of TopoIIβ affected transcription of early-response genes even upon stimulus induction. However, the positive role of DSB on transcription activation was so strong that targeted DSB could induce gene expression even in absence of TopoIIβ activity (Madabhushi et al 2015). These study proposes a model where, upon neuronal activity stimulation, promoters of early response genes undergo TopoIIβ-mediated DSB to rapidly resolve topological constraints to facilitate transcription (Figure 4).
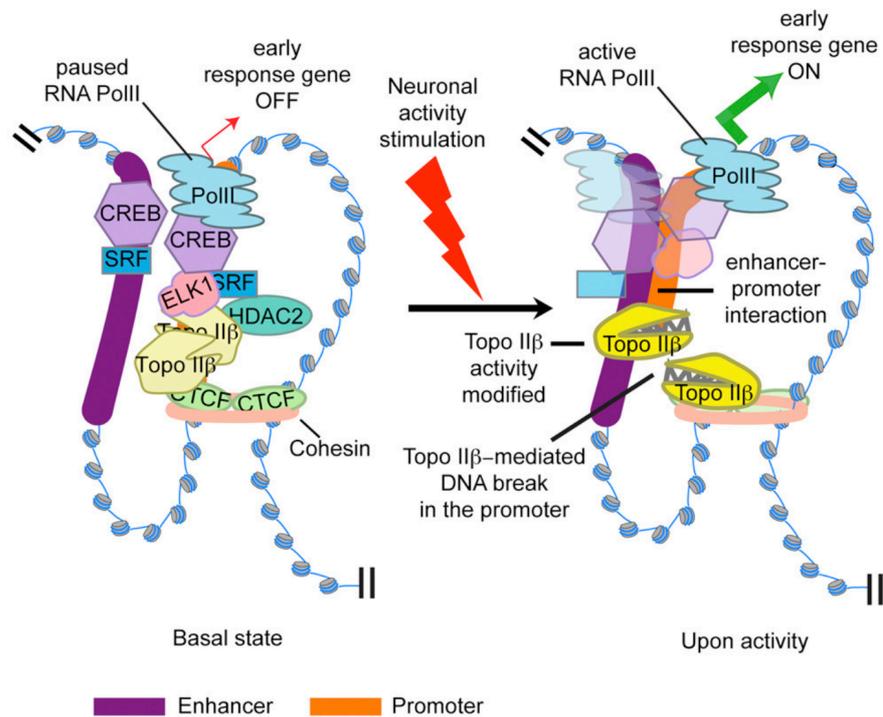
Figure 4. Upon neuronal stimulation, TopoIIβ-operated DSB on promoter of mouse early-response genes eliminate topological constraints to facilitate RNA PolII elongation and transcription activation.

Additional evidence that sustain the link between transcription and DNA damage (and DNA breaks in particular), come from the study recently published by Puc et al. that showed the requirement of topoisomerase 1 (TOP1) nicking function in activation of enhancers transcription units (eRNAs), a class of non-coding RNA that may have a role in transcriptional regulation. Their data demonstrate that, upon androgen stimulation, androgen receptor (AR) and TOP1 are rapidly recruited on a large cohort of AR-regulated enhancer. TOP1 acts relaxing DNA supercoils by operating transient single-strand breaks (SSB) for the passage of individual strands through one another, and by the following rejoining of the backbone of DNA. Upon TOP1 nicking activity, a robust eRNA synthesis is observed.
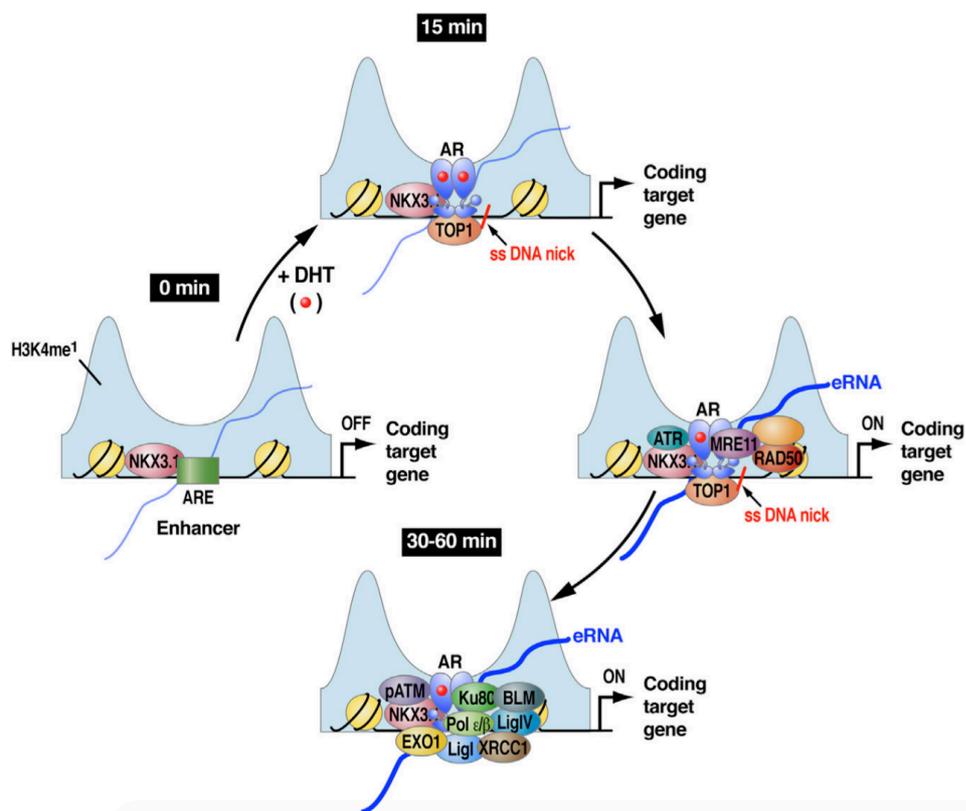
Figure 5. Following androgen stimulation, AR and TOP1 are recruited to enhancer region, where enzymatic activity of TOP1 is stimulated, leading to a nick of DNA on a single strand. The generation of break leads to eRNA synthesis and recruitment of ATR and MRN repair components.

However, SSBs operated by TOP1 must be kept under surveillance because they can turn into DSB. For this reason, TOP1 acivity is accompanied by the recruitment of DNA damage response factors, such as ATR and MRN complex, that are well known to be implicated in the repair of DNA breaks (Figure 5) (Puc et al 2015).

These data clearly confirm that DNA breaks, and perhaps DNA damage in general, can no longer be considered as a harmful event for cell survival, but rather a necessary step for gene transcription.

1.4 Oxidative DNA Damage

A common type of DNA lesion is represented by oxidative DNA damage. Oxidation is a common insult to DNA consisting in the presence of oxidized bases in the nucleotide chain. It can be generated both by endogenous and exogenous sources deriving from cellular metabolism and from external environment, respectively (Klaunig et al 2010).
All four genomic bases can chemically interact with reactive oxygen and radical species, and when this occurs, the nucleotide is oxidized and damage is generated.
More than one hundred oxidative DNA adducts (purine, pyrimidine, and the deoxyribose backbone) have been identified (Figure 6) (JE Klaunig 2010), and among these, 8-oxo-7,8-dihydro-2′-deoxyguanine (8oxodG) is the most studied (Figure 7) (Cooke et al 2003; Scott et al 2014).

8oxodG is a critical biomarker of oxidative stress, and elevated genome levels have been associated with tumour; an inefficient repair of 8oxodG can lead to GC → TA transversions, potentially forming mutations even in oncosuppressor genes (Cooke et al 2003).
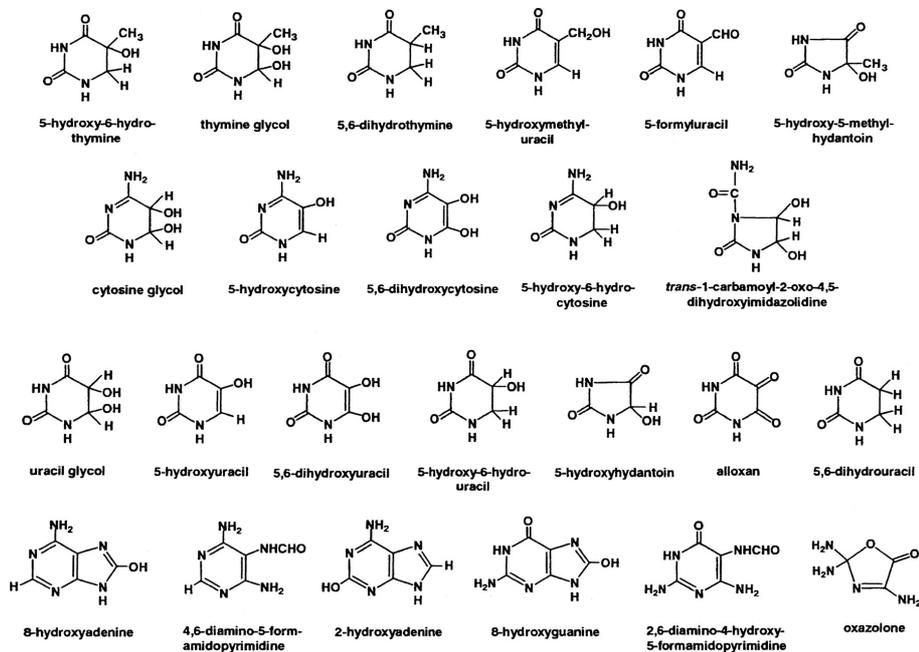
Figure 6. DNA base products of interaction with reactive oxygen and free radical species.

Base Excision Repair (BER) is the pathway implicated in the repair of oxidative damage, and it was associated to transcription activation as it involves the formation of DNA breaks that are thought to relax DNA and favour RNA Polymerase elongation (Fong et al 2013).
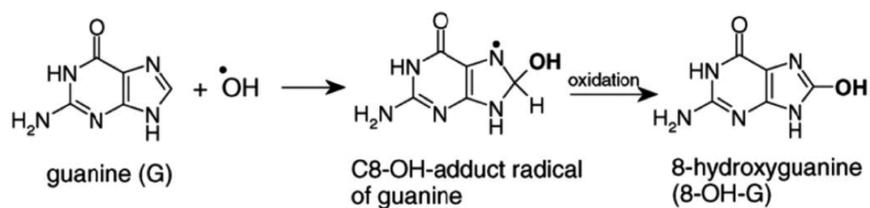
Figure 7. Generation of 8oxodG induced by interaction of hydroxyl radical.

1.4.1 Source of Oxidative DNA Damage

Cell metabolism is the main source of endogenous oxidative DNA damage. Production of energy by oxidative phosphorylation occurring in mitochondria and chemical reactions catalyzed by various oxidase enzymes, produce reactive oxygen species (ROS) - such as superoxide anion radical (O_2^-), peroxynitrite ($ONOO^-$), hydrogen peroxide (H_2O_2) – that, if not disposed, accumulate in the cell creating an oxidative environment (Cooke et al 2013).

During mitochondrial oxidative metabolism, most of oxygen molecules are reduced to water, while about 5% is converted to superoxide anion ($O2^-$); superoxide dismutase (SOD) reduces $O2^-$ to hydrogen peroxide ($H2O2$) that is one of the major contributor to cellular oxidative damage; H_2O_2 is later converted to water, but this conversion is not 100% efficient, and residual $H2O2$ molecules persist in the cell forming a source of damage to DNA (Klaunig et al 2010).

Although it is not clear how ROS generated in mitochondria can travel into the nucleus, it is known that hydrogen peroxide is a highly diffusible molecule, hence more prone to shuttle in the nucleus and be involved in the formation of oxidized bases through Haber-Weiss reactions, which generates hydroxyl radical (-OH) from interaction between O_2- and H_2O_2 (Scott et al 2014).

However, regardless the ability of ROS to shuttle from mitochondria, nucleus as well can produce ROS through oxidation reactions catalysed by oxidase enzymes, such as flavin adenine dinucleotide (FAD)-dependent oxidative reactions, in which the cofactor FAD is reduced to $FADH_2$ and then reoxidized to FAD by oxygen with the generation of H_2O_2 (Amente et al 2010).

These reactions often occur nearby the DNA, like the ones catalysed by chromatin remodelling enzymes, thus creating an oxidative environment that can lead to the formation of oxidized bases with consequent oxidative DNA damage.

1.4.2 8oxodG in Genome Instability

As previously cited, the most frequent product of oxidative damage to DNA bases is 8oxodG, that is formed when guanosine of nucleotide chain reacts with ROS, mainly H_2O_2, and is oxidized to 8oxodG.

Elevated levels of 8oxodG have been associated to cancer, aging, cardiovascular disease, and therefore it is the most studied DNA adduct deriving from oxidative damage (Cooke et al 2013).

In its stable *syn* conformation, 8oxodG can pair with both cytosine and adenine, and if the A:G mismatch is not repaired, a G:C → T:A transversion will occur, thus creating a mutation. This mutagenicity can even support initiation, promotion and progression of carcinogenesis; indeed, GC → TA tranversion potentially derived from 8oxodG have been observed in vivo in the *ras* oncogene and in the *p53* tumor suppressor gene in lung and liver cancer (Ohnishi et al 2011; Hollstein et al 1995).

Furthermore, 8oxodG can even induce chromosome rearrangements through its repair by BER proteins, whose action (discussed in detail later) contemplates the production of DNA strand breaks to allow the substitution of 8oxodG with correct G.

If not immediately repaired, these breaks can lead to the rejoining of the broken ends to produce new chromosomal arrangement of genes, thus causing genome instability.

DNA breaks mediated by BER pathway represent the link between two opposite aspects of 8oxodG-marked oxidative DNA damage: on one hand, they explain the mutagenic potential of this damage that could result in mutations and genome instability as just said; on the other hand, they are required to facilitate transcription of selected genes through relaxation of DNA helix.

1.4.3 Repair of Oxidative DNA Damage

In mammals, removal of 8oxodG residues and its substitution with the correct G residues is accomplished by the Base Excision Repair pathway. However, BER is involved not only in repair of 8oxodG-marked DNA oxidation, but also in the correction of base adducts arising from alkylation, demamination, depurination and depyrimidination. The basic components of BER are DNA glycosylases, AP-endonucleases, DNA polymerases, and DNA ligases.

It acts via two possible pathways, short-patch and long-patch. The short-patch pathway repairs a tract of a single nucleotide, while the long-patch produces a flap composed by at least two nucleotides starting from the one containing the base adducts (Robertson et al 2009). Currently, it is not clear how BER decides to operate the short-patch rather than the long-patch; in this discussion we will focus on the short-patch pathway, that is the one involved in repairing of 8oxodG lesions.

In the repair of 8oxodG-marked damage the first step of BER pathway is the recognition of base adduct by 8-Oxoguanine glycosylase (OGG1). This glycosylase catalyzes the cleavage of an N-glycosidic bond, thus removing the damaged base; this reaction, leaves an apurinic site (AP site) in correspondence of the removed oxidized guanine. Next, the DNA backbone is cleaved by Apurinic/apyrimidinic endonuclease 1 (APE1), that creates a single-stranded DNA nick 5′ to the AP site. Importantly, this break contains a 3′-hydroxyl and a 5′-phosphate, that is the substrate compatible for the subsequent action of DNA polymerase β (POLB). POLB fills in the gap adding the correct nucleotide. Finally, the DNA nick with 3′-OH and 5′-P ends previously generated by APE1, is ligated by a DNA ligase to restore the integrity of the helix (Figure 8) (Robertson et al 2009).

Although 8oxodG has a relatively low mutagenic potential (Shikazono et al 2006), BER is a very efficient pathway, as it guarantees the removal of almost all base adducts that form in the
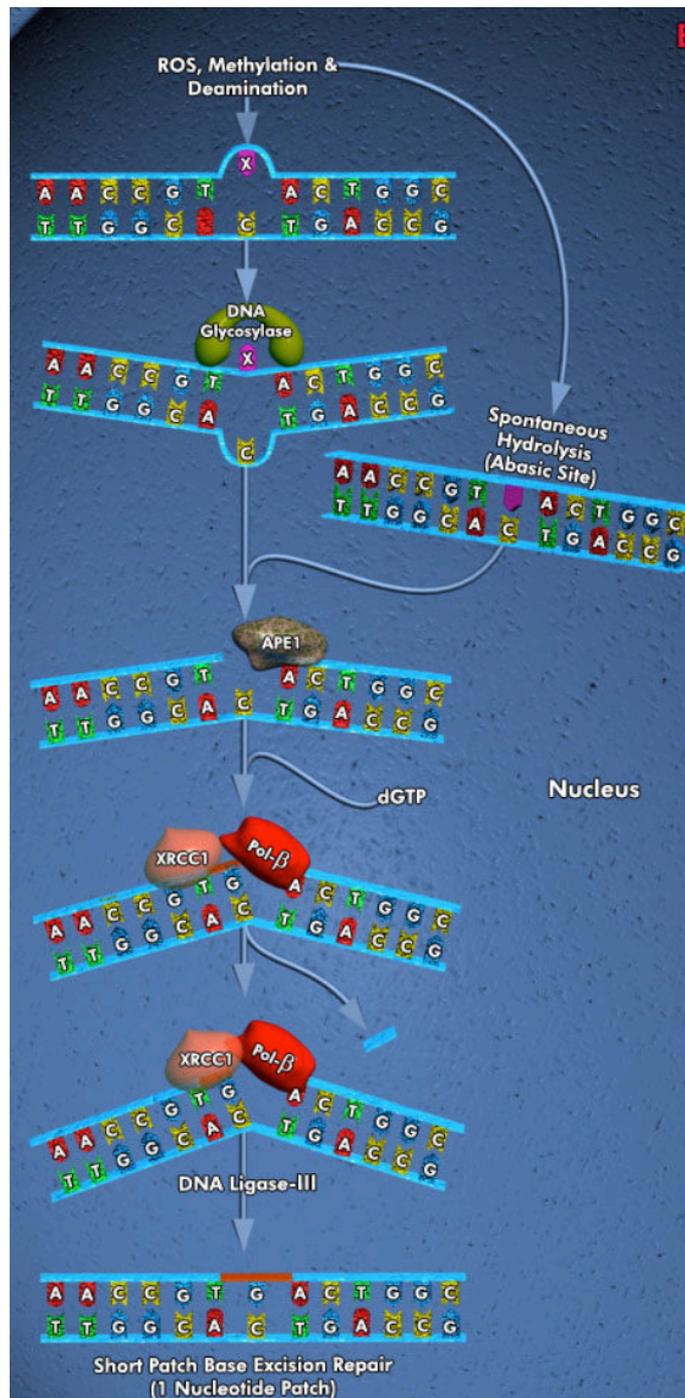
Figure 8. Schematic representation of BER pathway. DNA glycosilase catalyzes the excision of the damaged base, creating an abasic (AP) site. Endonuclease APE1, which can act also upon a spontaneous hydrolysis of a DNA base, catalyzes the incision of the DNA backbone 5' to the AP site. PolB displaces the AP site and polymerizes DNA to fill in the gap. Finally, DNA Ligase III catalyzes the formation of a phosphodiester bond, completing the repair pathway.

genome (Cooke et al, 2003). Paradoxically, DNA breaks operated by APE1 endonuclease are even more toxic than the base adduct that is removed, as they can lead to unscheduled recombination events and chromosome rearrangements.

Conversely, as it will be discussed in more detail below, it is proposed that this transient breaks are required for chromatin relaxation in order to facilitate transcription. Indeed, DNA single-strand breaks can function as entry point for Topoisomerases, the enzymes known to catalyse the transient breaking and rejoining of two strands of duplex of DNA, which allows the strands to pass through one another, thus relieving torsional stress that occurs during transcription.

1.5 Role of Oxidative DNA Damage in Transcription

Generation of 8oxodG seems to be a crucial step for transcription activation. As previously explained, repair of 8oxodG contemplates the formation of DNA breaks by BER pathway, that can be used by topoisomerases to eliminate supercoiling generated by the progression of transcription fork. This allows the DNA double-helix to be relaxed, thus facilitating elongation of RNA polymerase throughout the gene body.

A clear evidence of the correlation between oxidative damage and transcriptional activation comes from Zarakowska et al, who carried out a chromatin fractionation experiment basing on differential solubility of H1-containing and H1-free nucleosomes; notably, they found level of 8oxodG in transcriptionally active euchromatin that was approximately 5-times higher compared to transcriptionally silenced heterochromatin (Zarakowska et al 2013).

Another important example of correlation between DNA oxidation and transcription is represented by lysine speficic

demethylase (LSD1)-mediated transcription, as DNA oxidation generated by LSD1 was reported to drive transcriptional activation of estrogen-induced genes and Myc-responsive genes (Perillo et al 2008; Amente et al 2010).
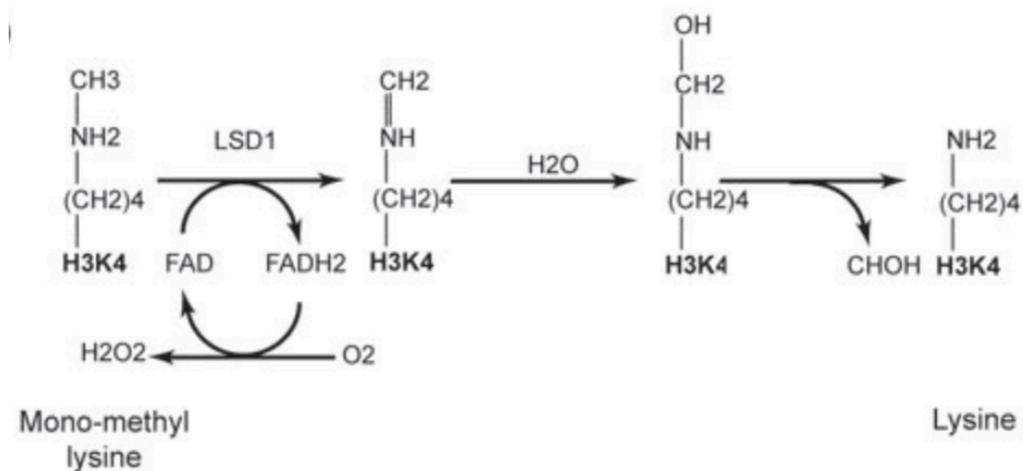
Figure 9. LSD1 demethylates H3K4me2/me1 via an amine oxidation reaction using FAD as a cofactor and producing H_2O_2. The imine intermediate is hydrolyzed to an unstable carbinolamine that spontaneously degrades to release formaldehyde.

LSD1 is a FAD-containing enzyme which demethylases both mono- (H3K4me1) and di-methylated (H3K4me2) H3K4me via an amine oxidation reaction that uses FAD as cofactor producing H_2O_2 (Figure 9). It has been demonstrated that LSD1 is recruited by DNA-bound estrogen receptor on hormone-responsive genes, and H_2O_2 produced by localized demethylation reaction converted nearby guanine residues into 8oxodG. 8oxodGs generation was rapidly followed by rectruitment of OGG1 to trigger BER pathway, and by endonuclease topoisomerase IIβ (TopoIIβ), that alters topological states of DNA during transcription. Interestingly, recruitment of OGG1 and TopoIIβ was dependent on action of LSD1, thus suggesting a model contemplating that removal of the oxidized bases generates transient nicks that function as entry points for topoisomerase. In this way, this enzyme can relax DNA strands and favor chromatin bending to accommodate the transcription

initiation complex (Perillo et al 2008). Moreover, the DSBs resulting from the endonuclease activity of TopoIIβ are sensed by repair enzyme like PARP-1 and DNA-PK to keep under control this damage until it is required for transcription (Figure 10).
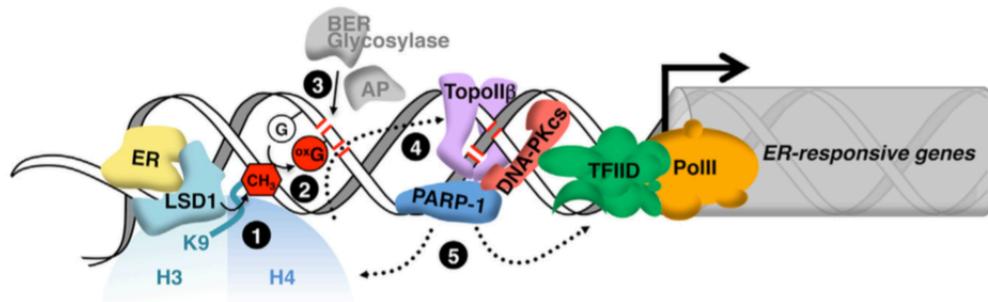
Figure 10. Upon ligand binding, estrogen receptor (ER) activates LSD1 at responsive genes (1). The demethylation reactions releases H_2O_2 that converts nearby guanines G into 8oxodG (^{ox}G) (2). ^{ox}G removal by BER creates DNA nicks (3) that facilitate the entrance of TopoIIβ (4). TopoII recruits repair enzymes, e.g. PARP-1 and DNA-PKcs, which induce a permissive chromatin architecture for transcription initiation (5).

Similarly, LSD1-mediated local DNA oxidation was proposed as driving force for Myc-responsive genes as well (Amente et al., 2010).

Myc is one of the most common activators of cell proliferation used by cancer cells to drive disease progression, and LSD1-BER-coupled epigenetic regulation via demethylation of H3K4me2 by LSD1 at the promoters of Myc-target genes drives their expression. Indeed, activation of Myc triggers a cascade of events on its target genes that starts with the recruitment of LSD1 that demethylases H3K4me2 epigenetic mark. Since LSD1-mediated demethylation produces hydrogen peroxide, an accumulation of 8oxodG on gene promoter is observed; subsequent processing of oxidized bases by OGG1 and Ape1 seems to be crucial for transcription, as their silencing drastically affected the expression of Myc-responsive genes.

Figure 11 shows a representation of the different steps suggested for Myc mediated transcription activation, where H3K4me/K4me2 is a

signature for Myc activated targets with preloaded RNA Polymerase; Myc-LSD1 complex is recruited on Myc-target gene promoters, and 8oxodGs are generated and repaired by BER factors; recruitment of transcription elongation factor P-TEFb leads to the phosphorylation and activation of RNA Pol II, thus triggering gene transcription (Amente et al 2010).

Although it is not yet clear how cells can distinguish between these "scheduled" DNA damage events that are linked to gene activation from those that arise spontaneously in the genome causing undesirable consequences, localized DNA damage is the price that cells can afford to pay in order to accomplish a vital process that is transcription.

Thus, the hypothesis according to which 8oxodG-marked oxidative damage could have a role in gene transcription and not only be considered a type of DNA lesion represents a relatively new concept, and data that will be discussed below together with literature reports seem to support it, making 8oxodG a potential new "epigenetic marker" that regulates gene transcription.
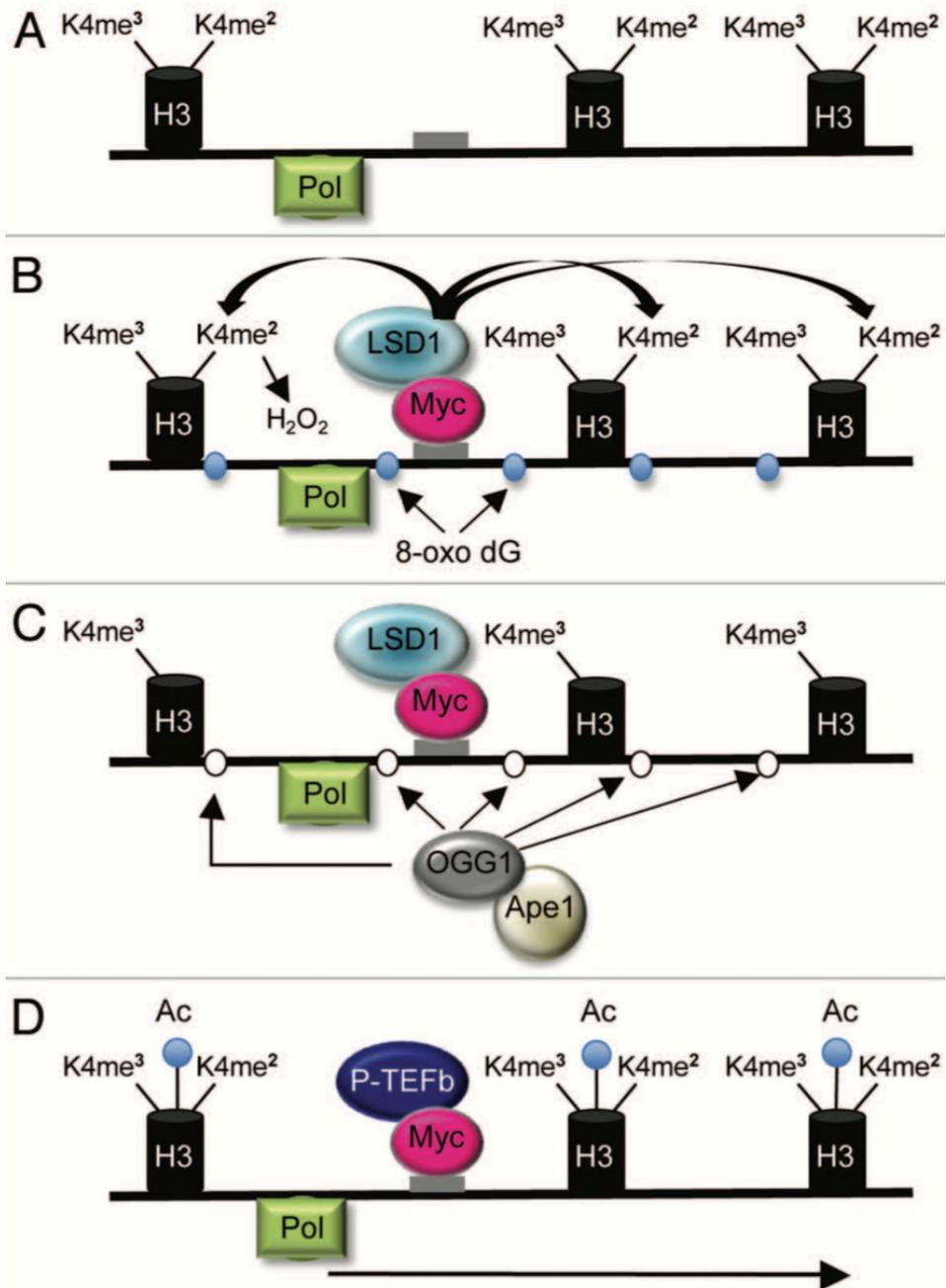
Figure 11. Schematic representation of Myc transcription activation model. (A)H3K4me/K4me2 is a signaturefor Myc activated targets with preloaded RNAPolII. (B) Myc binding induces chromatin relaxing: Myc-LsD1 complex is recruited on the e-box, transient de-methylation of H3K4me2. (C) BER requirement for repairing 8oxodG. (D) Histone acetylation and P-TEFb recruitment allow efficient transcription of Myc target genes.

2. AIMS OF THE STUDY

DNA damage and transcription are no longer considered as two distinct processes, since it is becoming more clear that they are much more intertwined than expected. DNA repair proteins act as transcription factors, and transcription can cause DNA damage. Moreover, transcription activation can require "scheduled" DNA damage, thus providing to DNA lesions not only a harmful role, but also effects that are indispensable for an important cell function that is transcription.

DNA oxidation is an example of damage correlated to transcription. Generation of oxidized guanine residues (8oxodG), triggers the activation of the BER pathway that contemplates the formation of DNA breaks; these breaks are thought to unwind chromatin structure to facilitate transcription.

The aim of this work is focused to provide an accurate and specific map of oxidative DNA damage distribution along human and mouse genome, to correlate discrete oxidized sites to double-strand break-marked regions and to the event of transcription. We set up OxiDIP (DNA ImmunoPrecipitation)-Seq technique for genome-wide distribution of 8oxodG, and ChIP-Seq experiments for genomic mapping of γH2AX and NBS1, two markers of DNA double-strand breaks.

We used specific computational tools that allowed us to elaborate the large amount of data generated by next-generation sequencing approach. The knowledge aquired should provide further evidences to sustain the correlation between DNA damage and transcription.

3. 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, 100 ng/ml cholera enterotoxin, and 20 ng/ml epidermal growth factor, and incubated at 37°C in humidified atmosphere with 5% CO_2. For UV treatment, exponentially growing MCF10A cells were exposed to $40J/m^2$ UV light (254 nm). Medium was refreshed after irradiation and the cells incubated for 30 minutes. Then, cells were washed twice with ice-cold phosphate buffered saline (PBS) and collected 30 minutes after irradiation. For NAC treatment, 1 mM *N*-acetyl cysteine (A7250, SIGMA-ALDRICH) was added to the medium for 2 hours before being collected as previously described (Amente et al 2010).

MEF cells were grown in DMEM supplemented with 10% north-american fetal bovine serum (FBS), 2 mM glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin, at 37°C in 5% CO2.

OxyDIP (DNA ImmunoPrecipitation) Protocol

Genomic DNA from cultured cells was extracted by using DNeasy Blood&Tissue kit (Cat. no. 69504, QIAGEN). 10μg of genomic DNA per immunoprecipitation were sonicated in 100μl TE buffer (100 mM Tris-HCl pH 8.0, 0.5 M EDTA pH 8.0) to produce random fragments ranging in size between 200 and 800bp using Bioraptor (Bioruptor Plus UCD-300). 4μg of fragmented DNA in 500 μl TE Buffer were denatured for 5 min at 95 °C and immunoprecipitated over night at 4°C with 4μl of polyclonal antibody against 8-Hydroxydeoxyguanosine (AB5830 Millipore) in a final volume of 500μl IP buffer (10 mM NaPi, 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) under constant rotation. The immunoprecipitation complex was incubated with 50μl Dynabeads Protein G (Cat. No. 10003D, ThermoFisherSCIENTIFIC), previously saturated with 0,5% Bovine Serum Albumine diluted in PBS for 3 hours at 4 °C, rotating, and

washed three times with 1ml Wash buffer (10 mM NaPi ph 7.4, 0.15 M NaCl, 0.05% Triton X-100. The complex beads-antibody-DNA was then disrupted adding 200µl Lysis buffer (50 mM Tris-HCl ph 8, 10 mM EDTA ph 8, 1% SDS, 0.5 mg/mL Proteinase K for 4 hrs at 37 °C, and again with 100µl Lysis buffer for additional 1 hr increasing the temperature to 52°C. The recovered oxidized DNA was purified by using MinElute PCR Purification kit (Cat. No. 28004, QIAGEN) in a final volume of 72µl EB buffer (provided by the kit). To avoid possible interference of light-sensitivity of 8oxodG residues, all the steps of OxyDIP protocol, until the washes of the immunocomplex, were carried out in low-light conditions. Furthermore, 50 µM *N-tert*-Butyl-α-phenylnitrone (B7263, Sigma) was added to Dneasy Blood&Tissue, IP and wash buffers, to preserve the oxidized state of DNA.

OxiDip-Sequencing and quantitative 8oxodG immunoprecipitation assay

DIP-seq libraries were prepared from 10ng of DIP (or Input) DNA with TruSeq ChIP Sample Prep Kit (Illumina) according to the manufacturer's instructions. Prior to sequencing, libraries were quantified using Quibit (Invitrogen) and quality-controlled using Agilent's Bioanalyzer. 50bp single-end sequencing was performed using Illumina HiSeq 2000 platform (Genomix4life S.R.L., Baronissi, Salerno, Italy) according to standard operating procedures. Alignments were performed with Bowtie to hg18 reference genome using default parameters. SAMtools (ref) and BEDtools (ref) were used for filtering steps and file formats conversion. The peaks were identified from uniquely mapped reads without duplicates using MACS and the p-value cutoff used for peak detection was 1e-5. The normalized reads counts, representing DNA oxidation values, were computed using DNA Input as control. UCSC genome browser was used for data visualization.

For qPCR analysis, 3µl of 8oxodG immunoprecipitated DNA (antibody AB5830, Millipore) was analyzed in duplicate by quantitative PCR,

using SYBR Green 2X PCR Master Mix (Applied Biosystems). The following primer sets were used:

Positive region (genomic position chr2: 233294905 - 233294981)

 FW 5′-CCAACATCTTAAATTTGTCAACTCTC;

REV 5′-TGCTGGCAGAAGTGTGATTT.

Negative region (genomic position chr2: 232053796-232053862) FW 5′-AAGCTGGAGGCAGAGTGG;   REV   5′-TCTGACAACCCTGTTCACTACC.

γH2AX and NBS1 ChIP-sequencing

Chromatin extracts of MCF10A cells were performed as described (Ambrosio S et al 2015).

Rabbit polyclonal against H2A.X (phospho S13) antibody (ab11174) and rabbit polyclonal against p95 NBS1 (phospho S343) antibody (ab47272) were used.

ChIP-seq libraries were prepared from 10 ng of ChIP (or Input) DNA with TruSeq ChIP Sample Prep Kit (Illumina) according to the manufacturer's instructions. 50bp single-end sequencing was performed using Illumina HiSeq 2000 platform. Reads were quality checked and filtered with ngsqctoolkit. Alignments were performed with Bowtie to hg18 using default parameters. SAMtools and BEDtools were used for filtering steps and file formats conversion. The peaks were identified from uniquely mapped reads without duplicates using SICER and the FDR cutoff used for peak detection was 0,01. The normalized reads counts, representing the γH2AX positioning scores, was calculated using DNA Input as control. UCSC genome browser was used for data visualization.

RNA-Seq analysis

Fastq data MCF10A RNA-Seq study were retrieved from GSM1100206 (GEO Datasets, NCBI). In particular, we selected samples from MCF10A cells where RNA samples are sequenced in duplicate at 100M

reads. RNA-Seq was analyzed with RAP pipeline (D'Antonio et al 2015): briefly, RNA-seq reads were quality checked and filtered with ngsqctoolkit; sequenced reads were mapped against hg18 reference genome using TopHat. Transcript assembly and abundance estimation were performed with Cufflinks. Transcripts relative abundance was measured in FPKM.

TSS selection

TSS selection was performed as recently described (Scala 2014). Briefly, genomic coordinates of human TSSs were downloaded from the UCSC track ''switchDbTss''. We retained TSSs whose confidence score was greater than or equal to 20; N_{TSS}=527,487 (21% of the total).

Profile of 8-xodG around TSSs

Let *OXO(i,j)* be the set of DNA oxidation values, associated with sites placed inside *Bin(i,j)*, and let *|OXO(i,j)|* be its cardinality. If we denote with *score(i,j)* the sum of all DNA oxidation values from *OXO(i,j)*, we can compute the "average Oxidation score" OXO(j) for the j-th Bin as:

$$OXO(j) = \frac{\sum_{i=1}^{N_{TSS}} score(i,j))}{\sum_{i=1}^{N_{TSS}} |OXO(i,j))|}.$$

Profile of gH2AX around TSSs

Let *gH2(i,j)* be the set of yH2AX positioning scores, associated with sites placed inside *Bin(i,j)*, and let *|gH2(i,j)|* be its cardinality. If we denote with *score(i,j)* the sum of all yH2AX scores from *gH2(i,j)*, we can compute the "average yH2AX Positioning score" (gH2(j)) for the j-th Bin as:

$$gH2(j) = \frac{\sum_{i=1}^{N_{TSS}} score(i,j))}{\sum_{i=1}^{N_{TSS}} |gH2(i,j))|}$$

Taking advantage of the six months period that I spent at the Department of Systems Biology of Columbia University in the lab of Dr. A. Califano, I acquired a large number of expertise to perform bioinformatics analyses through the use of computational tools, such as: MACS 1.4.2 for peak calling analyses, PAVIS tool (update 10-28-2015) for location of previously found peaks relative to different genomic features, and SeqMINER 1.3.3 software to generate datasets signal heatmap, and K-means cluster was also processed.

4. RESULTS

4.1 OxiDIP-Seq Detects Genome-Wide Location of 8oxodG

To date, mapping of 8oxodG in mammalian cells has been faced using two different approaches. In the first one, genome-wide distribution of 8oxodG was performed by fluorescence in situ detection of 8oxodG on human metaphase chromosomes with signal resolution of megabases (Ohno et al 2006). More recently, a microarray technology-based protocol, allowed a 8oxodG distribution along rat genome with a resolution of kilobases (Yoshihara et al, 2014).

Here, we set up a novel technique, named OxiDIP-Seq, in order to profile the genome distribution of 8oxodG at a single nucleotide level, by using a specific antibody to immunocapture oxidized genomic fragments.

As depicted in figure 12, genomic DNA containing 8oxodG residues is extracted and fragmented by sonication; fragments are then denatured and immunoprecipitated using a specific antibody targeting 8oxodG residues. The step of denaturation is fundamental to expose DNA bases to the antibody, that reside inside the double-helix structure. After the immunoprecipitation step, the eluted DNA will be enriched in fragments containing 8oxodGs that can be analyzed by qPCR or sequenced by high-throughput sequencing. For the first time, we apply the NGS approach to immunocapturing of genomic oxidation, to have the most accurate map of 8oxodG genome distribution.

NGS offers several advantages compared to microarray. First, the material is directly sequenced and not interrogated by hybridization. Second, NGS offers single-nucleotide resolution allowing the comparison between samples differing in as little as one nucleotide of sequence. Finally, the signal from NGS approach is represented as an absolute number of sequence tags, and this allows to detect changes in

rare as well as in highly expressed sequences.
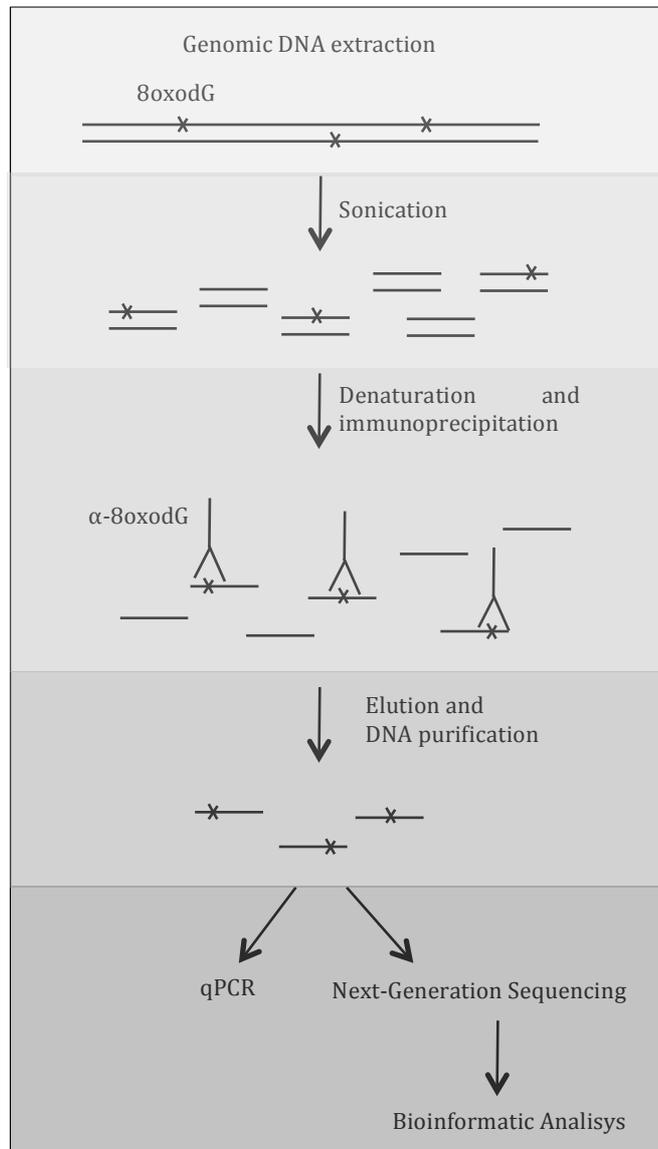
Figure 12. Steps for OxiDIP-Seq experimental procedure. Denatured genomic DNA of desired fragment length (generated by sonication) is incubated with antibody directed against 8oxodG, and oxidized DNA is isolated by immunoprecipitation (IP). Enrichment of target sequences in the oxidized fraction can be quantified by standard DNA detection methods, such as REAL-TIME PCR, or sequenced by Next Generation Sequencing.

To precisely map 8oxodG distribution in human genome, we carried out OxiDIP-Seq in *in vitro* immortailzed and non-tumorigenic human epithelial MCF10A cell line, because appropriate study of the DNA

damage requires cells that preserved intact the majors signaling pathways involved in DNA damage and repair. Data from two different biological experiments were analysed to create a first complete chromosomal map of DNA oxidation in human genome. Oxidized regions from the two datasets obtained were visualized by using University of California-Santa Cruz (UCSC) Genome Browser tool (Figure 13 shows a screenshot of a 2 Mbp window along chromosome 19).
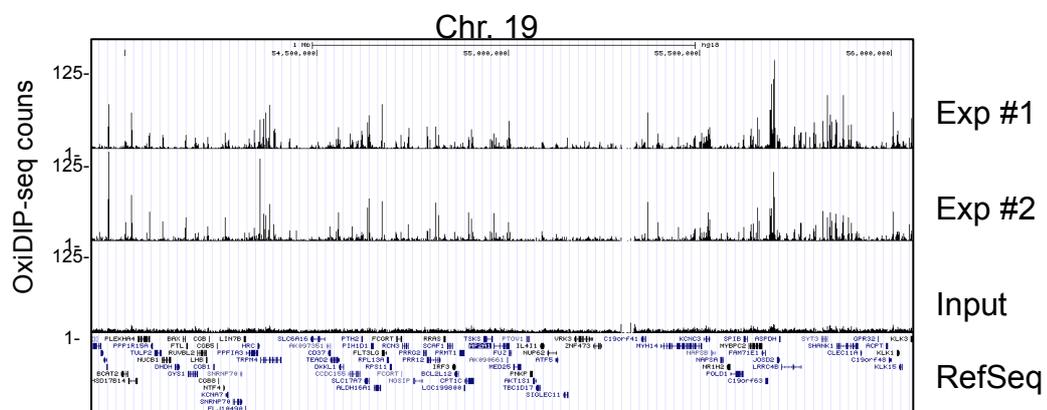
Figure 13. UCSC Genome Browser screenshot of OxiDIP-Seq peak spanning a 2 Mbp region along chromosome 19. Reads counts of two independent experiments are reported plus Input sample. Reference Sequence (RefSeq) collection of annotated genes is also reported.

This is the first high-resolution map of DNA oxidation in human genome, so we used these datasets to identify small amplicons on oxidized and non-oxidized genomic regions to validate DNA immunoprecipitaion by qPCR to have a quantitative validation of OxiDIP, in order to confirm the specificity of the antibody and the reliability of this technique.

We carried out OxiDIP experiment on untreated MCF10A analysing the immunoprecipitated fragments on genomic positive and negative regions indicated in Figure 14. Untreated sample shows an enrichment of 8oxodG, represented as percentage of input, that reflects the oxidation level visualized by UCSC Genome Browser): the positive

region, corresponding to a discrete peak of oxidation, shows a 8oxodG enrichment of 0,2% of input, while it is barely detectable on negative region (Figure 15); MCF10A cells were also UV-irradiated and treated with N-acetylcysteine (NAC). UV irradiation is involved in intracellular photoreactions giving rise to reactive oxygen species (ROS) that can oxidize DNA (Waster et al 2009), while NAC is an anti-oxidant that inhibits production of ROS. Thus, these treatments were used as positive and negative control of 8oxodG immunoprecipitation, respectively. Immunoprecipitated fragments were amplified by qPCR using amplicons reported in Figure 14 for quantitative analysis of 8oxodG enrichment.
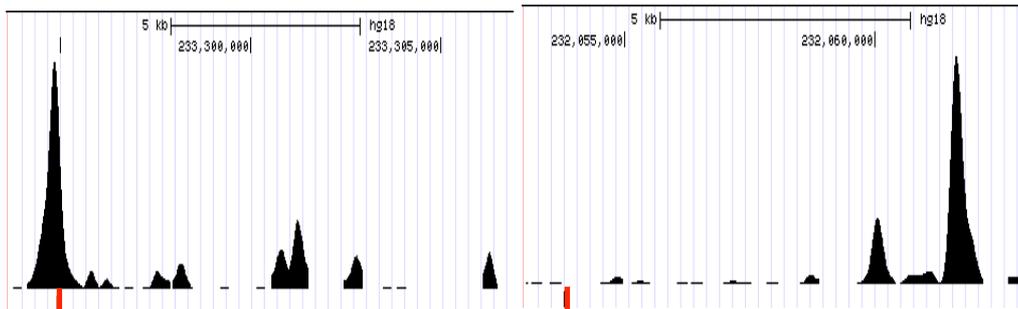
Figure 14. Oxidized (positive region) and non-oxidized (negative region) areas used to validate OxiDIP experiment. Red bar indicate the amplicon used for qPCR.

Interestingly, UV-irradiated sample shows an increase in 8oxodG level in both regions, suggesting that ROS produced by UV irradiation is also oxidizing the region that is not damaged in physiological conditions; as expected, the depletion of ROS molecules in NAC-treated cells drastically decreases the level of 8oxodG on positive region, showing level of oxidation comparable to non-oxidized DNA in physiological conditions (compare NAC positive region vs. Untr negative region).
These results indicate the specificity of the α-8oxodG antibody that we use for our OxiDIP experiments, as level of immunoprecipitated oxidized DNA increases in experimental conditions favouring DNA oxidation (UV-irradiation) and decreases in a condition preventing it (NAC treatment). Moreover, these results also show the reproducibility of our technique, that shows the same enrichment of 8oxodG in the

analyzed regions both by qPCR (Figure 15) and by sequencing (Figure 14).
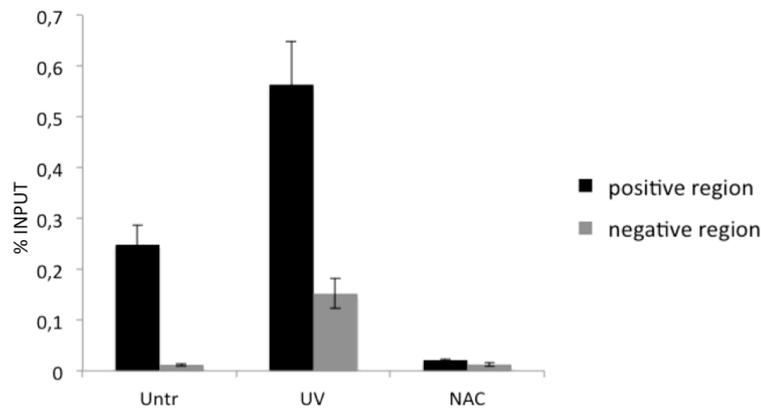
Figure 15. Enrichment of immunoprecipitated 8oxodG changes upon perturbation of redox cellular balance. 8oxodG antibody was used for OxiDIP in MCF10A cells. Immunoprecipitated sample was analyzed by qPCR using specific primers amplificating positive and negative regions reported in Figure 14. MCF10A were exposed to 40 J/m^2 UV light (254 nm) and treated with 1 mM N-acetyl cysteine (NAC) for two hours. Data from two independent OxiDIP assays were used to make % of input graphs presented along with standard deviations.

4.2   High-Resolution Genome-Wide Map of 8oxodG

For initial assessment of genome-wide OxiDIP-Seq signals, the averaged 8oxodG levels were plotted along the chromosomes and shown as ideograms. Map of chromosome 19 is shown in Figure 16, and the ideogram clearly indicates a not uniform distribution of regions with high 8oxodG levels and other regions conspicuously protected from DNA oxidation. To our knowledge, this map represents the most accurate 8oxodG genome distribution available so far. A genome-wide profiling of 8oxodG was recently obtained in rat genome, but the microarray-based approach used, provided a resolution of kilobases (Yoshihara et al 2014). The use of Next Generation Sequencing applied to DNA immunoprecipitation for genome-wide profiling of 8oxodG, allows to increase the resolution of the map.
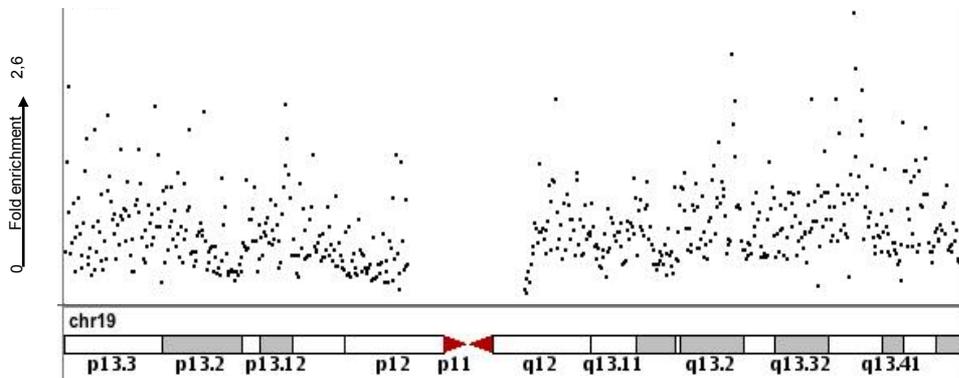
Figure 16. Assessment of genome-wide mapping of 8oxodG levels were plotted along the human chromosome 19 and shown as ideogram. Each dot represents an oxidized region and is expressed as fold enrichment respect to a sample control. The plots clearly indicate a not uniform distribution of regions with high 8oxodG levels and some regions conspicuously protected from DNA oxidation like centrosomes.

In order to identify discrete oxidized regions across the genome, a peak calling analysis was performed by using MACS (Model-based Analysis for ChIP-Seq) algorithm (Zhang et al 2008). MACS identifies regions in the genome that contain more sequencing reads than he would expect to see by chance. In this way, it detects discrete areas enriched in 8oxodG residues deriving from sequenced and aligned immunoprecipitated fragments.

MACS found 161,198 oxidized regions in MCF10A genome, with a distribution that is proportional to chromosome length. Indeed, number of peaks decreases with the decrease of chromosome length (Figure 17).
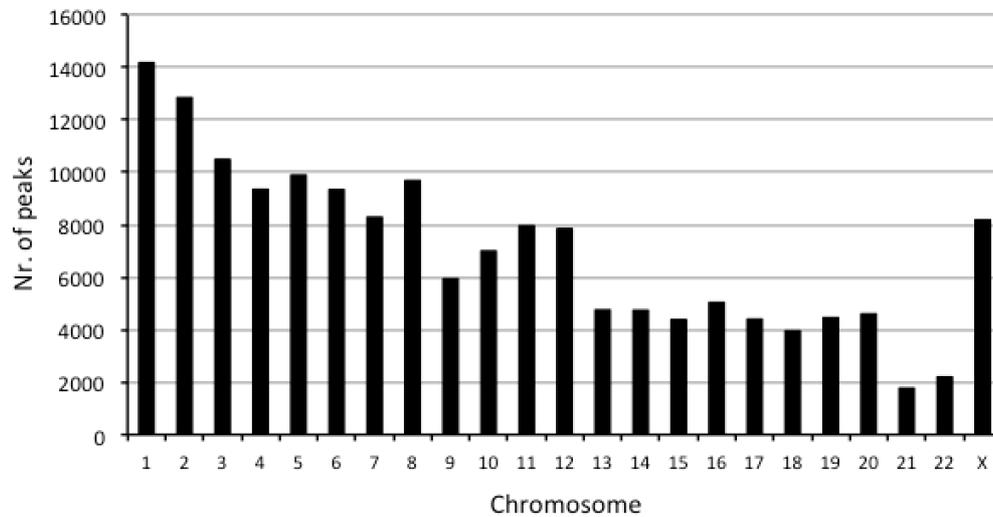
Figure 17. Graph reporting number of MACS-identified 8oxodG peaks for each human chromosome. Number of peaks is proportional to chromosome length.

Next, in order to assess whether the distribution of these peaks was correlated to some characteristics of genome, we annotated peaks found by MACS in respect to different genomic features by using Peak Annotation and VISualization (PAVIS) tool (Huang et al 2013). PAVIS allows to annotate and visualize peak location relative to different genomic feature. PAVIS reports the relative enrichment of peaks in these functionally distinct categories, and provides a summary plot of the relative proportion of peaks in each category.

We grouped 8oxodG peals in the context of genomic features: gene regions and intergenic regions. Gene region are further divided in Upstream (-5 Kbp from transcriptional start site), 5' UTR, exons, introns and Downstream (+5 Kbp from transcription termination site). Correlation of the positional information of these functional genome regions revealed that 8oxodG peaks localize with a slight preference in the genic region (52%, including properly genes and regulative sequences) rather than in the intergenic region (48%); within the gene-related sequences, oxidation accumulates mainly in the intron regions (38%) (Figure 18).
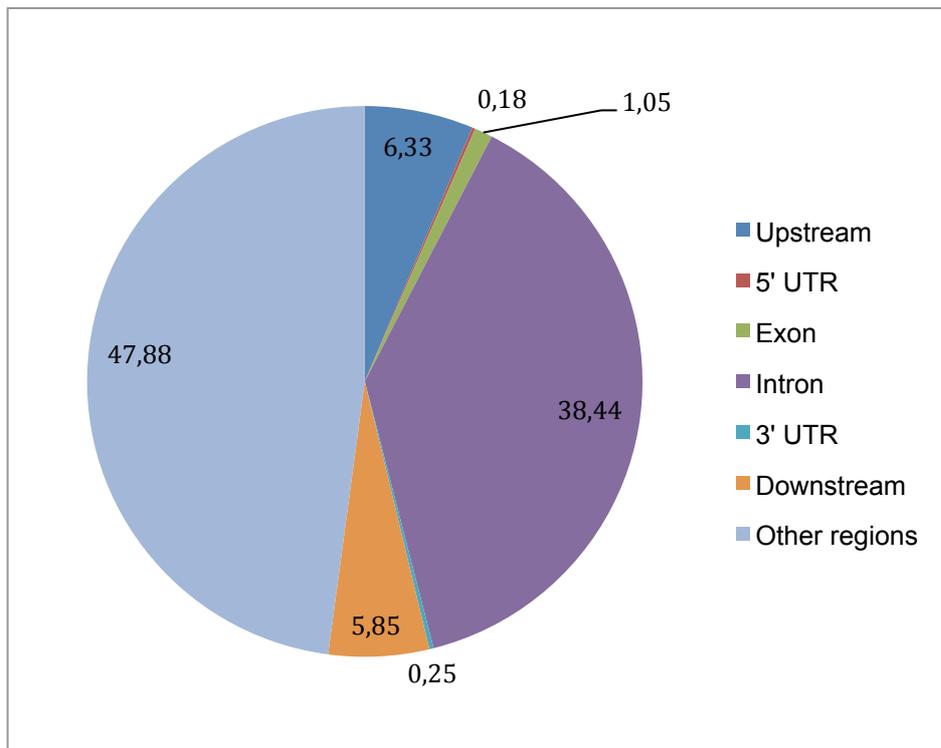
Figure 18. Visualization of the annotated 8oxodG peaks regions in human genome.Peaks of 8oxodG regions are grouped for their distribution in the context of genomic features: gene region and intergenic regions. Gene region are divided in Upstream (-5kb), 5'UTR, Exons; Introns; 3'UTR; Downstream (+5kb).

Then, PAVIS allows us also to measure 8oxodG distribution among different types of genes (protein coding genes, pseudogenes, genes transcribing ribosomalRNA, long intergenic non-coding RNA, microRNA, and other). Interestingly, we found that the vast majority of 8oxodG peaks localized in protein coding genes (74%), followed by genes transcribing long non-coding RNA (21%) (Fig 19).
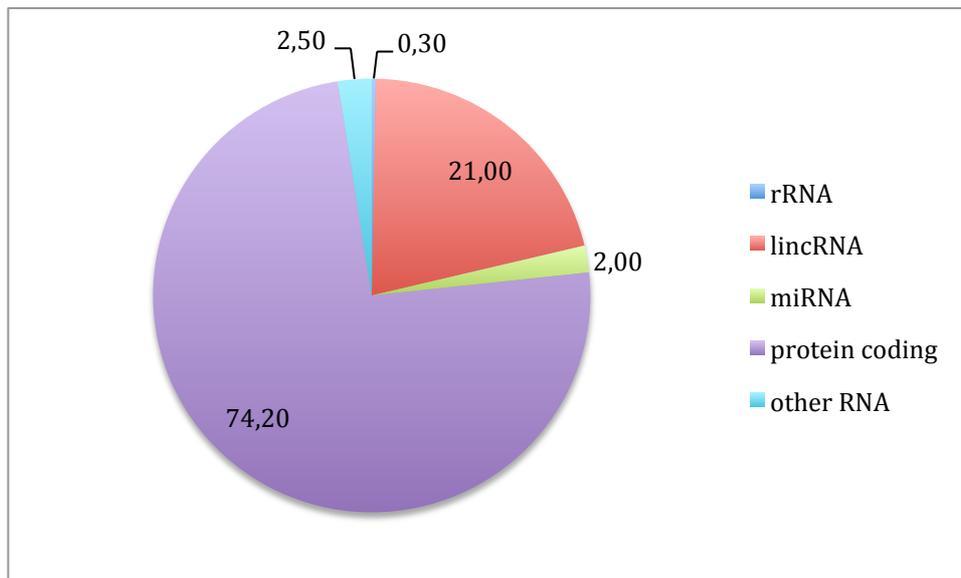
Figure 19. Visualization of the annotated 8oxodG peaks regions in human genome grouped for their distribution in the context of gene types: ribosomal RNA (rRNA), long intergenic non-coding RNA (lincRNA), micro RNA (miRNA), mRNA (protein coding), other RNA.

Taken together, these data suggest that distribution of 8oxodG-marked regions along human genome seems to be not stochastic, but rather preferentially occupies gene-related regions. Furthermore, among all transcribed genes, protein-coding genes reveal to be the most oxidized. The accumulation of 8oxodG peaks in genic regions and, in particular, in protein-coding genes, in fact, would confirm a role of DNA oxidative damage in facilitating transcription of genes transcribed by RNA polymerase II (RNAPolII).

As previously shown, the accumulation of 8oxodG in gene-related regions and, among these, particularly in protein-coding genes, prompt us to further investigate the role of DNA oxidation in transcription.

In order to further explore this aspect, we focused our analyses within genes. We investigated the distribution of 8oxodG residues around human Transcriptional Start Sites (TSS), measuring their frequency in a 10 Kbp region surrounding each start site. TSS human genomic

coordinates were retrieved from UCSC database, and to adopt a conservative approach we selected only TSSs having a confidence score ≥20. Accordingly to this criterion, we analyzed a total of 27,487 TSSs. Analysis revealed a significant 8oxodG frequency surrounding human TSSs. In particular, showing an evident depression of signal on TSS, followed by a clear peak of oxidized residues within 1 Kbp downstream TSS. Intriguingly, this pattern is specific for TSS-surrounding region (Figure 20A), as no particular peak of DNA oxidation is present around the transcriptional end region (Figure 20B).
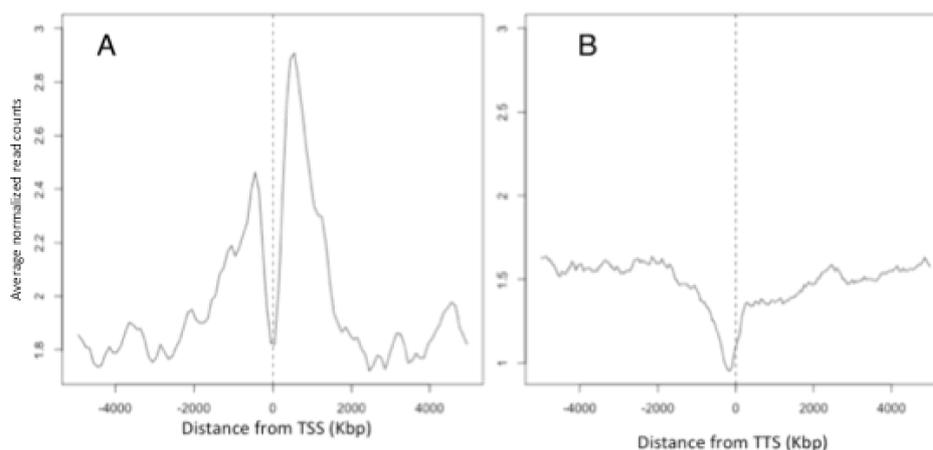
Figure 20. Profile of 8oxodG around TSS (A) and TTS region (B). Average read counts were calculated in a region spanning 5 Kbp downstream and upstream TSS and TTS.

These analyses show that 8oxodG-marked DNA oxidation preferentially occupies the region where transcription complex is assembled and where RNA Polymerase begins to elongate, further confirming that DNA oxidation preferentially occupies transcription-related genomic regions.

4.3   Genomic Distribution of γH2AX

Based on the assumption that oxidative DNA damage facilitates transcription through generation of DSB, we also performed in MCF10A cells a genome-wide analysis (ChIP-Seq) of γH2AX. The function of this phosphorylated variant histone of H2A is to demarcate DSB sites. When a DSB is generated, ataxia telangiectasia mutated (ATM) kinase phosphorylates H2AX in an early event of the cellular DNA damage response (DDR) to DSB. Phosphorylated H2AX (γH2AX) serves as docking station for the recruitment of the downstream effectors of DDR, leading to amplification of the signalling cascade. (Valdigleasias et al 2012). Thus, γH2AX is considered a universal biomarker of DSB presence.

Using the same bioinformatics tool, we investigated γH2AX genome distribution in the same genomic features previously queried for 8oxodG. We identified 20,440 enriched regions of γH2AX preferentially localized once again in the genic regions (60%) compared to intergenic region (40%) (Figure 21), thus showing a distribution that is similar to the oxidized peaks.
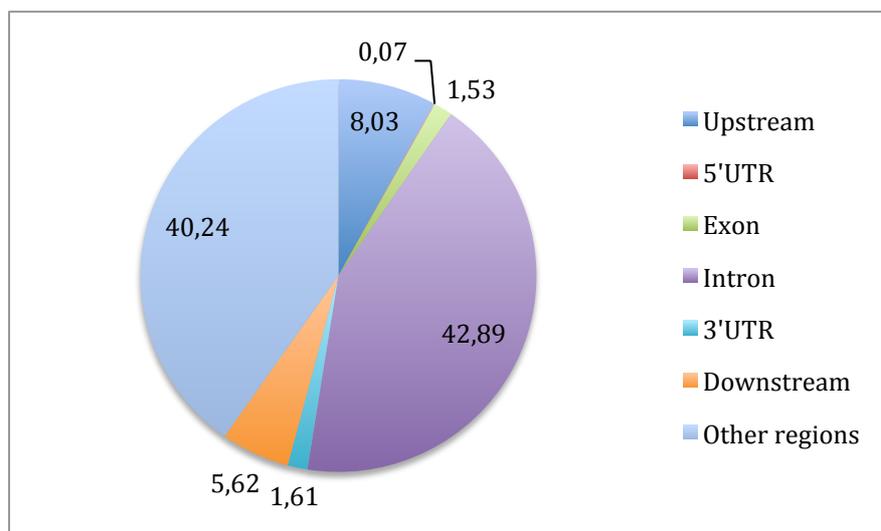
Figure 21. Visualization of the annotated γH2AX peaks regions in human genome.Peaks of γH2AX regions are grouped for their distribution in the context of genomic features: gene region and intergenic regions. Gene region are divided in Upstream (-5kb), 5'UTR, Exons; Introns; 3'UTR; Downstream (+5kb).

Next, we repeated for γH2AX the same TSS analyses previously shown for 8oxodG. Interestingly, γH2AX shows a similar profile of distribution, with TSS that is protected by γH2AX (Figure 22A) as well as 8oxodG accumulation, compared to surrounding regions. Also in this case, accumulation of γH2AX is higher around TSS region that TTS region (Figure 22B).
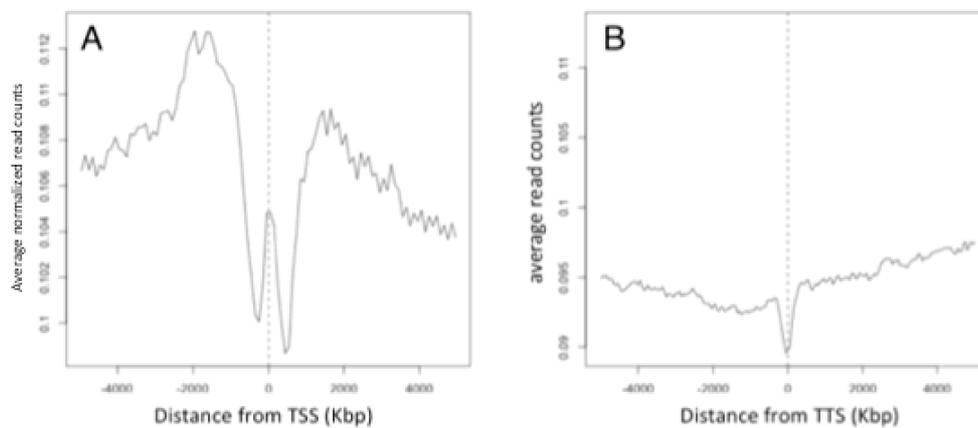
Figure 22. Profile of γH2AX around TSS (A) and TTS region (B). Average read counts were calculated in a region spanning 5 Kbp downstream and upstream TSS and TTS.

Similarity of TSS profiles of 8oxodG and γH2AX is likely to suggest that processing of oxidized bases could actually lead to the generation of DNA breaks as reported in literature.

Further analyses to clarify this aspect will be discussed below.

4.4 Correlation between DNA Damage and Transcription

To further confirm the hypothesis that oxidative DNA damage correlates with gene transcription, we decided to compare the relative occupancy of DNA damage markers (8oxodG, γH2AX and NBS1) with transcription markers (RNA Polymerase II phosphorylated on CTD-Ser5, on CTD-Ser2, and mRNA) along all annotated human genes. We used our datasets for 8oxodG, γH2AX, NBS1 and RNAPol2-Ser5 (unpublished), and datasets already available in literature (Pol2-Ser2: GSE45715, Gardini et al 2014; RNA-Seq: GSE45258, Kang et al 2013) (Table 2). Among all datasets of RNA-Seq available in the databases for MCF10A cells, we chosen the one with the highest depth of sequencing (100 milions reads) in order to obtain a more accurate gene expression profile required for our analyses.

Dataset	Source
OxiDIP-Seq α-8oxodG	this work
ChIP-Seq α-γH2AX	this work
ChIP-Seq α-NBS1	this work
ChIP-Seq α-RnaPol2-Ser2	GSE45715
ChIP-Seq α-RnaPol2-Ser5	this work
RNA-Seq	GSM1100206

Table 2. List of datasets used in this work.

The two phosphorylated forms of RNA PolII, Ser5 and Ser2, mark respectively initiating and elongating RNA Pol II (Cheng et al 2003). Nijmegen breakage syndrome 1 (NBS1), as well γH2AX, is considered a biomarker of DSB. NBS1 forms a multimeric complex with DDR factors MRE11/RAD50 and recruits or retains them at the vicinity of sites of DNA damage by direct binding to phosphylated histone H2AX, where

it proceeds to rejoining of the broken DNA ends (Kobayashi et al 2004).

To determine the relative occupancy along all annotated genes of datasets reported in Table 2, we computed a bioinformatics analysis by using seqMINER, a computational platform that allows comparison and integration of multiple datasets and extraction of qualitative and quantitative information (Ye et al 2010).
In Figure 23, data are represented as heatmap, where each single horizontal line shows the read density mapped on TSS, gene body, TTS, and a 5 Kbp flanking region of each unique genes (intensity of color red in the heatmap is proportional to signal intensity of the corresponding dataset). The five datasets used are clustered according to the similarity of signal intensity and, based on signal densities of RNA-Seq and Pol2 datasets, human genes were grouped in three clusters basing on different expression level and:

1) Cluster 1 contains low expressed (or not expressed) genes;
2) Cluster 2 includes moderately expressed genes;
3) Cluster 3 comprises most expressed genes.

To further define the differences between the 3 clusters, data are also presented as plots in Figure 24. Plots represent the average gene profiles of read density, calculated averaging the read density mapped on total genes grouped in each cluster for each dataset.
As reported in Figure 24, average level of two RNA polymerase perfectly reflects the increasing level of transcription in the three clusters (Figure 24, A, B and C). Indeed,

a) in Cluster 1 (Figure 24A), level of both RNAPOL2 are lower in the gene body compared to outside 5′ and 3′ flanking region, suggesting an absent or very low level of transcription;
b) in Cluster 2 (Figure 24B), we observe an increase of general level of RNAPOL2s within the gene body, with an accumulation of Ser2 form on TSS and an accumulation of Ser5 along the whole gene, while the outside level are the same of Cluster 1; this

demonstrates a higher level of transcription in this Cluster than Cluster 1;
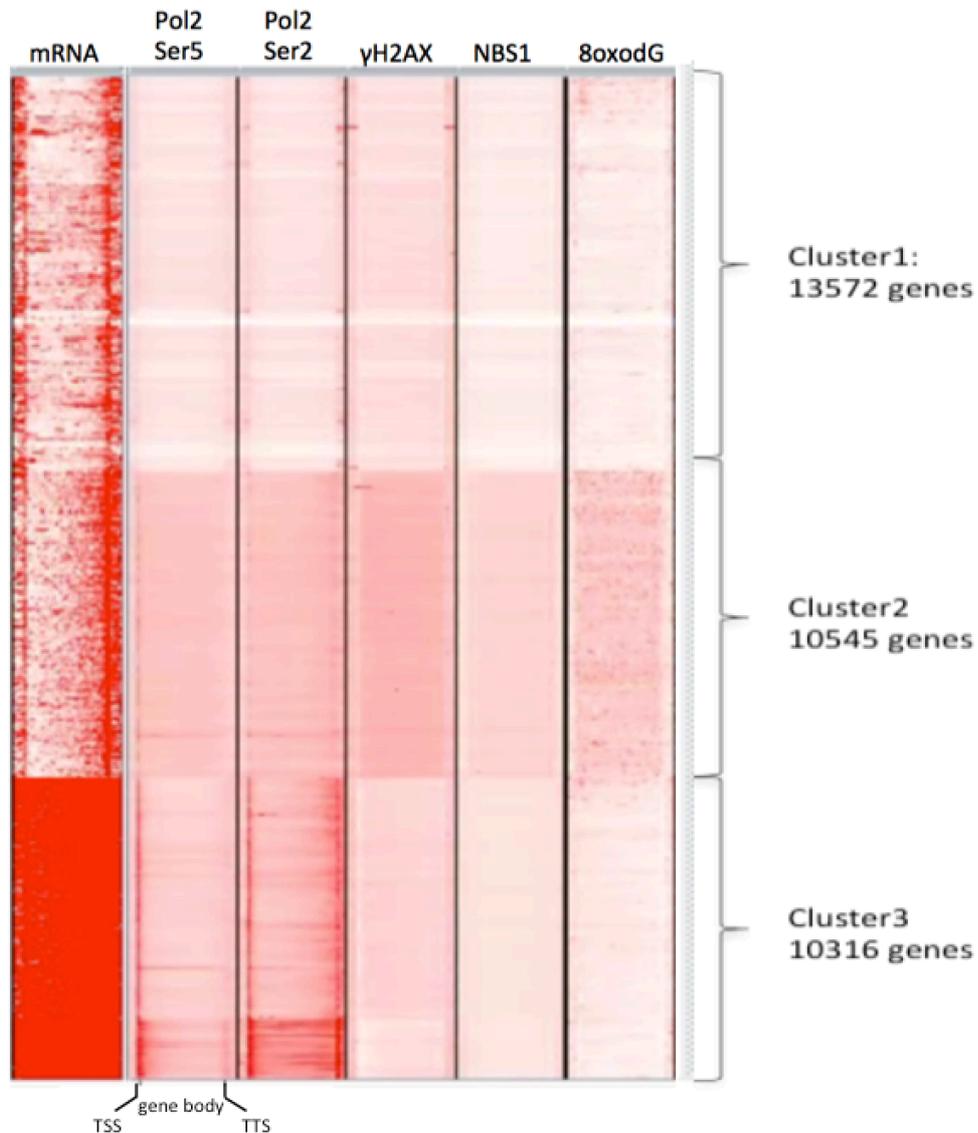
Figure 23. Clustering analysis of ChIP-Seq of γH2AX, NBS1, RNAPol2-Ser5, RNAPol2-Ser2, OxiDIP-Seq of 8oxodG and RNA-Seq data from MCF10a cells were subjected to unbiased clustering, in a region encompassing 10 Kbp of unique RefSeq genes, using seqMINER bioinformatics tool. Original gene length was divided in arbitrary 160 bins to obtain the same virtual length. Kmeans linear clustering was used for the analysis.

c) in Cluster 3 (Figure 24C), this increase is more evident, showing a much higher increase of transcription level; again, this increase is specific for gene-body region.

Comparing the RNAPol2-defined gene expression level of the three clusters with the considered DNA damage markers, we observe that:

i) low or not transcribed genes of Cluster 1 show intragenic level of 8oxodG, γH2AX and NBS1 that are very low, especially if compared to 5′ and 3′ flanking region of the genes, implying that this subset of genes is not damaged (Figure 24D);

ii) expressed genes in Cluster 2 show a peculiar increase of the damage markers along the gene body respect to flanking regions (Figure 24E);

iii) Highly expressed genes (Cluster 3) show a signal intensity largely similar in both flanking and gene-body regions (Figure 24F).
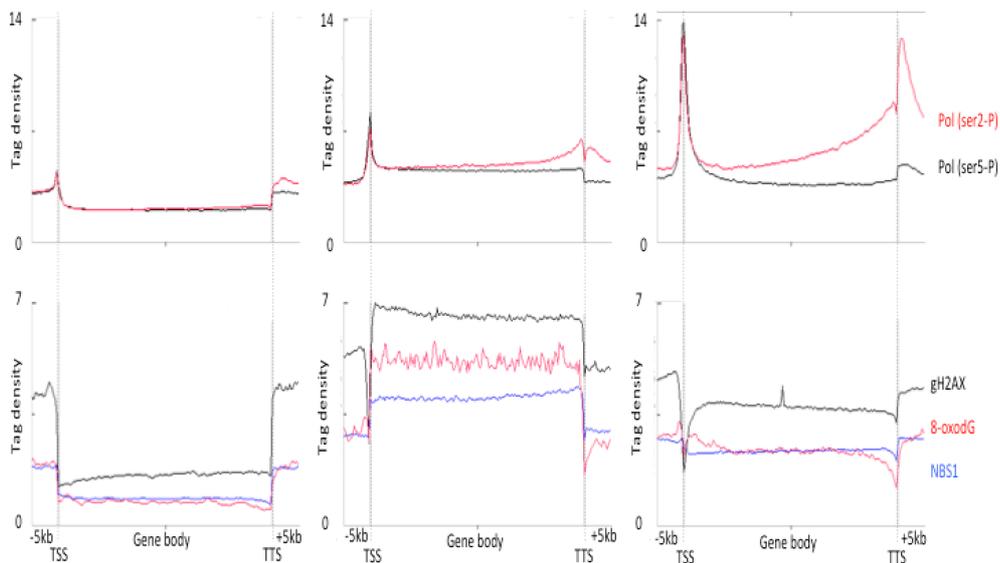
Figure 24. Average profile of Pol2-Ser5 and Pol2-Ser2 in Cluster 1, 2 and 3 (A, B, C), and γH2AX, NBS1 and 8oxodG in Cluster 1, 2 and 3 (D, E, F) for selected clusters was reported in a region encompassing 10 Kbp of RefSeq genes. Original gene length was divided in arbitrary 160 bins to obtain the same virtual length.

Taken together, these data show that there is a significant correlation among occupancy of DNA damage markers (8oxodG, γH2AX, NBS1) in the gene body, suggesting that DSB is actually a consequence of oxidative DNA damage; furthermore, this intragenic occupancy of DNA damage marker is associated to transcription. Indeed, low or not transcribed genes (Cluster 1) do not accumulate damage, that instead increases in the gene body compared to outside region in expressed genes (Cluster 2).

Interestingly, only high expressed genes (Cluster 3) do not show a further accumulation of DNA damage, and this accumulates at the same level inside and outside the gene.

This behavior, never observed before, likely suggests that cells could repair transcription-induced oxidative damage more efficiently when genes are highly expressed than moderately expressed (cluster 3 vs. cluster 2), probably because highly expressed genes are hotspot of damage and require an immediate repair in order to maintain genome integrity. Collectively, our findings suggest the model, according to which scheduled DNA oxidative damage facilitates transcription, is correct when referred to a subset of genes, such as moderately expressed genes, that act a "scheduled" DNA damage to accomplish their expression (Di Palo et al, manuscript in preparation).

4.5   Mouse Genome-Wide Distribution of 8oxodG

In order to assess whether 8oxodG genomic profile is conserved among species, we carried out OxiDIP-Seq experiments and preliminary bioinformatics analyses also in mouse embryonic fibroblasts (MEFs).
Sequenced fragments were aligned on mouse genome (assembly NCBI37/mm9), and a screenshot of 4 Mbp visualized by UCSC

Genome Browser is shown in Figure 24, reporting tracks of two independent experiments, input sample and Gencode v22 collection genes annotation.
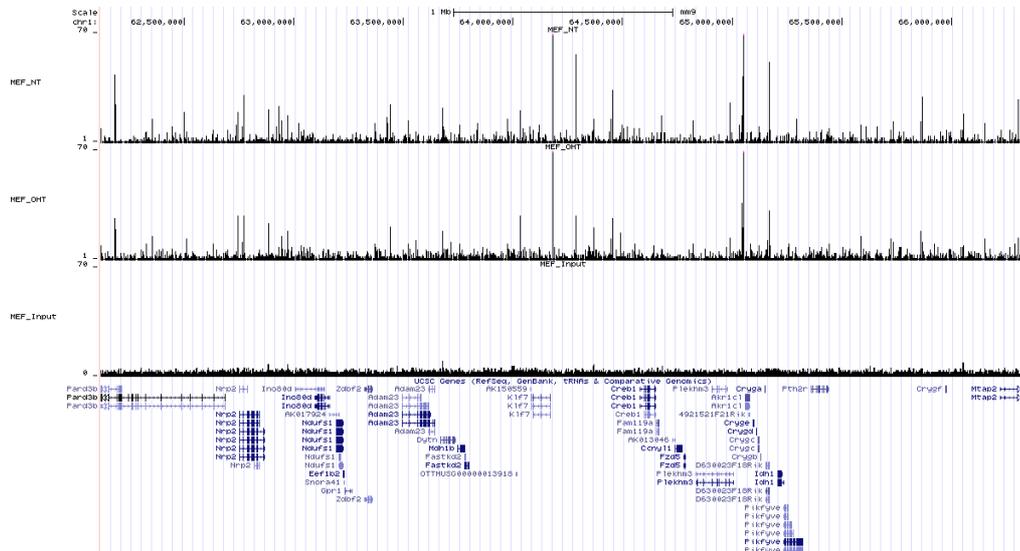
Figure 24. UCSC Genome Browser screenshot of OxiDIP-Seq peak spanning a 2 Mbp region along mouse chromosome 1. Reads counts of two independent experiments are reported plus Input sample. Gencode v22 collection of annotated genes is also reported.

We found 76,033 oxidized regions in MEF genome, with a distribution that is globally proportional to chromosome length (Figure 25).
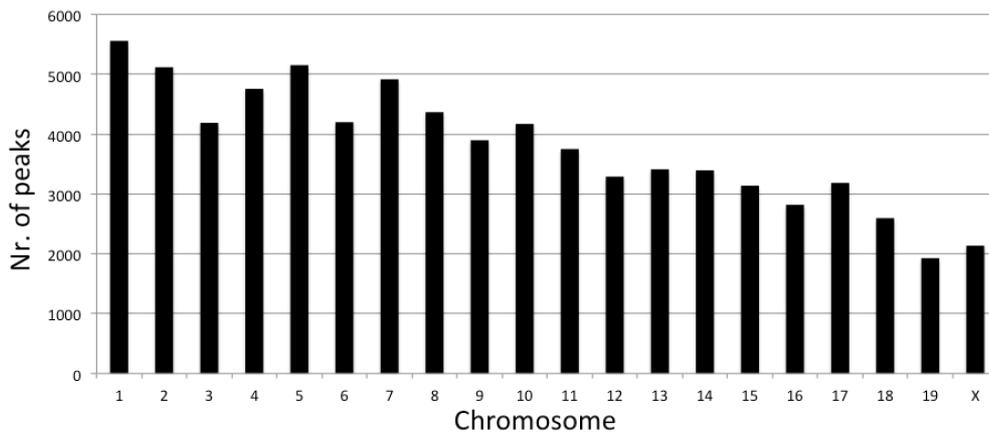
Figure 25. Graph reporting number of MACS-identified 8oxodG peaks for each human

chromosome. Number of peaks is proportional to chromosome length.

Also in this case, we annotated oxidized peaks in the context of genomic features: gene regions and intergenic regions. Figure 26 shows that in mouse genome oxidation peaks are equally distributed between intergenic and genic regions. Moreover, similarly to human genome, introns are the areas most enriched in oxidation within genic regions.
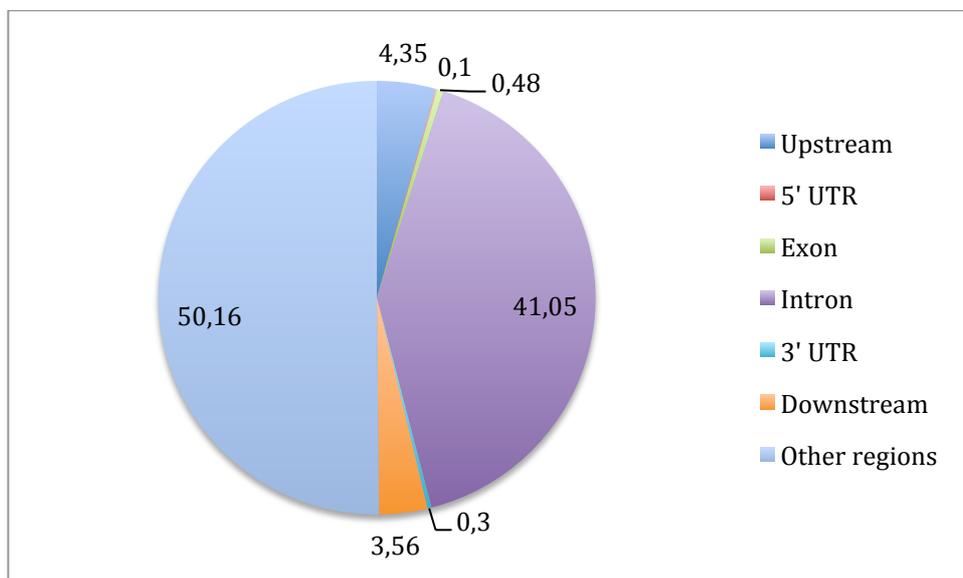
Figure 26. Visualization of the annotated 8oxodG peaks regions in mouse genome. Peaks of 8oxodG regions are grouped for their distribution in the context of genomic features: gene region and intergenic regions. Gene region are divided in Upstream (-5kb), 5'UTR, Exons; Introns; 3'UTR; Downstream (+5kb).

Furthermore, analysis of 8oxodG profile around transcriptional start site reveals also in this case a depletion of oxidation in correspondence of TSS; however, this characteristic does not seem to be specific of 5'-end of mouse genes, as we observe a similar profile in 3'-end as well (Figure 27).
Furthermore, in the 5 Kbp region encompassing TSS, we do not observe a specific region enriched in oxidation (like the peak of oxidation within 1 Kbp downstream human TSS, Figure 21), but rather a spread signal.

This different behavior of 8oxodG distribution around TSS of mouse and human genes may be due to a biological explanation, as MEFs and MCF10A cells are very different from each other in the context of gene expression programs, epigenetic state, morphology, thus distribution of DNA oxidation might be cell-specific.

Further analyses are required to correlate DNA oxidation and transcription in mouse genome as well, and to carry out a comparison between human and mouse context.
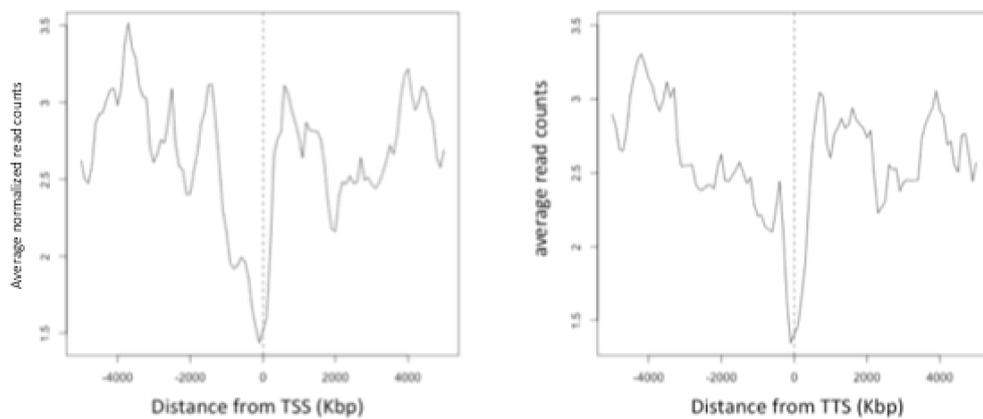
Figure 27. Profile of 8oxodG around mouse TSS (A) and TTS region (B). Average read counts were calculated in a region spanning 5 Kbp downstream and upstream TSS and TTS.

5. DISCUSSION

In this study, we used a genome-wide approach to provide further evidence supporting the correlation existing between DNA damage and transcription.

We focused on oxidative DNA damage, obtaining for the first time a high-resolution map of 8oxodG distribution along human genome; data presented show an accumulation of oxidation in gene-related regions (Figure 18) and, among these, in protein coding genes (Figure 19).

Next, we carried out γH2AX and NBS1 ChIP-Seq experiments, and we compared the occupancy of DNA damage markers depending on transcription. We found a strong correlation of 8oxodG and double-strand break markers (γH2AX, NBS1) with moderately expressed genes compared to low or not transcribed genes. Interestingly, level of these damage markers is not enriched in a specific intragenic region, but rather accumulates throughout the gene body (Figure 24).

Taken together, the data suggest a requirement of DSB-related DNA damage for transcription activation.

Although 8oxodG is the most studied oxidative DNA damage, little is known about how it is distributed in human genome. Literature reports in detail how it is generated and repaired, and many studies show a quantitative correlation of oxidative DNA stress in cancer, aging, neurodegenerative and cardiovascular disease (Cooke et al, 2003). However, so far, successful attempts to identify oxidative DNA stress in the genome are represented by immunological detection of 8oxodG on human metaphase chromosome, that gives a signal resolution of Megabases and, more recently, by a microarray technology applied to immunoprecipitation of 8oxodG in rat genome allowed to increase the resolution to kilobases (Yoshihara et al, 2014). These information are still insufficient to try to correlate DNA oxidation with genomic features, like gene-related regions.

To further increase the resolution of genomic map of DNA oxidation, we developed an innovative methodology (OxiDIP-Seq) that applies high-throughput next-generation sequencing to 8oxodG immunoprecipitation.

In this way, we could use genome-wide approach to the study of oxidative DNA damage associated to transcription in a steady-state condition, without perturbing cell context with external stimuli that could change cellular transcriptional programs.

Previous works, in fact, show a requirement of DNA oxidation for transcription only in a specific subset of genes, whose transcription was induced in vitro. Perillo et al. reported how DNA breaks, generated upon processing of 8oxodG residues, were necessary for transcription activation of oestrogen-responsive genes (Perillo et al 2008); similarly, Amente et al. demonstrated that LSD1-mediated activation drives the transcription of Myc target genes (Amente et al 2010).

Conversely, the genome-wide approach used in our work, allows for the first time to study association between DNA damage (and oxidative DNA damage in particular) with transcription globally, no longer restricted to a specific subset of genes. This method allows to answer the question whether dependence of transcription on DNA damage could be universally applied.

Furthermore, the large amount of data coming from sequenced immunoprecipitated DNA were processed by using modern computational tools, in order to investigate DNA oxidation in different genomic features. Our analyses show a higher susceptibility of protein coding genes to be oxidized (Figure 20). This seems to suggest that DNA oxidation is a driving force for transcription carried out by RNA polymerase II, likely due to the recruitment of factors, e.g.: histone demethylaseLSD1, that produce ROS molecules during their enzymatic activity.

Despite how and by which factor oxidative damage is generated in the DNA, its processing by BER pathway contemplates the formation of DNA strand breaks that were demonstrated to be necessary for chromatin relaxing mediated by topoisomerases. Once again, several

studies report the requirement of DNA breaks to activate transcription of only a subset of genes. Puc et al. and Madhabhushi et al, for example, demonstrated that neuronal stimulus-inducible genes and androgen-responsive genes were activated upon DSBs operated by endonuclease activity of topoisomerase I and II, respectively.

Our γH2AX and NBS1 ChIP-Seq data show that oxidized DNA is also marked by these two factors universally considered markers of DSB presence (Figure 23). This suggests that 8oxodG-marked DNA can actually be processed by BER factors and topoisomerases to produce DNA breaks. Interestingly, this damage is equally wide-spread throughout the gene, compared to the outside flanking regions, suggesting that the whole gene body is equally prone to be damaged by oxidation. The fact that γH2AX and NBS1 also spread throughout the whole gene body does not necessary means that each oxidized residue is converted to a DSB. γH2AX, in fact, is well known to spread for megabases around a DSB, while NBS1 was demonstrated to mediate also chromatin relaxing (Saito et al 2016), suggesting that this role of NBS1 could be required along the gene body.

Additionally, we investigated the level of these DNA damage markers in dependence of transcription, grouping all annotated human genes in 3 Clusters characterized by progressively higher level of transcription (Figure 23). Data presented show that the event of transcription is actually accompanied by a strong increase of DNA damage, as not (or very low) expressed genes do not show any damage in the gene body (compare Cluster 2 and Cluster 1 of Figure 23). However, DNA damage level is not proportional to transcription intensity, as highest transcribed genes do not show highest accumulation of damage (compare Cluster 3 and Cluster 2 of Figure 23). This phenomenon could be explained by the fact that, since highly expressed genes are more prone to be damaged, perhaps because their chromatin architecture makes them more susceptible to damaging agents, DNA damage response is more efficient, and damage accumulates to a lesser extent.

We can not state that damage is functional to transcription, but a clear correlation between these two events is evident.

In order to assess whether this correlation is an evolutionary conserved feature or rather species- or cell type-specific, we applied analyses of 8xodG distribution in mouse genome. Unlike human genes, preliminary data of 8oxodG distribution do not show a profile that is specific of promoter region, but a background noise seems to characterize the 10 Kbp region flanking TSS and TTS. Thus, data reported are still insufficient to clearly say that DNA damage and transcription correlate also in this species, but further analyses are on going to establish this issue in mouse genome.
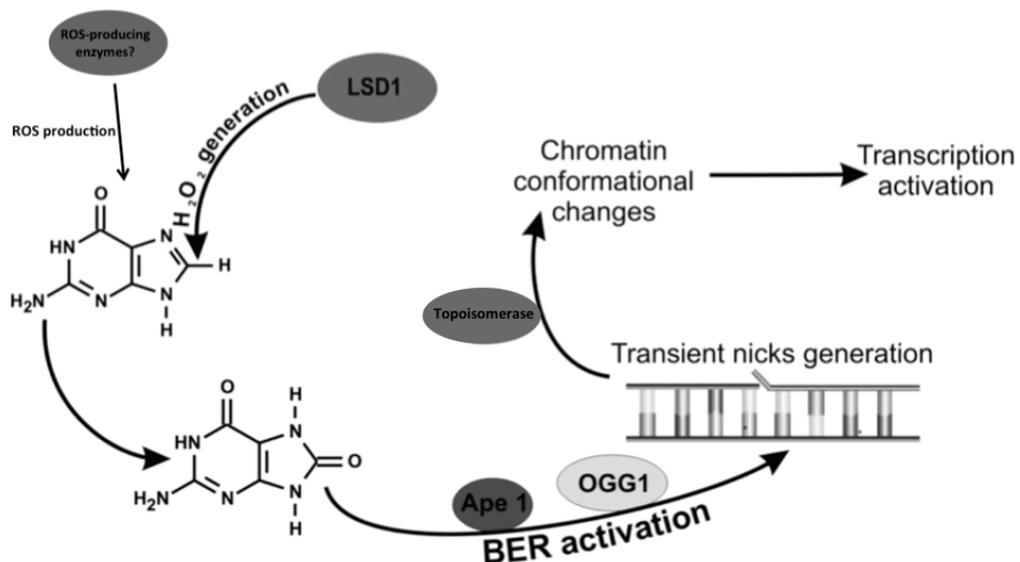
Figure 28. Possible involvement of 8-oxoGua generation in chromatin relaxation and transcription activation.

In conclusion, our findings add further evidences in support to the model according to which 8oxodG-marked DNA oxidation, generated by nuclear ROS-producing enzymes (e.g.: demethylase LSD1), is converted into DNA breaks by BER factors, and this break can be used by topoisomerase to induce chromatin conformational changes that are required to activate transcription (Figure 28) (Zarakowska et al 2013).

After all, 8oxodG would not be the first modified base with a role in regulation of gene expression. 5-methylcytosine, the cytosine nucleotide modified by the addition of a methyl group to its 5th carbon, is a very important repressor of transcription since, when present in promoters, it is associated with stable, long-term transcriptional silencing, and this occur by either blocking positive transcription factors, or promoting the binding of negative ones (Defossez and Stancheva, 2011).

We planned further experiments to definitely validate the DNA oxidation-mediated transcriptionIn model, and if this will occur, we should start to consider 8oxodG as a new "epigenetic marker", because of its involvement in the regulation of gene transcription.

6. CONCLUSIONS

Here, we report a study of correlation between DNA damage and transcription.

We developed OxiDIP-Seq methodology to precisely map oxidized guanine residues (8oxodG) in human and mouse genomes.

By using a genome-wide approach, we found enrichment of 8oxodG in gene-related regions rather than intergenic regions. This oxidative DNA damage strongly correlates with biomarkers of DNA double-strand breaks along human genes,. Indeed, the overlap between OxiDIP-Seq, γH2AX and NBS1 ChIP-Seq data suggests that 8oxodG can be converted in DNA breaks by factors of base excision repair pathway.

Furthermore, increase of DNA damage is also associated with the entrance of RNA polymerase II in the gene body, suggesting a role of 8oxodG in regulation of protein coding genes transcription.

Same analyses are on going to investigate this issue also in mouse genome.

In conclusion, the findings reported in this study add further support to the role of DNA damage and transcription. Furthermore, this is the first example that defines the oxidative DNA damage as a function of transcription, thus providing circumstantial evidence that 8oxodG may play a novel undefined new role as "epigenetic marker".

Additional works are clearly required for a deep understanding of this proposal.

7. REFERENCES

Ambrosio S, Di Palo G, Napolitano G, Amente S, Dellino GI, Faretta M, Pelicci PG, Lania L, Majello B. Cell cycle-dependent resolution of DNA double-strand breaks. Oncotarget. 2016; 7(4):4949-60.

Amente S, Bertoni A, Morano A, Lania L, Avvedimento EV, Majello B. LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc-induced transcription. Oncogene. 2010; 29(25):3691-702.

Amente S, Lania L, Avvedimento EV, Majello B. DNA oxidation drives Myc mediated transcription. Cell Cycle 2010; 9(15):3002-4.

Aguilera A, Garcia-Muse T. R loops: from transcription byproducts to threats to genome stability. Mol. Cell 2012; 46, 115-124

Bradsher J, Auriol J, Proietti de Santis L, Iben S, Vonesch JL, Grummt I, Egly JM. CSB is a component of RNA pol I transcription. Mol Cell 2002; 10, 819-829.

Cooke MS, Evans MD, Dizdaaroglu M, Lunec J. Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J. 2003; 17(10), 1195-1214.

Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell 2011; 146, 67-79.

D'Antonio M, D'Onorio De Meo P, Pallocca M, Picardi E, D'Erchia AM, Calogero RA, Castrignanò T, Pesole G. RAP: RNA-Seq Analysis Pipeline, a new cloud-based NGS web application. BMC Genomics. 2015;16:S3.

Defossez PA, Stancheva I. Biological functions of methyl-CpG-binding proteins. Prog Mol Biol Transl Sci. 2011;101:377-98.

Fong YW, Cattoglio C, Tjian R. The intertwined roles of transcription and repair proteins. Mol. Cell 2013; 52(3): 291-302.

Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol. Cell 2012; 45(6): 814-825.

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

Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. Nature 2011; 411, 366-374.

Hollstein M, Shomer B, Greenblatt M, Soussi T, Hovig E, Montesano R, Harris CC. Somatic Point Mutations in the *p53* Gene of Human Tumors and Cell Lines. Nucl. Acids Res. 1996; 24 (1): 141-146.

Huang W, Loganantharaj R, Schroeder B, Fargo D, Li L. PAVIS: a tool for Peak Annotation and Visualization. Bioinformatics. 2013 Dec 1;29(23):3097-9

Iyama T, Wilson DM. DNA repair mechanisms in dividing and non-dividing cells. DNA Repair 2013; 12, 620-636.

Ju BG, Lunyak VV, Perissi V, Garcia-Bassets I, Rose DW, Glass CK, Rosenfeld MG. A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. Science 2006; 312, 1798-1802.

Klaunig JE, Kamendulis LM, Hocevar BA. Oxidative stress and oxidative damage in carcinogenesis. Toxicol. Pathol. 2010; 38, 96-109.

Kobayashi J, Antoccia A, Tauchi H, Matsuura S, Komatsu K. NBS1 and its functional role in the DNA damage response. DNA Repair (Amst). 2004; 3(8-9):855-61.

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

Lemon B, Tjian R. Orchestrated response: a symphony of transcription factors for gene control. Genes Dev. 2000; 14, 2551-1569.

Lin C, Yang L, Tanasa B, Hutt K, Ju BG, Ohgi K, Zhang J, Rose DQ, Fu XD, Glass CK, Rosendfeld MG. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell 2009; 139, 1069-1083.

Lindahl T, Barnes DE. Repair of endogenous DNA damage. Cold Spring Harb. Symp. Quant. Biol. 2000; 65, 127-133

Madabhushi R, Gao F, Pfenning AR, Pan L, Yamakawa S, Seo J, Rueda R, Phan TX, Yamakawa H, Pao PC, Stott RT, Gjoneska E, Nott A, Cho S, Kellis M, Tsai LH.Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. Cell 2015; 161(7), 1592-1605.

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

Ohnishi S, Saito H, Suzuki N, Ma N, Hiraku Y, Murata M, Kawanishi S. Nitrative and oxidative DNA damage caused by K-ras mutation in

mice. Biochem. Biophys. Res. Commun. 2011; 413(2):236-40.

Ohno M, Miura T, Furuichi M, Tominaga Y, Tsuchimoto D, Sakumi K, Nakabeppu Y.
A genome-wide distribution of 8-oxoguanine correlates with the preferred regions for recombination and single nucleotide polymorphism in the human genome. Genome Res. 2006; 16(5):567-75.

Perillo B, Ombra MN, Bertoni A, Cuozzo C, Sacchetti S, Sasso A, Chiariotti L, Malorni A, Abbondanza C, Avvedimento EV. DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. Science 2008; 319, 202-206.

Puc J, Kozbial P, Li W, Tan Y, Liu Z, Suter T, Ohgi KA, Zhang J, Aggarwal AK, Rosenfeld MG. Ligand-dependent enhancer activation regulated by topoisomerase-I activity. Cell 2015; 160(3), 367-380.

Rahmouni AR and Wells RD. Direct evidence for the effect of transcription on local DNA supercoiling in vivo. J. Mol. Biol. 1992; 223, 131-144.

Robertson AB, Klungland A, Rognes T, Leiros I. Base excision repair: the long and short of it. Cell. Mol. Life Sci. 2009; 66, 981-993.

Saito Y, Zhou H, Kobayashi J. Chromatin modification and NBS1: their relationship in DNA double-strand break repair. Genes Genet Syst. 2016 Jan 6;90(4):195-208.

Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu. Rev. Biochem. 2004; 73, 39-85.

Scott TL, Rangaswamy S, Wicker CA, Izumi T. Repair of oxidative DNA damage and cancer: recent progress in DNA base excision repair. Antioxid. Redox Signal. 2014; 20(4), 708-726.

Shikazono N, Pearson C, O'Neill P, Thacker J. The roles of specific glycosylases in determining the mutagenic consequences of clustered DNA base damage. Nucl. Acids Resc 2006; 34(13), 3722-3730.

Skourti-Stathaki K, Proudfoot NJ. A double-adged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. 2014; 28(13), 1384-1396.

Valdiglesias V, Giunta S, Fenech M, Neri M, Bonassi S. γH2AX as a marker of DNA double strand breaks and genomic instability in human population studies. Mutat Res. 2013; 753(1):24-40.

Venema J, van Hoffen A, Natarajan AT, van Zeeland AA, Mullenders LH. The residual repair capacity of xeroderma pigmentosum complementation group C fibroblasts is highly specific for transcriptionally active DNA. Nucleic Acids Res. 1990; 18, 443.448.

Wäster PK, Ollinger KM. Redox-dependent translocation of p53 to mitochondria or nucleus in human melanocytes after UVA- and UVB-induced apoptosis. J Invest Dermatol. 2009 Jul;129(7):1769-81.

Ye T, Krebs AR, Choukrallah MA, Keime C, Plewniak F, Davidson I, Tora L. seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res. 2011; 39(6):e35.

Yoshihara M, Jiang L, Akatsuka S, Suyama M, Toyokuni S. Genome-wide profiling of 8-oxoguanine reveals its association with spatial positioning in nucleus. DNA Res. 2014; 21(6):603-12.

Zarakowska E, Gackowski D, Foksinski M, Olinski R2.
Are 8-oxoguanine (8-oxoGua) and 5-hydroxymethyluracil (5-hmUra) oxidatively damaged DNA bases or transcription (epigenetic) marks? Mutat Res Genet Toxicol Environ Mutagen 2014;764-765:58-63.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS.
Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9(9):R137.

8. ANNEXIS

During my PhD, I have also been involved in parallel research lines, and my work contributed to the publication of 4 articles, reported below.

1)

Ambrosio S, Di Palo G, Napolitano G, Amente S, Dellino GI, Faretta M, Pelicci PG, Lania L, Majello B. Cell cycle-dependent resolution of DNA double-strand breaks. Oncotarget. 2016 Jan 26;7(4):4949-60. doi: 10.18632/oncotarget.6644.

Cell cycle-dependent resolution of DNA double-strand breaks

Susanna Ambrosio[1], Giacomo Di Palo[2], Giuliana Napolitano[1], Stefano Amente[2], Gaetano Ivan Dellino[3,4], Mario Faretta[3], Pier Giuseppe Pelicci[3,4], Luigi Lania[2], Barbara Majello[1]

[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

[3]Department of Experimental Oncology, European Institute of Oncology, Milan, Italy

[4]Department of Oncology and Haemato-oncology, University of Milan, Italy

Correspondence to: Luigi Lania, e-mail: lania@unina.it
Barbara Majello, e-mail: majello@unina.it

ABSTRACT

DNA double strand breaks (DSBs) elicit prompt activation of DNA damage response (DDR), which arrests cell-cycle either in G_1/S or G_2/M in order to avoid entering S and M phase with damaged DNAs. Since mammalian tissues contain both proliferating and quiescent cells, there might be fundamental difference in DDR between proliferating and quiescent cells (or G_0-arrested). To investigate these differences, we studied recruitment of DSB repair factors and resolution of DNA lesions induced at site-specific DSBs in asynchronously proliferating, G_0-, or G_1-arrested cells. Strikingly, DSBs occurring in G_0 quiescent cells are not repaired and maintain a sustained activation of the p53-pathway. Conversely, re-entry into cell cycle of damaged G_0-arrested cells, occurs with a delayed clearance of DNA repair factors initially recruited to DSBs, indicating an inefficient repair when compared to DSBs induced in asynchronously proliferating or G_1-synchronized cells. Moreover, we found that initial recognition of DSBs and assembly of DSB factors is largely similar in asynchronously proliferating, G_0-, or G_1-synchronized cells. Our study thereby demonstrates that repair and resolution of DSBs is strongly dependent on the cell-cycle state.

INTRODUCTION

Eukaryotic genome is constantly being challenged by various endogenous and exogenous insults. These damaging events include crosslinks, base modifications, base mismatches, stalled replication forks, single-strand breaks (SSBs), and particularly dangerous double-strand breaks (DSBs). To deal with such dangerous insults, eukaryotes possess an array of distinct pathways to monitor and repair damaged DNA.

The initial phases of DSB recognition and recruitment of repair factors are now quite elucidated. Following DSB the MRE11/RAD50/NBS1 complex senses a DSB within seconds and then activates PI3K-like kinases ATM (ataxia-telangiectasia mutated) protein kinase, a large Ser/Thr kinase of the PI3K-like kinase family, which also includes DNA-PKcs (DNA-dependent protein kinase catalytic subunit), and ATR [1–4]. ATM then phosphorylates histone H2AX on Ser139 (named γH2AX when phosphorylated) in DSB adjacent chromatin. The primary function of γH2AX is to recruit its decoder, MDC1 (mediator of DNA damage checkpoint protein 1), which recognizes the phosphorylated Ser139 epitope on γH2AX. γH2AX-bound MDC1 recruits in turn more MRN complexes (via an interaction with NBS1) and thus initiates a positive ATM feedback loop that leads to the amplification of the γH2A.X chromatin domain [5–8]. Concomitant with the assembly of DNA repair factors at DNA lesion, DSB response activates DNA-damage checkpoints (DDR), and diffusible signaling events that can arrest cell cycle progression either in G_1 or G_2 to allow for DNA repair

Within a collaboration with Department of Experimental Oncology of IEO in Milan, I contributed to demonstrate the cell cycle-dependence of DSB repair in non-tumorigenic mammalian cells (MCF10A), with experiments of ChIP-Seq demonstrating the accumulation of γH2AX upon DSB induction in both proliferating and quiescent cells, and with a flow citometry approach showing a delayed cell cycle re-entry when quiescent cells are DSB-damaged, compared to proliferating DSB-damaged cells.

Through ChIP-sequencing I found that proliferating and G_0 arrested cells show similar enrichment of γH2AX upon induction of, suggesting that cell cycle phase does not interfere with the initial steps of DDR; however, the persistent accumulation of γH2AX at DSB sites in G_0 cells indicates that DNA repair is compromised in non-proliferating cells.

Similarly, flow citometry approach shows a different timing in cell cycle re-entry when DSB occurs in proliferating or quiescent cells. Proliferating asynchronous cells undergo a G1/s or G2/M arrest upon DSB induction, that is completely resolved within 72h allowing the cell cycle re-entry; conversely, quiescent cells show a delayed cell cycle re-entry upon DSB induction, with accumulation of G1 and G2 phase even after 72h. These data indicate that quiescent cell is not able to repair DSB with the same efficiency of proliferating cells, and that G1/S transition is required for complete damage resolution (Ambrosio et al, 2015).

2)

Amente S, Milazzo G, Sorrentino MC, Ambrosio S, Di Palo G, Lania L2, Perini G, Majello B. Lysine-specific demethylase (LSD1/KDM1A) and MYCN cooperatively repress tumor suppressor genes in neuroblastoma. Oncotarget. 2015 Jun 10;6(16):14572-83.

Lysine-specific demethylase (LSD1/KDM1A) and MYCN cooperatively repress tumor suppressor genes in neuroblastoma

Stefano Amente[1,2,*], Giorgio Milazzo[3,*], Maria Cristina Sorrentino[1], Susanna Ambrosio[1], Giacomo Di Palo[2], Luigi Lania[2], Giovanni Perini[3,4] and Barbara Majello[1]

[1] Department of Biology, University of Naples, Naples, Italy

[2] Department of Molecular Medicine and Medical Biotechnologies, University of Naples, Naples, Italy

[3] Department of Pharmacy and Biotechnology, University of Bologna, Italy

[4] CIRI Health Sciences and Technologies (HST), Bologna, Italy

[*] These authors have contributed equally to this work

Correspondence to: Barbara Majello, email: majello@unina.it

Giovanni Perini, email: giovanni.perini@unibo.it

ABSTRACT

The chromatin-modifying enzyme lysine-specific demethylase 1, KDM1A/LSD1 is involved in maintaining the undifferentiated, malignant phenotype of neuroblastoma cells and its overexpression correlated with aggressive disease, poor differentiation and infaust outcome. Here, we show that LSD1 physically binds MYCN both in vitro and in vivo and that such an interaction requires the MYCN BoxIII. We found that LSD1 co-localizes with MYCN on promoter regions of CDKN1A/p21 and Clusterin (CLU) suppressor genes and cooperates with MYCN to repress the expression of these genes. KDM1A needs to engage with MYCN in order to associate with the CDKN1A and CLU promoters. The expression of CLU and CDKN1A can be restored in MYCN-amplified cells by pharmacological inhibition of LSD1 activity or knockdown of its expression. Combined pharmacological inhibition of MYCN and LSD1 through the use of small molecule inhibitors synergistically reduces MYCN-amplified Neuroblastoma cell viability in vitro. These findings demonstrate that LSD1 is a critical co-factor of the MYCN repressive function, and suggest that combination of LSD1 and MYCN inhibitors may have strong therapeutic relevance to counteract MYCN-driven oncogenesis.

INTRODUCTION

Neuroblastoma (NB) is a pediatric tumor with poor outcome and highly refractory to therapeutic treatment. The molecular bases of NB development and progression are still poorly understood. The best-characterized genetic markers include amplification of the proto-oncogene MYCN, amplification and mutation of ALK gene and chromosomal alterations [1-7]. Classical risk factors include the age at diagnosis, MYCN amplification and stage of the disease. MYCN is a member of the MYC family (MYC, MYCN and MYCL) proteins that are basic Helix-Loop-Helix Leucine Zipper (bHLHZip) transcription factors, which forms transcriptionally active hetrodimers with another bHLHZip protein called MAX [8-9]. Dimerization with Max endows Myc with sequence specific DNA binding ability, preferentially to sites containing the E-box sequence CACGTG. Activation of MYC oncogenes simultaneously coincides with global modifications in chromatin structure and subsequent robust changes in MYC targets gene expression. MYC proteins have been found to orchestrate epigenetic alterations by recruitment of higher order chromatin complexes that activate or repress transcription.

MYC/MYCN have been found to associate with different chromatin modifying complexes and their role in transcription depends on both histones tails modifications already present at promoters of its target genes and on the biochemical composition of protein complexes that MYC can recruit in different cellular environment [9,

63

I used flow citometry approach in order to assess the effect of MYCN and LSD1 pharmacological inhibitors on Neuroblastoma cells survival. Data show that co-treatment of both drugs causes a reduction of S-phase and an increase of the sub-G1 population, suggesting the induction of apoptosis. We propose that combinatorial targeting of both MYCN and LSD1 in Neuroblastoma could be taken under consideration for a therapeutic approach.

3)

Ambrosio S, Amente S, Napolitano G, Di Palo G, Lania L, Majello B. MYC impairs resolution of site-specific DNA double-strand breaks repair. Mutat Res. 2015 Apr;774:6-13. doi: 10.1016/j.mrfmmm.2015.02.005. Epub 2015 Mar 4.

MYC impairs resolution of site-specific DNA double-strand breaks repair

CrossMark

Susanna Ambrosio [a,1], Stefano Amente [b,1], Giuliana Napolitano [a], Giacomo Di Palo [b], Luigi Lania [b,*], Barbara Majello [a,**]

[a] Department of Biology, University of Naples 'Federico II', Naples, Italy
[b] Department of Molecular Medicine and Medical Biotechnologies, University of Naples 'Federico II', Naples, Italy

ABSTRACT

Although it is established that when overexpressed, the MYC family proteins can cause DNA double-stand breaks (DSBs) and genome instability, the mechanisms involved remain unclear. MYC induced genetic instability may result from increased DNA damage and/or reduced DNA repair. Here we show that when overexpressed, MYC proteins induce a sustained DNA damage response (DDR) and reduce the wave of DSBs repair. We used a cell-based DSBs system whereby, upon induction of an inducible restriction enzyme AsiSI, hundreds of site-specific DSBs are generated across the genome to investigate the role of MYC proteins on DSB. We found that high levels of MYC do not block accumulation of γH2AX at AsiSI sites, but delay its clearance, indicating an inefficient repair, while the initial recognition of DNA damage is largely unaffected. Repair of both homologous and nonhomologous repair-prone segments, characterized by high or low levels of recruited RAD51, respectively, was delayed. Collectively, these data indicate that high levels of MYC proteins delay the resolution of DNA lesions engineered to occur in cell cultures.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Activated oncogenes generally associate with the induction of DNA damage response (DDR) [1,2]. This phenomenon named oncogene induced DNA damage (OID) is exemplified by the c-MYC proto-oncogene. The Myc proteins (MYC, MYCN and MYCL in human) are basic helix-loop-helix leucine zipper transcription factors driving a wide range of cellular responses depending on the cellular context. In most human cancers MYC expression is deregulated and/or significantly increased [3,4]. It has been clearly demonstrated that supraphysiological levels of MYC play a causative role in the onset and progression of many types of cancers. In addition to its canonical role in transcription, MYC overexpression in mammalian cells increases genomic instability [5,6]. Although it is not clear what type of physical alterations are induced at the DNA level, recent observations propose that at least two types of DNA damage can be associated with MYC. First, MYC overexpression in mammalian cells results in loss of chromo-

somal integrity associated with chromosomal aberrations [5–9]. An additional source of MYC-induced DNA damage is replication stress, a poorly understood perturbation of DNA replication. Replication stress generates aberrant DNA replication intermediates that induce the activation of DDR at sites of active DNA replication [10–12]. Such non-transcriptional Myc effects are likely crucial to its role in tumorigenesis. However, in both cases the molecular mechanisms involved remain unclear.

It is likely that MYC-induced genomic abnormalities might be generated by defects in the repair of double-strand-breaks (DSB). DSBs are one of the most challenging forms of DNA damage, which in turn if not correctly repaired, can trigger the onset and progression of cancer cells. A number of studies have been focused on DNA-damage response (DDR) mechanisms induced by chemical compounds or radiations. A limitation of this approach is that the sites of DNA damage within the genome occur random and therefore differ from cell to cell, precluding efforts to determine co-association at specific DSBs. To overcome such limitations several systems have been developed that rely on inducible restriction enzymes capable to generate unambiguously positioned sequence-specific DSBs. However, sequence-specific DSB-inducible systems that induce either a single DSB (HO, I-SceI and FokI systems) [13,14] or several DSBs in ribosomal DNA (I-PpoI system) [15] hamper the comparison of DSB repair occurring at different locations.

* Corresponding author. Tel.: +39 081 7643029.
** Corresponding author. Tel.: +39 081 679062.
E-mail addresses: lania@unina.it (L. Lania), majello@unina.it (B. Majello).
[1] These authors contributed equally to this work.

In this work, I carried out a cell cycle analysis to investigate the role of oncoprotein Myc overexpression on repair of DNA double-strand breaks in osteosarcoma cells. Data show that damaged Myc

overexpressing cells undergo a prolonged cell cycle arrest compared to cell expressing physiological level of Myc, suggesting that it could inhibits resolution of DSB, thus suppressing DNA repair.