UNIVERSITY OF NAPLES "FEDERICO II"

DOCTORATE SCHOOL IN
MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY
XXIX CYCLE

DNA repair during estrogen-induced transcription

TUTOR
Prof. Vittorio Enrico Avvedimento

CANDIDATE
Dr. Mariarosaria De Rosa

PhD COORDINATOR
Prof. Vittorio Enrico Avvedimento

ACADEMIC YEAR 2015-2016

DNA repair during estrogen-induced transcription

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

3' UTR 3´ untranslated region

5-AzadC 5-Aza-2'-deoxycytidine

5-caC 5-carboxylcytosine

5-fC 5-formylcytosine

5mC 5-methyl Cytosine

8-oxo-G 8-Oxo-2'-deoxyguanine

AP site Apurinic/Apirimidinic site

APE1 AP endonuclease 1

BCL2 B-cell lymphoma 2 gene

BER Base excision repair

C Cytosine

CF Cleavage factor

chr chromosome

CPF Cleavage and polyadenylation factor

CTD Carboxyl-Terminal Domain

DBD DNA binding domain

DNMTs DNA methyltransferases

E2 17β-estradiol

ERE Estrogen response element

ERs Estrogen receptors

FAD Flavin adenine dinucleotide

FEN1 Flap endonuclease-1

G Guanine

GG-NER global genomic NER

H3K27me3 Lysine 27 on histone H3 tri-methylated

H3K36me3 Lysine 36 on histone H3 tri-methylated

H3K4me2/me3 Lysine 4 on histone H3 di- and tri-methylated

H3K9me2/me3 Lysine 9 on histone H3 di- or tri-methylated

hmC 5-hydroxymethyl-2′-deoxycytidine

HMTs Histone methyltransferases

JmjC Jumonji C

K Lysine

LBD Ligand binding domain

LSD1/KDM1a Lysine specific demethylase 1

MPG N-Methylpurine DNA Glycosylase

mRNA messenger RNA

NER Nucleotide excision repair

OC Open complex

OGG1 8-Oxoguanine DNA glycosylase 1

PCNA Proliferation-cell-nuclear antigen

PIC preinitiation complex

Polβ DNA polymerase β

RNA pol RNA polymerases

ROS Reactive oxygen species

SAM/Ado-Met S-Adenosyl methionine

TAFs TBP-associated factors

TBP TATA-binding protein

TDG Timidine DNA Glycosylase

TET ten-eleven- translocation

TFF1 Trefoil factor 1 (pS2 gene)

TFIIA Transcription Factor for RNA polymerase II A

Topo IIβ Topoisomerase IIβ

TC-NER transcription coupled NER

TSS Transcription Start Site

UNG Uracil DNA glycosylase

XRCC1 X-ray repair cross-complementing group 1

ABSTRACT

Estrogen-induced transcription is characterized by localized histone demethylation and DNA oxidation, followed by the recruitment of repair and DNA and histone methyl-transferase enzymes at target sites. The functional link between these different proteins recruited at estrogen chromatin sites is unclear and it is unknown which steps catalyzed by these proteins are essential for productive estrogen-mediated transcription.

Here we report that specific genomic regions that synchronously recruit estrogen receptor complexed with the demethylase LSD1 and DNA repair protein such as OGG1 (a component of base excision repair) or topoisomerase IIβ. These enzymes are recruited at the estrogen regulatory regions in a precise temporal order and are essential for the assembly of the transcription initiation complex induced by estrogens. We find that DNA methyltransferase 3a (DNMT3a) couples BER and NER repair enzymes at promoter sites and stimulates estrogen-induced transcription initiation. The orderly recruitment of DNA and histone methyltransferases and repair enzymes greatly reduces the mutational burden induced by DNA oxidation associated with transcription.

1. BACKGROUND

1.1 Overview of transcription

Genomic DNA is the macromolecule that holds genetic information. It codes for several functional macromolecules, among which messenger RNA (mRNA) stands out for its role as a substrate in protein synthesis.

Transcription, as the first stage of gene expression, is the process of copying the DNA sequence of a gene to produce an RNA molecule with all the information necessary to obtain a protein. DNA is a double-stranded molecule, but only the antisense or "minus" strand of the DNA is used as template for transcription, while the non-transcribed or "plus" strand sequence is identical to the RNA transcript (the sequence only differs for the presence of Uracil instead of Thymine).

RNA polymerases (RNA pol), large multi-subunit enzymes that attach to the antisense DNA strand and catalyze the production of complementary RNA, are the main enzymes involved in transcription. Eukaryotes have three distinct nuclear enzymes, Pol I, II, and III that synthesize different classes of RNA. The Pol II transcription machinery is the most complex, with almost 60 polypeptides, but also the most used by cells, considering that it transcribes nearly all protein-coding genes and also miRNA genes.

1.1.1 A typical RNA Pol II transcription cycle

During *initiation*, the first step of transcription, RNA Pol II must recognize a promoter sequence onto DNA, separate the duplex to expose the template strand, and initiate RNA synthesis using nucleotide triphosphates. The core promoter is the location at which the RNA pol II machinery initiates transcription. To recognize the core promoter, RNA pol II requires additional factors, commonly named "general" or "basal" transcription factors, which include TFIIA (Transcription Factor for RNA polymerase II A), TFIIB, TFIID, TFIIE, TFIIF, and TFIIH (Juven-Gershon et al., 2010).

Pol II is positioned at the core promoter by a combination of TFIID, TFIIA, and TFIIB to form the preinitiation complex (PIC). Many polymerase II promoters have a consensus sequence TATAA (TATA box) located 25 to 30 nucleotides upstream of the transcription start site. This sequence is recognized by transcription factor

TFIID, which consists of the TATA-binding protein (TBP) and TBP-associated factors (TAFs). TFIIB connects TBP and Pol II, associating with TFIIF to form, finally, a complex with TFIIE and TFIIH (Horn et al., 2016). TFIIH then melts 11–15 bp of DNA, using energy provided from ATP hydrolysis. Hence, the single-strand template stands in the Pol II cleft and the open complex (OC) is assembled to initiate RNA synthesis (Grünberg et al., 2013; Li et al., 2007). Once short abortive RNA molecules are produced, after the synthesis of about 30 bases of RNA, Pol II releases its contacts with the core promoter and the rest of the transcription machinery to reach the stage of transcription *elongation*. In this instance, all the TFs are released from PIC except for TFIIF. Moreover, the Carboxyl-Terminal Domain (CTD) of the Pol II largest subunit is phosphorylated, fulfilling a critical role for elongation. Then, elongating Pol II can recruit all factors needed to obtain productive RNA chain synthesis, RNA processing, RNA export, and chromatin modification (Hahn, 2004). Meanwhile, all the unemployed factors that are not useful anymore for the elongation stage, are recycled to start a new initiation of transcription in the so-called *Scaffold Complex*. This complex assumes an advantageous role because it promotes the *Reinitiation* of transcription, bypassing the slow recruitment step of all general factors (Figure 1).

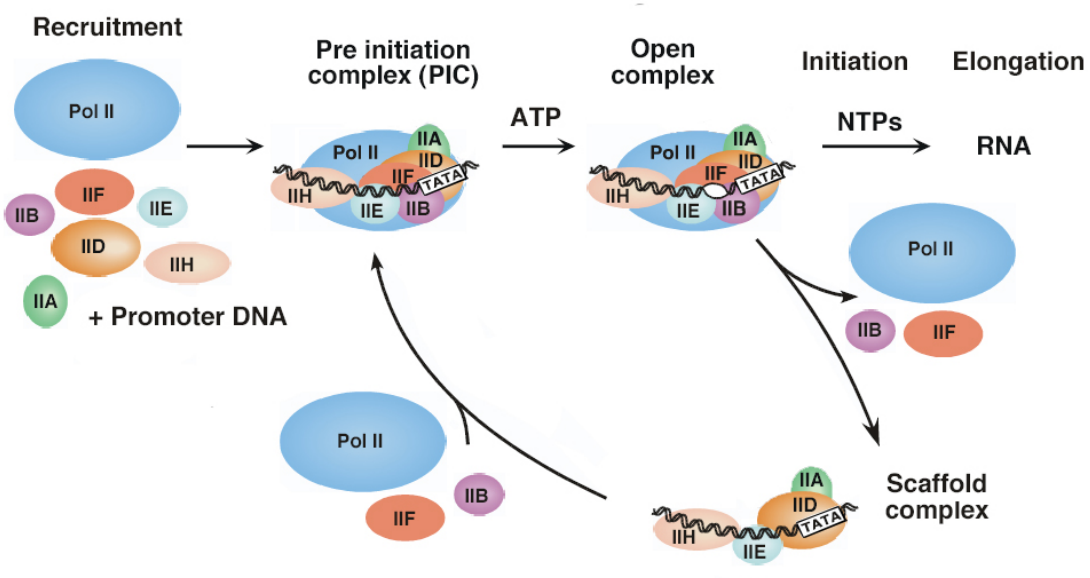


Figure 1. Structure and mechanism of the RNA Polymerase II transcription machinery

General TFs and RNA Pol II are recruited on the promoter region to assemble the Pre Initiation Complex (PIC); once the *transcription bubble* is open and some abortive synthesis cycles are performed, the process can

progress in the elongation stage. The active recycling of the general TFs allows them to maintain productive transcription cycles (Adapted from Hahn S, 2004).

Transcription *termination* occurs when the polymerase and the nascent RNA are released from the DNA template: this event is due to co-operation between components of the cleavage and polyadenylation factor (CPF) and cleavage factor (CF) complexes, which recognize specific sequences in the 3´ untranslated region (UTR) of the transcript (Porrua O et al., 2015).

1.2 Epigenetics

1.2.1 Chromatin organization

Chromosomal DNA achieves a high degree of compaction into the microscopic space of the eukaryotic nucleus, thanks to its packaging into chromatin fibers. Chromatin is made of a basic repeating unit, called the nucleosome, that consists of 146 bp of DNA wrapped around an octamer of histones (Virani S et al., 2012). Two copies of each histones H2A, H2B, H3 and H4, combined in a H3-H4 tetramer and two H2A-H2B dimers, form an octamer. Histones are small basic proteins, highly conserved throughout evolution, constituted of a globular domain flanked by a carboxyl-terminal domain and an amino-terminal tail rich in lysine residues that protrudes out of the nucleosome. The H3 and H4 N-terminal tails confer a higher-order to the nucleosome structure through their bond with DNA or the acidic stretch on neighboring nucleosomes (Li Z et al., 2016). Chromatin is compartmentalized in distinct domains: the term *heterochromatin* is attributed to a transcriptionally silent structure with thickly packed nucleosomes; *euchromatin*, instead, is more lightly packaged and can harbor the transcriptional machinery (Figure 2) (Bártová E et al., 2008).

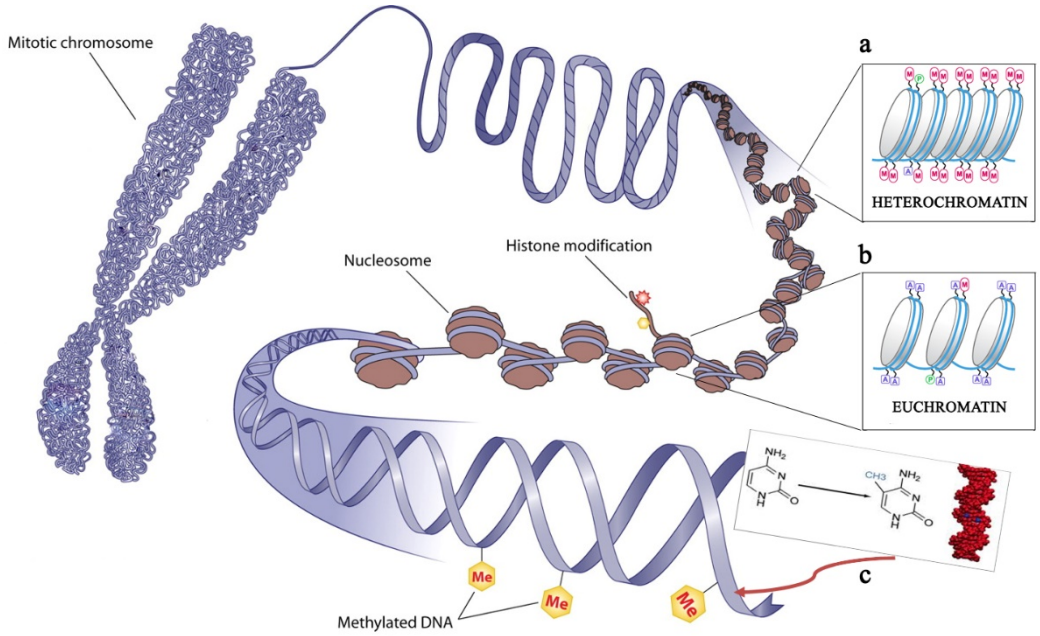


Figure 2. Overview of chromosomal DNA organization

a) Heterochromatin: regions with low or no transcriptional activity that are densely packed; **b)** Euchromatin: regions with high transcriptional activity that are loosely packed; **c)** naked DNA, which can be methylated at cytosine level. In a) and b) panels, the markings represent different modifications on histones: red – methyl groups, purple – acetyl groups, green – phosphoryl groups (Adapted from Zaidi SK et al., 2010).

Specific modifications, which take place at the amino-terminal histone tails, are responsible for the segregation of the genome into distinct domains and they influence the transcription availability of genes.

1.2.2 Epigenetic modifications

Covalent modifications of DNA and histone proteins determine variations in the degree of chromatin condensation. Several types of post-translational modification can affect the protruding histone tails: methylation, acetylation, phosphorylation, ADP-ribosylation and ubiquitination. All of them can affect interactions between DNA and histones, altering gene transcription (by regulating the binding of transcription factors to DNA sequences), DNA replication, DNA repair, chromosome organization and disease processes (Hamilton JP, 2011). These modifications create the so-called *histone code*, an epigenetic marking system read by specific proteins, which subsequently change the structure of chromatin, governing gene expression. Indeed, expression of genes in mammals is controlled by genetic as well as epigenetic mechanisms: the term *epigenetics* specifies the study of heritable changes in the phenotype that are not encoded in the genome of a

cell (Jin B et al., 2011). Epigenetic alterations encompass histone and other chromatin protein modifications and DNA methylation.

- **Histone methylation and de-methylation**

  Among the major posttranslational histone modifications, are the methylation of certain lysine (K) and arginine residues of histones H3 and H4, managed by histone methyltransferases (HMTs). Lysines can be modified by mono-, di- and tri-methylation, whereas arginines can be mono- and di-methylated (Yi X et al., 2015). S-Adenosyl methionine (SAM/Ado-Met) provides the methyl groups to HMTs. Lysine HMTs are extremely specific: their modifications take place on one single lysine on a single histone but the resulting effect can be either activation or repression of transcription. Generally, condensed heterochromatin is characterized by repressive marks, like di- and tri-methylation of lysine 9 on histone H3 (H3K9me2/me3) or tri-methylation of lysine 27 on histone H3 (H3K27me3). Histone H3 di- and tri-methylation of lysine 4 (H3K4me2/me3) and tri-methylation of lysine 36 (H3K36me3) are instead associated with transcription activation (Kouzarides T, 2007). Conversely to other modifications (like lysine acetylation, which abrogate the positive charge of the amino acid and eliminate the electrostatic bond between DNA and histones), methylation affects the binding of chromatin-associated proteins. Moreover, different readers can be recruited when the methylation is reversed: methyl groups can be removed from histone lysine residues by enzymes called *histone lysine demethylases*. These are the jumonji C (JmjC)-domain-containing, iron-dependent dioxygenases and the amine oxidases, such as the lysine-specific histone demethylase 1 (LSD1) also known as KDM1 (Morera L et al., 2016). Post-translational histone modifications can clearly influence each other by either enhancing or inhibiting transcription, therefore an hypothesis arose that they compose the so-called *combinatorial code*, regulating and determining some phenotypic traits by the recruitment of different chromatin-modifying proteins (Figure 3) (Cieślik M et al., 2014).

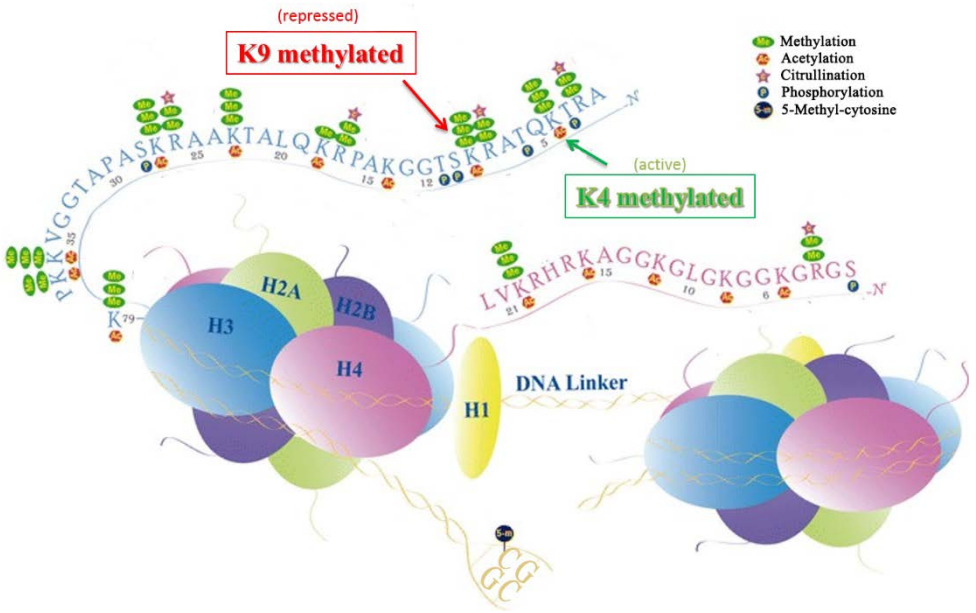


Figure 3. The *histone code*

DNA is wrapped around an octamer of histones H2A, H2B, H3 and H4, combined in two H2A-H2B dimers and a H3-H4 tetramer. The protruding amino-terminal histone tails can be differently modified: relevant is the methylation of histones, which can result in repression or activation of transcription, depending on the proteins involved in "writing" and "reading" the modification (Adapted from Morera L et al., 2016).

- **DNA methylation**

  DNA methylation consists in the covalent attachment of a methyl group at the 5'-carbon position of the cytosine (C) residues. In mammals, DNA is methylated primarily on the C of CpG dinucleotides in the DNA chain. Methylation of C which are not present in CpG dinucleotides, is present only in the body of actively transcribed genes of stem cells. Even if 60–80% of the CG residues appear methylated throughout the entire genome, it is noteworthy that in CpG islands and active regulatory regions only 10% of the CGs are methylated. Active promoters must be protected from methylation, while other genomic regions, like repetitive DNA sequences randomly dispersed or clustered near centromeres, must be repressed by methylation to maintain genome integrity (Du J et al., 2015). DNA methylation is considered a key player in epigenetic silencing of transcription due to the stable feature of its repressive mark and the ability to crosstalk with chromatin modification status (Jin B et al., 2011). Three DNA methyltransferases (DNMTs) are the enzymes responsible of the methyl-

group transfer from SAM to cytosine: DNMT1 accounts for the *maintenance* methylation, copying pre-existing methylation patterns to a newly synthesized strand after recognition of hemi-methylated sites; DNMT3a and DNMT3b are the *de novo* methyltransferases effective on unmethylated DNA (Castillo-Aguilera O et al., 2017). Though DNMT3b has approximately 30 isoforms and DNMT3a has two different isoforms, they share common features (Figure 4): a variable region at the N terminus, followed by a PWWP (Pro-Trp-Trp-Pro) motif most likely involved in nonspecific DNA binding, a Cys-rich 3-Zn-binding domain and a well-conserved C-terminal domain bearing the catalytic activity (Cheng X et al., 2008).

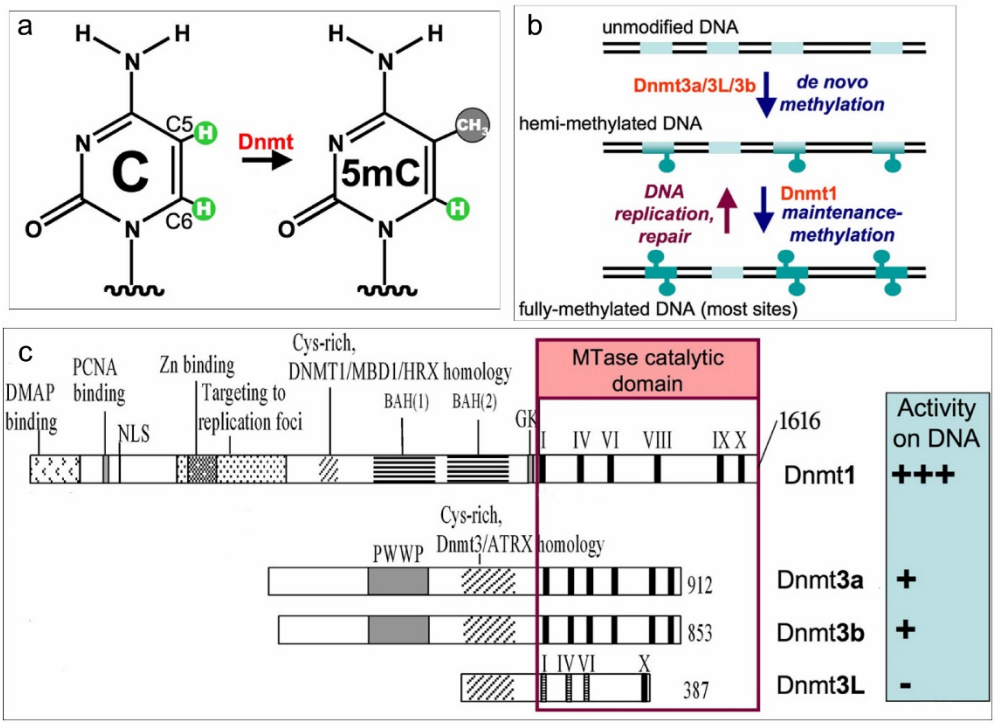


Figure 4. DNA Methylation and DNMTs domains

(a) Methylation of DNA at ring carbon C5 of cytosine. **(b)** De novo versus maintenance methylation. The pale-blue sections are substrate sequences (usually CpG), and the turquoise circles represent methyl groups on the cytosines. After replication or repair, the duplex is only hemi-methylated. **(c)** DNMTs family members: schematic representation of DNMT1 and DNMT3a/b/L. Roman numerals indicate conserved motifs of DNA MTases: DNMT3L lacks the conserved residues required for DNA methyltransferase activity in the C-terminal domain (Adapted from *Xiaodong Cheng, Robert M. Blumenthal. 2008 Structure*).

As a stable repressive mark, DNA methylation is considered a key player in epigenetic silencing of genes with high CpG dinucleotide frequency near their promoters (Bird AP et al., 1999); its function is also recognized in imprinting (Li E et al., 1993), in development and differentiation (Li E et al., 1992) and in X chromosome inactivation (Jaenisch R et al., 1998). Aberrant methylation has been detected in cancer cells (Cho YH et al., 2010); moreover, the association of DNA methylation with DNA damage and repair was demonstrated (Cuozzo C et al., 2007), as was the decreased methylation rate of repaired DNA regions caused by transcription (Morano A et al., 2014).

DNA methylation can be reverted trough *passive* or *active* demethylation: the passive mechanism takes place during replication and involves DNMT1, which does not methylate the newly synthesized DNA strands. Oxidation of 5-methyl cytosine (5mC) by enzymes from the ten-eleven- translocation (TET) family, instead, is the main active mechanism of DNA demethylation: these hydroxylases convert 5mC successively into 5-hydroxymethyl-2′-deoxycytidine (hmC), 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC), using ATP energy to restore the unmethylated C (Tahiliani M et al., 2009).

1.3 Regulation of transcription by estrogens

1.3.1 Estrogen receptors and their ligands

Estrogen receptors (ERs) are nuclear hormone receptors ligand-dependent. ERα and ERβ are two alternative isoforms, expressed in mammals as products of separate genes located on different chromosomes. Classical transcription factors do not need activation with a specific ligand to exert their function; conversely, to recognize specific sequences within the promoters of target genes, ERs need to bind estrogens (Farooq A, 2015). Estrogens are steroid hormones derived from cholesterol, able to diffuse through the membrane and bind to their receptors in the nucleus thanks to their lipophilic character. Together with its receptors, 17β-estradiol (E2) mediates a plethora of cellular functions from sexual development and reproduction to

regulation of metabolic processes. ERα and ERβ are members of a family of ligand-modulated transcription factors and they share a similar structure (Figure 5):

- the N-terminal A/B domain (AF1), able to transactivate transcription in a hormone-independent manner;
- the DNA binding domain (DBD), highly homologous in ERα and ERβ, that recognizes specific DNA sequences named Estrogen Response Elements (ERE);
- the hinge domain, important for the translocation of ER into the nucleus;
- the ligand binding domain (LBD), located at the C-terminus, which contains a ligand recognition pocket and also a second transactivation domain (AF2) (Nilsson S et al., 2001).

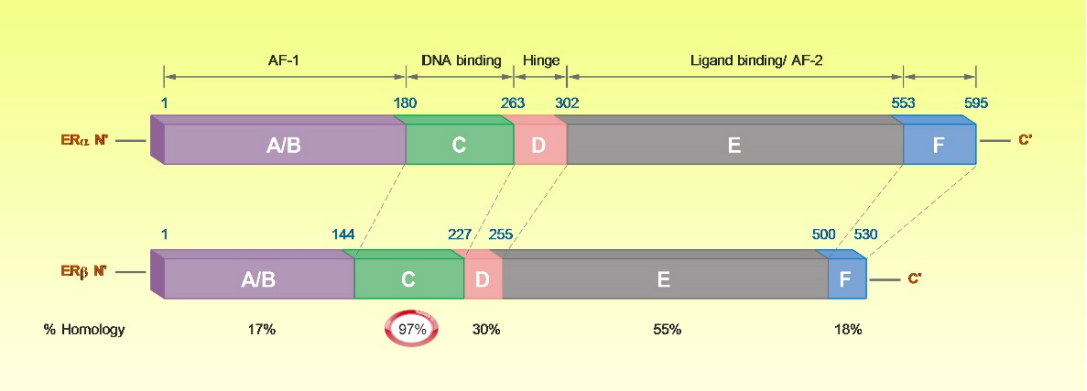


Figure 5. ERα and ERβ domain structures

As explained in the text, ERα and ERβ have a similar structure; however, the DNA binding domain is the most conserved and shares 97% homology between the two receptors.

When the receptor binds the hormone, it undergoes a conformational change resulting in dimerization. Then, the dimer enters the nucleus and, thanks to the presence of two C4-type zinc fingers in the DBD, it binds to the palindromic sequences known as ERE, modulating transcription of multiple responsive genes (Figure 6) (Deroo BJ et al., 2006).

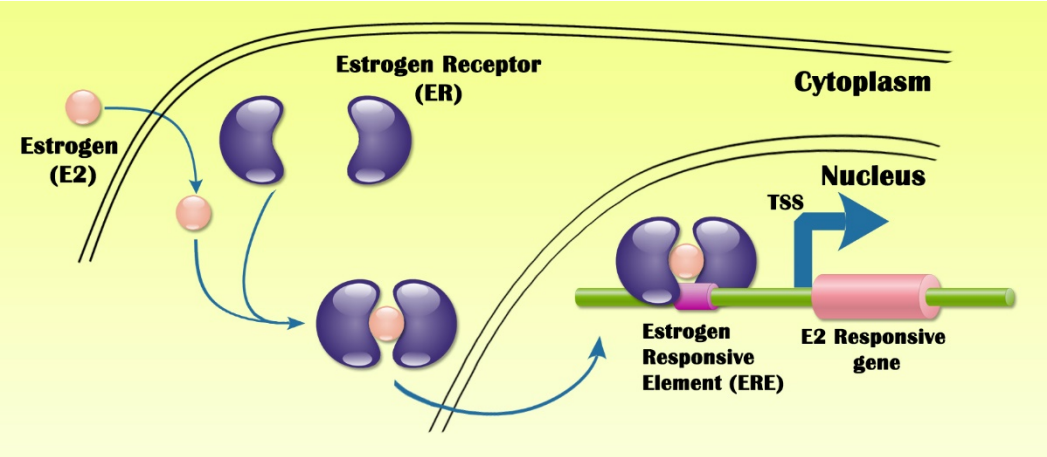


Figure 6. Estrogen-induced transcription

Schematic representation of E2-induced transcription: extracellular estrogen enters the cell and binds to ER monomers, causing dimerization and transport into the nucleus as a receptor-ligand complex. Once recognized the ERE, transcription is activated: the RNA Pol II in its PIC form (here not shown) can start RNA polymerization from the Transcription Start Site (TSS) (Image re-edited from *http://www.ifom-ieo-campus.it/research/petersen-mahrt.php*).

The whole-genome mapping of ERα binding sites showed that only 5% of the receptors are enriched within 50 kb around the transcription start sites (TSS). The majority of DNA binding sites are located in intronic or distal regions, suggesting transcriptional mechanisms that involve physically distant regions via chromatin looping (Lin CY et al., 2007). Moreover, it was shown that ERs collaborate with other transcription factors recruited at ERE sequences (Carrol JS et al., 2006).

1.3.2 Estrogen-induced genes

- **TFF1 gene**

  Trefoil factor 1 (TFF1), generically named *pS2* gene, is encoded by the TFF1 gene. It is located in a cluster on chromosome 21, together with two other members of the trefoil gene family. For several years, little information was known about the function of the trefoil proteins, apart from their increased expression in chronic inflammatory intestinal diseases (Wong WM et al., 1999). pS2 was first detected in breast cancer cell lines, although it is mainly expressed in gastrointestinal mucosa and is considered a gastric tumor suppressor. The pS2 gene product protects mucosa from insults, influences healing of the epithelium and stabilizes the mucus layer (Aihara E et al., 2016). Like the other members of the trefoil family, pS2 protein bears

a conserved 40-amino acid domain called *trefoil motif*, which contains three conserved disulfides. An ERE site in the 5'-flanking region of the gene enables the transcriptional regulation of pS2 by estrogen (Figure 7).

Figure 7. Structure of TFF1 gene (pS2)

pS2 gene contains three exons and an ERE in the 5'-flanking region of the promoter (legend: in green are represented the non-coding regions and introns; in red is shown the promoter region; in light pink are represented the exons; in dark pink can be identified the EREs; in blue are depicted PolyA regions)

- **BCL2 gene**

B-cell lymphoma 2 (BCL2) gene, localized in humans on chromosome 18, encodes an oncogenic and anti-apoptotic protein expressed in the outer mitochondrial membrane. BCl-2 protein modulates membrane permeability, thereby affecting mitochondria-dependent apoptosis. Two EREs, only 81 bp apart and located within the BCL2 coding region, at the end of the second exon (Figure 8), are responsible for the gene responsiveness to estrogen (Perillo B et al., 2000).

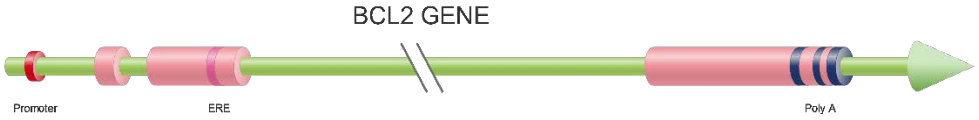


Figure 8. Structure of BCL2 gene

BCL2 gene is composed by three exons and a ERE is present in the second exon (legend: as in figure 7).

- **CAV1 gene**

The gene that encodes Caveolin-1 protein is *CAV1*, located on chromosome 7, near the *CAV2* gene. Both genes code for proteins that act as a hetero-oligomeric complex to constitute the plasma-membrane structure known as *caveolae*. E2 induces the expression of caveolin-1 mRNA (Park JH et al., 2009), through the recognition of an ERE located in the first intron (Figure 9).

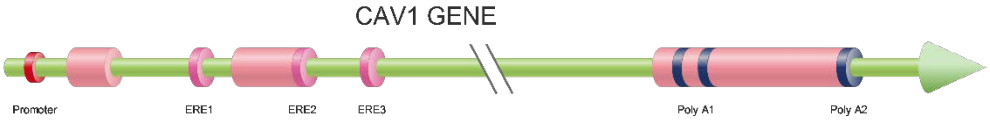


Figure 8. Structure of CAV1 gene

CAV1 gene is formed by three exons and three EREs spanning from the first to the second intron; the ERE recognized by ERs is located in the first intron (legend: as in figure 7).

1.4 DNA oxidation, repair and transcription: what is the connection?

1.4.1 DNA oxidative damage and repair

DNA is vulnerable to several kinds of damage, among which can be listed the oxidative damage caused by reactive oxygen species (ROS). As a consequence of normal metabolic processes and also interaction with radiations or toxic cancer-promoting substances. Either purines or pyrimidines can be oxidized, but the most commonly damaged base is the guanine, transformed in 8-oxo-2'-deoxyguanine (8-oxo-G). The ability of 8-oxo-G to base pair with deoxyadenosine instead of deoxycytidine explains why, if not properly repaired, this lesion can lead to G·C→T·A transversions, which can result in harmful point mutations. Generally, DNA modifications generated by environmental agents or endogenous sources of oxidative stress are recognized and removed by enzymes of the base excision repair (BER) pathway, while specific oxidative lesions, as cyclo-purines, are detected and repaired by nucleotide excision repair (NER) proteins.

- **Base excision repair**

  The BER pathway comprises numerous proteins that act in concert to eliminate non-bulky DNA lesions. An exception is the recently discovered NEIL3, able to unhook DNA inter-strand crosslinks during replication (Semlow DR et al., 2016). Generally, a single damaged base is recognized and excised by a DNA glycosylase, which forms an apurinic or apyrimidinic (AP) site. Subsequently, an AP endonuclease recognize the apurinic site in the DNA strand and cleaves at the 5´-side of the abasic site. At this stage, the newly formed single-strand break in the DNA can be resolved by either

short-patch or *long-patch BER*, with a DNA polymerase and a DNA ligase involved in the completion of the BER process (Figure 9) (Dyrkheeva NS at al., 2016).

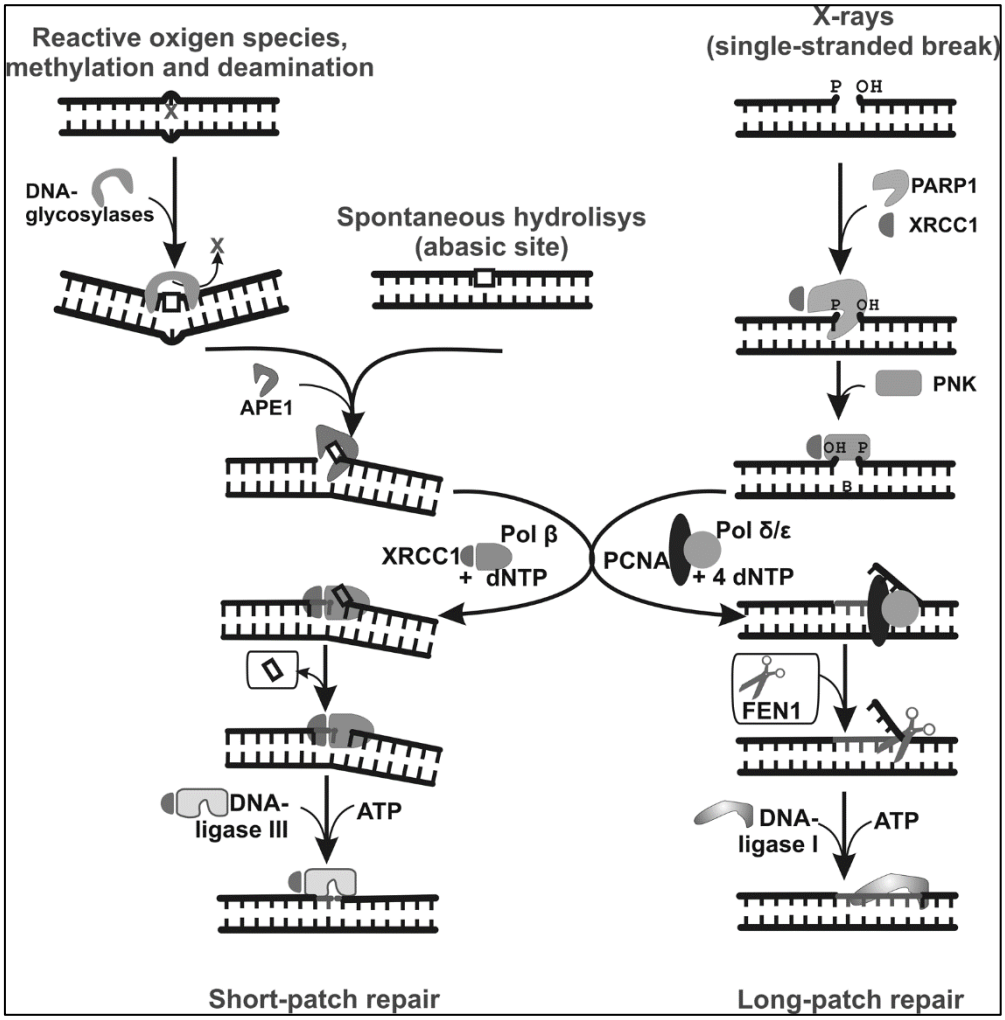


Figure 9. Short-patch and Long-patch BER

General scheme of the base excision repair (see the text for full details. Adapted from Dyrkheeva NS at al., 2016)

Oxidized bases are processed by different DNA glycosylases: 8-Oxoguanine DNA glycosylase I (OGG1) excises 8-oxo-G, thymine-DNA-glycosylase (TDG) removes modified cytosines, uracil-DNA-glycosylase (UNG) avoids C→T point mutations, eliminating the uracil arisen from cytosine deamination, and N-methylpurine-DNA glycosylase (MPG) excises modified purines. The 8-oxo-G base is a relevant example to understand the harmful feature of endogenously generated DNA modifications. For example, OGG1 is continuously active to maintain the levels of 8-oxo-G below 1 per million base pairs in physiological conditions (ESCODD, 2003);

OGG1-/- mice show modestly increased levels of 8-oxo-G in various organs and increased spontaneous mutation rates responsible for malignant transformation (Klungland A et al., 1999). Moreover, as mitochondria are the sites of oxidative phosphorylation, BER is essential also to preserve mitochondrial DNA, continuously exposed to endogenous oxidative damage (Caston RA and Demple B, 2016).

As depicted in Figure 9, in the *short patch BER*, the AP endonuclease 1 (APE1) recognizes the abasic site left by the glycosylase and creates a nick in the phosphodiester backbone of the DNA. Then, the DNA polymerase β (Polβ) uses its associated AP lyase activity to remove the deoxyribose-5-phosphate residue and its DNA polymerase activity to insert the correct nucleotide. For this purpose, Polβ is first recruited by Ape1, which remains bound to the cleaved AP site (Bennet et al., 1997), and is further helped to bind with DNA Ligase III by the scaffold protein X-ray repair cross-complementing group 1 (XRCC1) (Lindahl T and Wood RD, 1999).

The *long-patch BER* intervenes in the repair of nucleotide strands (starting from a minimum length of 2 nucleotides). The proliferation-cell-nuclear antigen (PCNA) acts as a scaffold protein for the repair enzymes, while the DNA polymerases Pol δ and Pol ε produce an oligonucleotide flap, further removed by flap endonuclease-1 (FEN1) (Klungland A and Lindahl T, 1997) The repair is then completed by DNA ligase I, which ligates the oligonucleotide and seals the break (Sung JS and Demple B, 2006).

- **Nucleotide excision repair**

  The NER pathway repairs damages long at least 2 nucleotides and responsible of a structural distortion of the DNA helix: these type of single strand breaks are mostly caused by exogenous insults, such as UV radiation and bulky DNA adducts (Balajee AS and Bohr VA, 2000). Mammals express 9 principal proteins involved in NER pathway. Deficiencies in some of these proteins lead to specific diseases and the specific protein names are associated with the disease. XPA, XPB, XPC, XPD, XPE, XPF, and XPG all derive from Xeroderma Pigmentosum and CSA and CSB are linked to Cockayne syndrome. Additional proteins such as RPA, RAD23A, ERCC1, and RAD23B also participate in nucleotide excision repair.

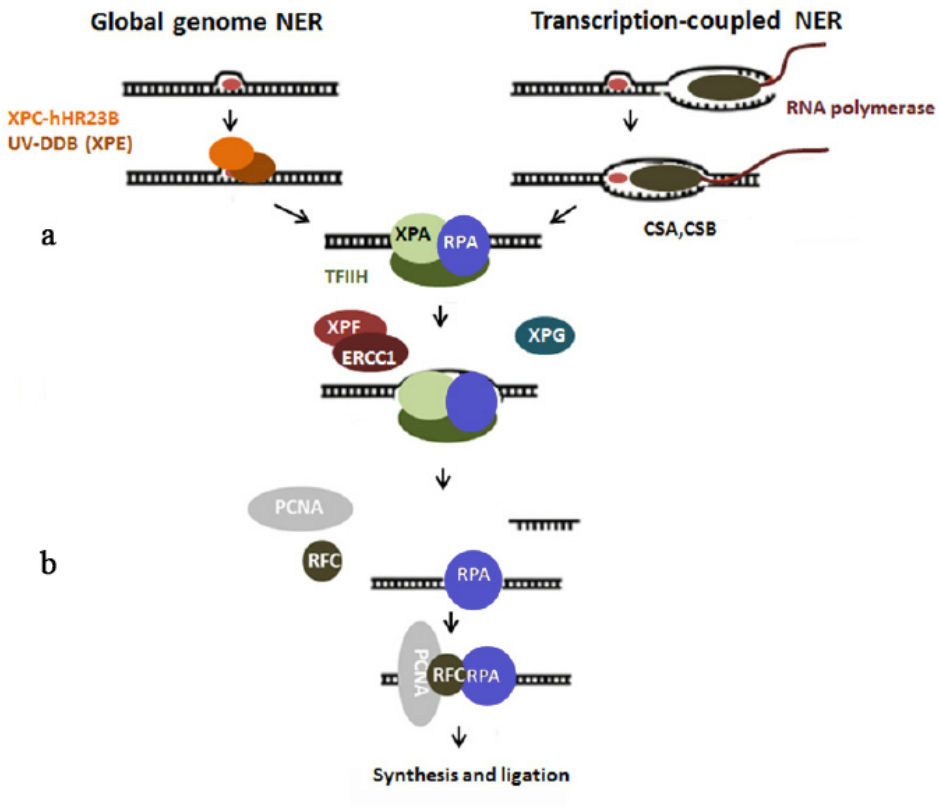


Figure10: A model of the nucleotide excision repair pathway, which includes global genomic repair and transcription-coupled repair **(a)** that both share a common point from the unwinding of the DNA **(b)** to the resolution of the damage and the repair of DNA (Adapted from Fousteri et al., 2013).

Eukaryotic NER is characterized by the presence of sub-pathways: global genomic NER (GG-NER) and transcription coupled NER (TC-NER) (Figure 10). Two different sets of proteins are involved in recognizing DNA damage for each sub-pathway. After damage recognition, the two pathways converge to the steps of dual incision, repair, and ligation. GG-NER is a process independent of transcription that repairs damage throughout the genome in both transcribed and un-transcribed genes. Some of the proteins involved in this pathway are the DNA-damage binding (DDB) and XPC-Rad23B complexes that constantly scrutinize the genome and recognize helix distortions: the XPC-Rad23B complex recognizes distortions, while DDB1 and DDB2 (XPE) are able to sense damages caused by UV light.

TC-NER is required, instead, when RNA pol is blocked at a lesion onto DNA: in this case, the detection of a helix distortion by the XPC-RAD23B

and DDB complexes is not necessary, because the stalled RNA pol serves as a damage signal. CSA and CSB proteins bind this type of DNA damaged structure instead of XPC-Rad23B. After the identification of the damaged site, several repair proteins are recruited to excise and fill in the damaged DNA surrounding the lesion. NER proteins repair the non-transcribed strands of transcriptionally active genes faster than non-transcribed strands and transcriptionally silent DNA.

1.4.2 DNA oxidation in the transcriptional activation of hormone receptors

DNA oxidation can result from transcription. This conclusion emerges from findings in different systems, in which the common denominator is the site-directed formation of oxidized bases in promoter regions. Perillo et al., in 2008 showed that the Lysine Specific Demethylase 1 (LSD1 or KDM1a) generates 8-oxo-G after estrogen treatment of MCF7 cells (ERα positive, thus estrogen-responsive cells). LSD1 is a flavin adenine dinucleotide (FAD)-containing enzyme, involved in the demethylation of mono- and dimethyl-lysine residues of histones. During its enzymatic activity, LSD1 generates H_2O_2 as a stoichiometric derivative product, causing the oxidation of guanines in the DNA (Culhane JC and Cole PA, 2007; Forneris F et al., 2005). It was shown that H3K9me2 demethylation by LSD1 triggers E2-induced transcription thanks to the recruitment of OGG1 to the 8-oxo-G formed in the promoter region of estrogen-responsive genes: the oxidized bases are then excised by OGG1 and Ape1 processes the resulting apurinic site leaving a transient nick in the DNA backbone. Finally, the action of Topoisomerase IIβ (Topo IIβ) and a DNA ligase reseals DNA. This process causes in the chromatin dynamic topological changes required for a productive transcription activation (Figure 11).

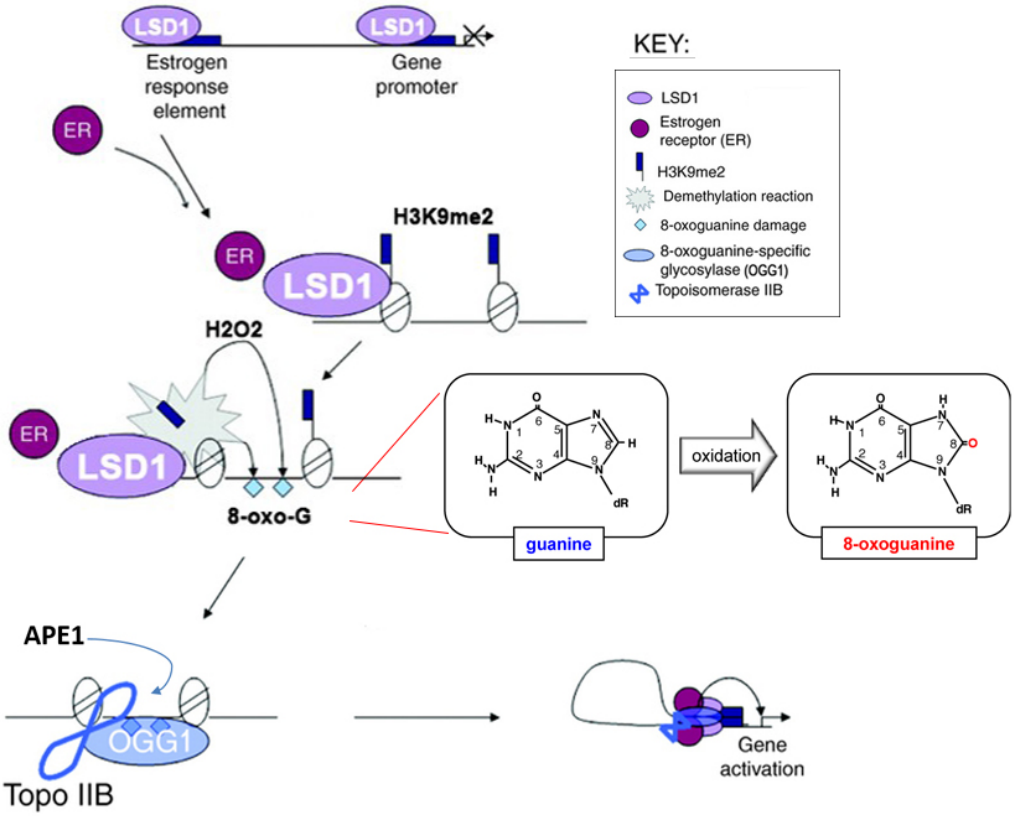


Figure 11. Model for regulation of E2-induced transcription by histone demethylation and oxidation of DNA.

LSD1 is constitutively associated with EREs and promoters of ER target genes. When the ER is activated by the ligand, it binds to both the enhancer and promoter, and drives LSD1-dependent demethylation of H3K9me2, production of H_2O_2 and oxidation of guanine (here depicted in details). This damage is recognized and repaired by OGG1 and Topo IIβ, which in turn triggers chromatin bending to promote gene transcription (Adapted from Nottke A et al., 2009 and Perillo et al., 2008).

An analogous mechanism of LSD1-dependent recruitment and activation of OGG1 has been shown in a Myc-induced transcription system, where the triggering modification is the transient demethylation of H3K4me2 (Amente et al., 2010). Moreover, H3K4me2 and H3K9 me2/me3 demethylation occurs in retinoic acid-induced transcription (Zuchegna et al., 2014). These epigenetic modifications have been found crucial also in an androgen receptor-mediated transcriptional system, where the demethylation of H3K4me2 followed by DNA oxidation and recruitment of involved proteins, activates the transcription of androgen target genes (Yang S et al., 2015). It is noteworthy that blocking DNA oxidation with antioxidants such as N-acetylcysteine lowers the induction of target genes in all these systems.

2. AIM OF THE STUDY

Estrogen-induced transcription of many genes occurs following demethylation of H3K9me2, performed by the demethylase LSD1. H_2O_2 released by LSD1 causes local oxidation of DNA, measured as 8-oxo-G formation. BER enzymes recognize and resolve oxidized guanines with the intervention of Topo IIβ and a DNA ligase that reseals DNA and allows chromatin bending and productive transcription (Perillo et al., 2008; Amente et al., 2010). The production and repair of 8-oxo-G is necessary to induce the transcription of estrogen target genes, and blocking DNA oxidation leads to impaired mRNA production (Perillo et al., 2008). It was already known that OGG1 is recruited to regulatory regions of estrogen-induced genes and recently, it has been reported that also GG-NER enzymes are found at inducible promoters after hormonal stimuli, participating in oxidative damage repair (Zuchegna et al., 2014). Moreover, it has been reported that CpG methylation cycles are induced within the promoter of estrogen target genes, together with the recruitment of DNMTs enzymes (Metivier et al., 2003).

The aim of this study is to analyze the action of estrogens on the coordinate recruitment of DNA repair enzymes (BER – NER) and a DNA methyltransferase enzyme (DNMT3a) during the assembly of the transcription initiation complex

3. MATERIALS AND METHODS

3.1 Cell cultures and treatments

Human breast cancer MCF-7 cells were grown at 37°C in 5% CO2 in Dulbecco's modified Eagle's medium (DMEM) supplemented with phenol red, L-glutamine (2 mM), insulin (10 μg/ml), hydrocortisone (3.75 ng/ml), and 10% fetal bovine serum (FBS, South America origin, Brazil, Invitrogen, Rockville, MD, USA). To evaluate the effect of Estrogen treatment, cells were grown in phenol red-free DMEM containing 10% dextran – charcoal-stripped FBS for 72 hours, before being challenged with 50 nM Estrogen for different times according to the experimental set up.

3.2 Transfections and silencing

All transfections were carried out using a Neon® Transfection System. Specific plasmids or siRNAs were introduced into each 3x10^6 dissociated cells in 100 ml volume according to the manufacturer's instructions. Pulse width was determined according to applied voltages: 1100V, 30 ms, 2 pulses. Electroporated cells were then seeded into 100-mm culture dishes containing 5ml of culture media. After 48 h, cells were treated with E2 (E8875 Sigma-Aldrich) at the times and concentrations indicated in the text.
The plasmids and siRNA used were the following: DNMT3a1-282C catalytic subunit expressing plasmid from TopoGEN (2881 – Lot#091112UCF); OGG1 wt and K341R plasmids (AddGene); DNMT3a siRNA (Santa Cruz sc-270087).

3.3 RNA extraction and qRT-PCR

Total RNA was extracted using Trizol (Invitrogen). Samples were quantified with NanoDrop 2000c (Thermo Fisher Scientific, Life technologies). cDNA was synthesized in a 20 μl reaction volume containing 1 μg of total RNA using a high-capacity reverse transcriptase kit (SensFAST cDNA Synthesis Kit, Bioline - BIO-65053) according to manufacturer's instructions. The mRNA expression level (50 ng for each cDNA) of the analyzed genes was measured by a StepOnePlus Real Time PCR System (Applied Biosystem) on DNA template (RT-PCR) using the SYBR Green-detection system (Biorad) according to

manufacturer's instructions. The complete list of oligonucleotides used is reported in Table 1.

3.4 Chromatin Immuno-Precipitation (ChIP) and Re-ChIP

Cells were transfected and/or treated as indicated in the legends of the figures. The cells (~2.5 x 10^6 for each antibody) were fixed for 10 min at room temperature by adding 37% formaldehyde (Sigma F8775) to a final concentration of 1%, the reaction was quenched by the addition of glycine to a final concentration of 125 mM. Fixed cells were harvested and the pellet was resuspended in 1 ml of Lysis Buffer (10 mM Tris- HCl pH 8.0, 10 mM NaCl, 0.2 % NP40) containing 1X protease inhibitor cocktail (Roche Applied Science). The lysates were sonicated in order to have DNA fragments from 300 to 600 bp. Sonicated samples were centrifuged and supernatants diluted 2 fold in the ChIP Buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris- HCl pH 8.0). An aliquot (1/10) of sheared chromatin was further treated with proteinase K (4U every 1 x 106 nuclei), extracted with 1 volume of phenol/chloroform/isoamyl alcohol (25:24:1) and precipitated in LiCl 0,4 M/ ethanol 75% to determine DNA concentration and shearing efficiency (input DNA). The ChIP reaction was set up according to the manufacturer's instructions. Briefly, the sheared chromatin was precleared for 2 h with 1 μg of non-immune IgG (Santa Cruz Biotechnology, Santa Cruz, CA, USA) and 20 μl of Protein A/G PLUS-Agarose (Santa Cruz Biotechnology) saturated with salmon sperm DNA (1 mg/ml). Precleared chromatin was divided into aliquots and incubated at 4 °C for 16 h with 1 μg of the specific antibody (for the codes, see below) and non-immune IgG respectively. The immuno-complexes were recovered by incubation for 3 h at 4 °C with 20 μl of protein-A/G PLUS agarose, beads were washed with wash buffers according to the manufacturer's instructions and immunoprecipitated DNA was recovered through phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation and redissolved in TE buffer (10 mM Tris-HCl, 1mM EDTA, pH 8,0). Samples were subjected to qPCR using the primers indicated in the legend of the specific figures; primers sequences are reported in Table 1. Real Time-qPCRs were performed using iTaq Universal SYBR Green Supermix (Biorad) with

cycle conditions as follows: 95 °C 10 min; 40x (95 °C 10 sec, 54 °C 30 sec, 72 °C 30 sec); 72 °C 10 min. Re-ChIP is a procedure in which formaldehyde-crosslinked, protein–DNA complexes are subjected to two sequential immunoprecipitations with antibodies of different specificity. This procedure is used to address whether two proteins can simultaneously co-occupy a stretch of DNA in vivo (Geisberg JV et al., 2004). Part of the immunoprecipitated samples was used to perform a second immunoprecipitation adding the specific antibodies; the procedure was followed identically as explained above for conventional ChIP experiments.

3.5 Human promoter microarray

Affymetrix array, human promoter 1.0 R array. The Human Promoter 1.0R Array is a single array comprising over 4.6 million probes tiled through 25,500 human promoter regions. Sequences used in the Array were selected from NCBI human genome assembly (Build 34). Probes are tiled at an average resolution of 35 base pair (bp), as measured from the central position of adjacent 25-mer oligos, leaving a gap of approx-imately 10 bp between probes. These arrays interrogate regions proxi-mal to transcription start sites and contain probes covering 59 percent of CpG islands annotated by UCSC in NCBI human genome assembly (Build 34). Each promoter region spans approximately from 7.5 kb up-stream through 2.45 kb downstream of 5'transcription start sites. For over 1,300 cancer-associated genes, coverage of promoter regions was expanded to include additional genomic content. This more extensive coverage spans from 10 kb upstream through 2.45 kb downstream of tran-scriptional start sites. Concerning the data analysis, Affymetrix has developed two software tools to analyse tiling array data. Affymetrix Tiling Analysis Software (TAS) provides the analysis for the GeneChip Tiling Arrays. TAS analyses feature intensity data stored in GCOS output .cell files and produces:

- Signal and p-values for each genomic position interrogated;
- Computation of genomic intervals based on computed signal and p-values;
- Visualizations for assessing the quality of the array data;

- Computation of summary statistics.

The results from TAS are imported into Affymetrix' Integrated Genome Browser to explore genomes and corresponding annotations from multiple data sources

3.6 Immunofluorescence assay

MCF-7 cells, grown on glass slides, were hormone-starved for 3 days and treated with E2 for the indicated times. Cells on coverslips were washed once with PBS, fixed for 20 min with paraformaldehyde (3%, w/v in PBS), permeabilized for 20 minutes with Triton X-100 (0.2%, v/v in PBS) and incubated for 1 h with PBS containing FCS (1%, v/v). For XPC and DNMT3a detection, coverslips were stained by incubation with anti-XPC or anti-DNMT3a antibodies diluted 1:250 in PBS for 1 h followed by three washings with PBS. Coverslips were then incubated in fluorescein isothiocyanate anti-rabbit (Jackson ImmunoResearch Laboratories, Inc, UK) diluted 1:200 in PBS. The coverslips were inverted

and mounted in Moviol (Calbiochem, CA) on glass slides. All images were captured with Zeiss confocal microscope 510.

3.7 AzadC trapping procedure

MCF-7 cells were hormone starved for 3 days and 30 min before E2 treatment had a pulse of 50 μM of 5-aza-2′-deoxycytidine. After PBS wash, cells were collected and genome extraction was performed using the MB buffer as indicated in the published protocol (Kiianitsa K and Maizels N, 2013). The genome was then sonicated and immunoprecipitated according with the ChIP assay procedure from Upstate Biotechnology. Antibodies used in these experiments were specific for DNMT3a or non-immune IgG. Immunoprecipitated DNA was analyzed by Real Time - PCR using sets of primers against PS2 ERE and PolyA regions. Normal serum and input DNA values were used to subtract/normalize the values from ChIP samples. All values represent the average of at least three independent experiments. A Slot-blot apparatus (Bio-Dot Biorad) was used to load on a nitrocellulose membrane the specified quantities of DNA and after blocking with 3% non-fat milk, the membrane was incubated with anti-DNMT3a antibody.

3.8 co-Immunoprecipitation assay and Western Blot

MCF7 cells, treated wit estrogens at the indicated times as described in the text, were lysed with Lysis Buffer containing 1% Triton, 300 mM NaCl, 50 mM Tris-HCl pH 7,5, 1mM EDTA pH 8, 2,5mM Na pyrophosphate (1%), 1mM NaVO4, 1mM NaF, 1x protease inhibitor (Roche) and 1mM PMSF. Lysates, clarified by centrifugation at 12000xg for 30' at 4°C, were co-immunoprecipitated with antibody against DNMT3a or normal IgG (as control), 1µg of antibody each mg

of total proteins, accordingly to the specific experimental needs. The lysate was treated with Micrococcal Nuclease (NEB) according to the manufacturer's instructions, in order to eliminate DNA-protein bindings and preserve only specifically physic bindings among proteins analyzed. Micrococcal nuclease works properly at 37°C, so in order to allow its function, even if at 4°C (protein lysates must be kept at low temperature), the enzyme was added and left in the samples until the end of the Co-IP. In each step, from lysis to immune-precipitated protein elution, an aliquot of Lysate was collected and analyzed on ethidium bromide-stained agarose gel to verify the gradual disappearance of DNA. The protein samples were separated by SDS-PAGE and subjected to western blot. The nitrocellulose membranes were immunoblotted with antibodies against DNMT3a, OGG1, XPC and Histone-H3 (or with anti-FLAG, anti-MCM7 or anti-DNMT3a for transfection/silencing the control), at the dilution of 1:1000 in TTBS (0,1% tween in TBS) over night at 4°C in gentle shacking. Image analysis for all gels was performed with ImageJ software using the "Gel Plot" plug-in.

3.9 Antibodies

ERalpha ab32063 (Abcam); LSD1 sc-271720 (Santa Cruz Biotechnology); TopoIIβ sc-13059 (Santa Cruz Biotechnology); DNMT3a ab2850 (Abcam); OGG1 sc-376935 (Santa Cruz Biotechnology); XPC sc-30156 (Santa Cruz Biotechnology); Ref-1 (Ape1) sc-9904 (Santa Cruz Biotechnology); Total H3 ab1791 (Abcam); Normal rabbit IgG sc-2027 (Santa Cruz Biotechnology); Normal mouse IgG sc-2025 (Santa Cruz Biotechnology).

3.10 Statistical analysis

All data are presented as mean ± standard deviation in at least three experiment in triplicate ($n \geq 9$). Statistical significance between groups was determined using Student's t test (matched pairs test or unmatched test were used as indicated in figure legends).

Table 1

	PRIMERS FOR mRNA	**LOCUS**
mRNA Fw	5’-CCAGACAGAGACGTGTACAGT-3’	pS2 / TFF1
mRNA Rev	5’-ATTCACACTCCTCTTCTGGA-3’	pS2 / TFF1
mRNA Fw	5’-GTGGTGGAGGAGCTCTTCAG-3’	BCL2
mRNA Rev	5’-CAAACTGAGCAGAGTCTTCAG-3’	BCL2
mRNA Fw	5’-GCAGACGAGCTGAGCGAGAAGC-3’	CAV1
mRNA Rev	5’-GAATAGACACGGCTGATGCACTG-3’	CAV1
18S Fw	5’-GCGCTACACTGACTGGCTC-3’	h18S
18S Rev	5’-CATCCAATCGGTAGTAGCGAC-3’	h18S
	PRIMERS FOR ChIP	**LOCUS**
ChIP ERE Fw	5’-CTAGACGGAATGGGCTTCAT-3’	pS2 / TFF1
ChIP ERE Rev	5’-TCTGAGAGGCCCTCCCGCCAG-3’	pS2 / TFF1
ChIP PolyA Fw	5’-CTACTCACTGCGGATGCCCCAG-3’	pS2 / TFF1
ChIP PolyA Rev	5’-GCTTCTGTATCCCTCCTCTGCTG-3’	pS2 / TFF1
ChIP II Intron Fw	5’-CCTTTTTATACGATGGGTTCTGA-3’	pS2 / TFF1
ChIP II Intron Rev	5’-CGGCCGTGACTCTGTGTAA-3’	pS2 / TFF1
ChIP ERE Fw	5’-CATTATAAGCTGTCGCAGAG-3’	BCL2
ChIP ERE Rev	5’-GAGGGTCAGGTGGACCACAG-3’	BCL2
ChIP PolyA Fw	5’-AGTAAATGTGCCCAGCCTCT-3’	BCL2
ChIP PolyA Rev	5’-TAGGGATGGTTCTCTGTTGC-3’	BCL2
ChIP ERE Fw	5’-GGATCTTAGATAAAGCTGGAAGG-3’	CAV1
ChIP ERE Rev	5’-GATCTCGCAGAGGACACCACAC-3’	CAV1
ChIP PolyA Fw	5’-GATGTGATTGCAGAACCA-3’	CAV1
ChIP PolyA Rev	5’-CAACAGCTTCAAAGAGTG-3’	CAV1
	PRIMERS FOR re-ChIP	**LOCUS**

ChIP ERE Fw	5’-CTAGACGGAATGGGCTTCAT-3’	pS2 / TFF1
ChIP ERE Rw (nucleosome)	5’-GCTTGGCCGTGACAACAGT-3’	pS2 / TFF1
ChIP PolyA Fw	5’-CTACTCACTGCGGATGCCCCAG-3’	pS2 / TFF1
ChIP PolyA Rev (nucleosome)	5’-GTTTCATCTCCTACGCCAAT-3’	pS2 / TFF1
ChIP II Intron Fw	5’-CCTTTTTATACGATGGGTTCTGA-3’	pS2 / TFF1
ChIP II Intron Fw	5’-CCTTGTGAGCCTTAATCCT-3’	pS2 / TFF1
ChIP ERE Fw (nucleosome)	5’-ATCCAGCCGCATCCCGGGAC-3’	BCL2
ChIP ERE Rev	5’-CATTATAAGCTGTCGCAGAG-3’	BCL2
ChIP PolyA Fw (nucleosome)	5’-ATGTTAGAAGCAATGAATGTA-3’	BCL2
ChIP PolyA Rev	5’-TAGGGATGGTTCTCTGTTGC-3’	BCL2
ChIP ERE Fw	5’-GGATCTTAGATAAAGCTGGAAGG-3’	CAV1
ChIP ERE Rev (nucleosome)	5’-TCGCTCTCGCCCTGAGCGCTT-3’	CAV1
ChIP PolyA Fw	5’-GATGTGATTGCAGAACCA-3’	CAV1
ChIP PolyA Rev (nucleosome)	5’-GATGAGTGCCATCGGGAT-3’	CAV1

4. RESULTS

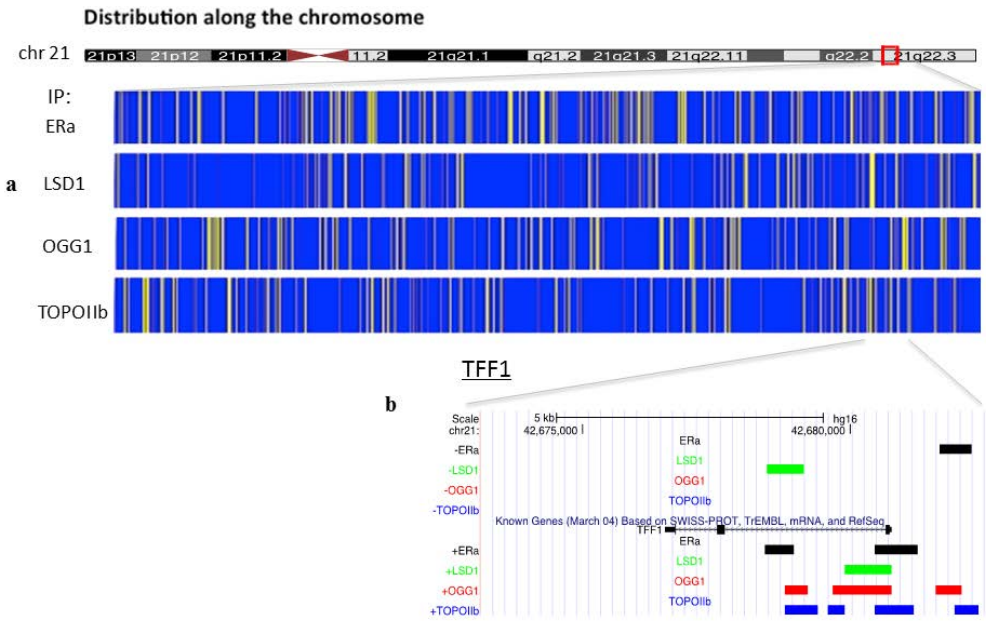
4.1 Estrogens synchronize the recruitment of the receptor ERα to discrete regions over the genome with enzymes that cause and resolve DNA oxidation

To find out the precise DNA regions that recruit ERα and enzymes involved in histone modification and the oxidation and repair of DNA (respectively, LSD1, OGG1 and TopoIIβ), we immunoprecipitated chromatin with antibodies specific for ERα, LSD1, OGG1 and TopoIIβ after stimulation of MCF7 cells with estrogens (50 nM E2 for 45 min). The resulting DNA-fragments from the independent immune-precipitations were utilized as probes to screen arrays of human promoters (Affymetrix, GeneChip Human Promoter 1.0R Array). These arrays are able to convey information about protein/DNA interactions in over 25,500 human promoters: with these powerful tools it is possible to analyze promoter regions from approximately 7.5 kb upstream through 2.45 kb downstream of 5' TSS. In particular, they interrogate genomic regions proximal to TSS and contain probes covering 59 percent of CpG islands annotated by UCSC Genome Browser (https://genome.ucsc.edu/index.html) in the NCBI human genome assembly (Build 34). The analysis uncovered several combinations of positive genomic sites for the diverse proteins analyzed (depending on the antibodies used for immune-precipitation), so we further analyzed regions concurrently bound by ERα, LSD1, OGG1 and Topo IIβ after estrogen challenging.

Here we show the map of three individual chromosomes (chr) that drew our attention because their sequences contain three prototypical estrogen-induced genes (TFF1 on chr21, CAV1 on chr7 and BCL2 on chr18).

- **Chromosome 21**

  We first evaluated the distribution of ERα, LSD1, OGG1 and TopoIIβ along chromosome 21, where TFF1 gene is located (pS2) (Figure 12a).



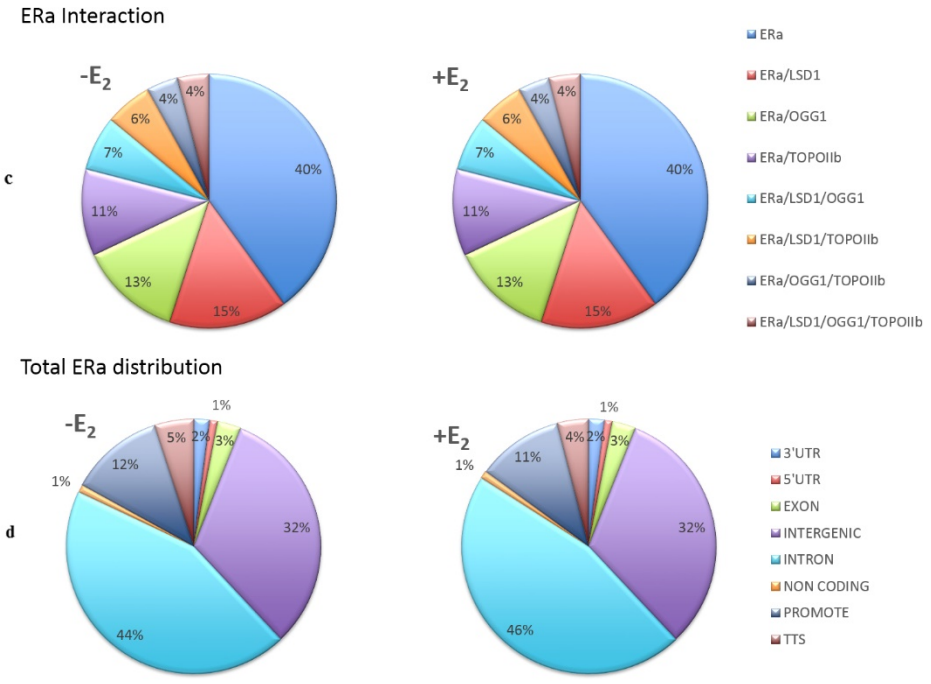


Figure 12. Screening human promoter array for ERa, OGG1, LSD1 and TOPOIIb sites induced by estrogen on chromosome 21.

a) Heatmap of ERa, LSD1, OGG1 and TOPOIIb binding relative to chromosome 21 at 45 min of E2 treatment. In yellow are the binding sites, while in blue the unbound regions. We consider only sites that fall within the 95 percentile. **b)** ERa, LSD1, OGG1 and TOPOIIb binding relative to TFF1 gene (located at the telomeric domain of chromosome 21). The black blocks represent ERa binding sites,

green blocks represent LSD1 sites, red blocks represent OGG1 binding sites and blue blocks represent TOPOIIb binding sites. The orientation is 3'UTR-5'UTR. The black arrows indicate the direction of the gene and predicted transcripts as reported in Genome Browser. **c)** Percentage of ERa interaction with other factors (LSD1, OGG1 and TOPOIIb) in complex with 2, 3 or 4 elements at 0 and 45 min of E2. The percent is calculated as the ratio of common number of bound sites to ERa-complex with respect to the total number of ERa binding sites (alone or in complex). **d)** Percentage of ERa distribution alone or in complex with LSD1, OGG1 and TOPOIIb along chromosome 21 at 0 and 45 min. We consider only sites that fall within the 95 percentile.

The yellow lines represent the occupancy of the regions and alignment of several lines indicates multiple recruitment of the enzymes. Comparing the blue (-E2) and the yellow (+E2), it appears that estrogen treatment induces redistribution of the receptor already bound to the chromatin to other sites with LSD1 and /or OGG1 and /or TOPOIIβ. These data imply that estrogens substantially synchronize the binding of these complexes to chromatin. To confirm this hypothesis, we analyzed the enrichment of ERα, LSD1, OGG1 and TOPOIIβ on a specific estrogen gene target: TFF1 gene. Figure 12b shows that estrogen induced two peaks of receptor recruitment, situated first at -500bp and second at +200bp from TSS. These peaks align with LSD1 and OGG1 recruitment, whereas we observe a reduction of TOPOIIβ signal.

- **Chromosome 18**

  We performed the same analysis of the distribution of ERα, LSD1, OGG1 and TOPOIIβ along chromosome 18, where the BCL2 gene is located (Figure 13). Figure 13a shows the distribution along chromosome 18 of ERα, LSD1, OGG1 and TOPOIIβ. The yellow lines represent the occupancy of the region. Figure 13c and 13d show an increase of the percentage of bound receptor, but not the changes in its distribution.

  To confirm this hypothesis, we analyzed the enrichment of ERα, LSD1, OGG1 and TOPOIIβ on the specific estrogen gene target present on chromosome 18, that is the BCL2 gene.

  Figure 13b shows that estrogen induced five peaks of receptor recruitment, situated first at -2500bp, -1500bp, -1300, -500 and +700bp from TSS. LSD1 recruitment is aligned with signal at -2500 and at +700bp. These peaks align with LSD1-OGG1 and TOPOIIβ and OGG1 recruitment.

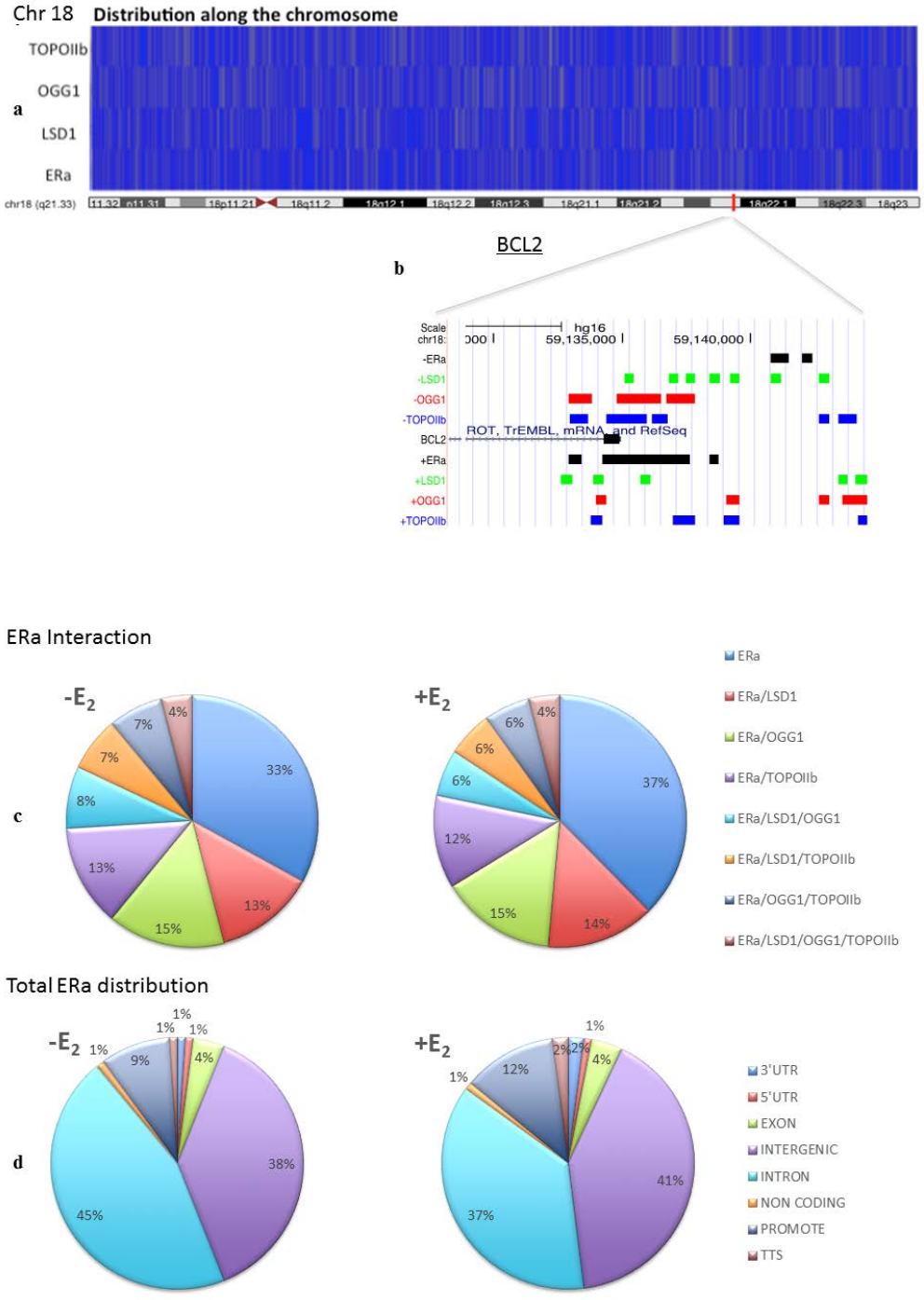


Figure 13. Screening human promoter array for ERa, OGG1, LSD1 and TOPOIIb sites induced by estrogen on chromosome 18.

a) Heatmap of ERa, LSD1, OGG1 and TOPOIIb binding relative to chromosome 18 at 45 min of E2 treatment. In yellow are the binding sites, while in blue the unbound regions. We consider only sites that fall within the 95 percentile. **b)** ERa, LSD1, OGG1 and TOPOIIb binding relative to BCL2 gene. The black blocks represent ERa binding sites, green blocks represent LSD1 sites, red blocks represent OGG1 binding sites and blue blocks represent TOPOIIb binding sites. The orientation is 3'UTR-5'UTR. The black arrows indicate the direction of the gene and predicted transcripts as reported in

Genome Browser. **c)** Percentage of ERa interaction with other factors (LSD1, OGG1 and TOPOIIb) in complex with 2, 3 or 4 elements at 0 and 45 min of E2. The percent is calculated as the ratio of common number of bonds sites to ERa-complex respect to the total number of ERa binding sites (alone or in complex). **d)** Percentage of ERa distribution alone or in complex with LSD1, OGG1 and TOPOIIb along chromosome 18 at 0 and 45 min. We consider only sites that fall within the 95 percentile.

- **Chromosome 7**

  As for chromosome 21 and 18, we evaluated the distribution of ERα, LSD1, OGG1 and TOPOIIβ along chromosome 7. Figure 14a shows the distribution of ERα, LSD1, OGG1 and TopoIIβ. The yellow lines represent the occupancy of the region. Also here, alignment of many lines indicates multiple recruitment of the enzymes at the ERE sites. We found that estrogen stimulation did not increase the interactions and did not change the receptor distribution, confirming that estrogen substantially synchronizes the binding of these complexes to chromatin (Figure 14c and 14d).

  To confirm this hypothesis, we analyzed the enrichment of ERα, LSD1, OGG1 and TOPOIIβ on specific estrogen gene target: CAV1 gene. Figure 9b shows that estrogen induced many peaks of receptor recruitment, situated at -500bp, on the TSS and in the second intron. These peaks align with LSD1, whereas OGG1 recruitment is present only at -500bp from TSS. TopoIIβ is also aligned with ERα/LSD1/OGG1 signal.

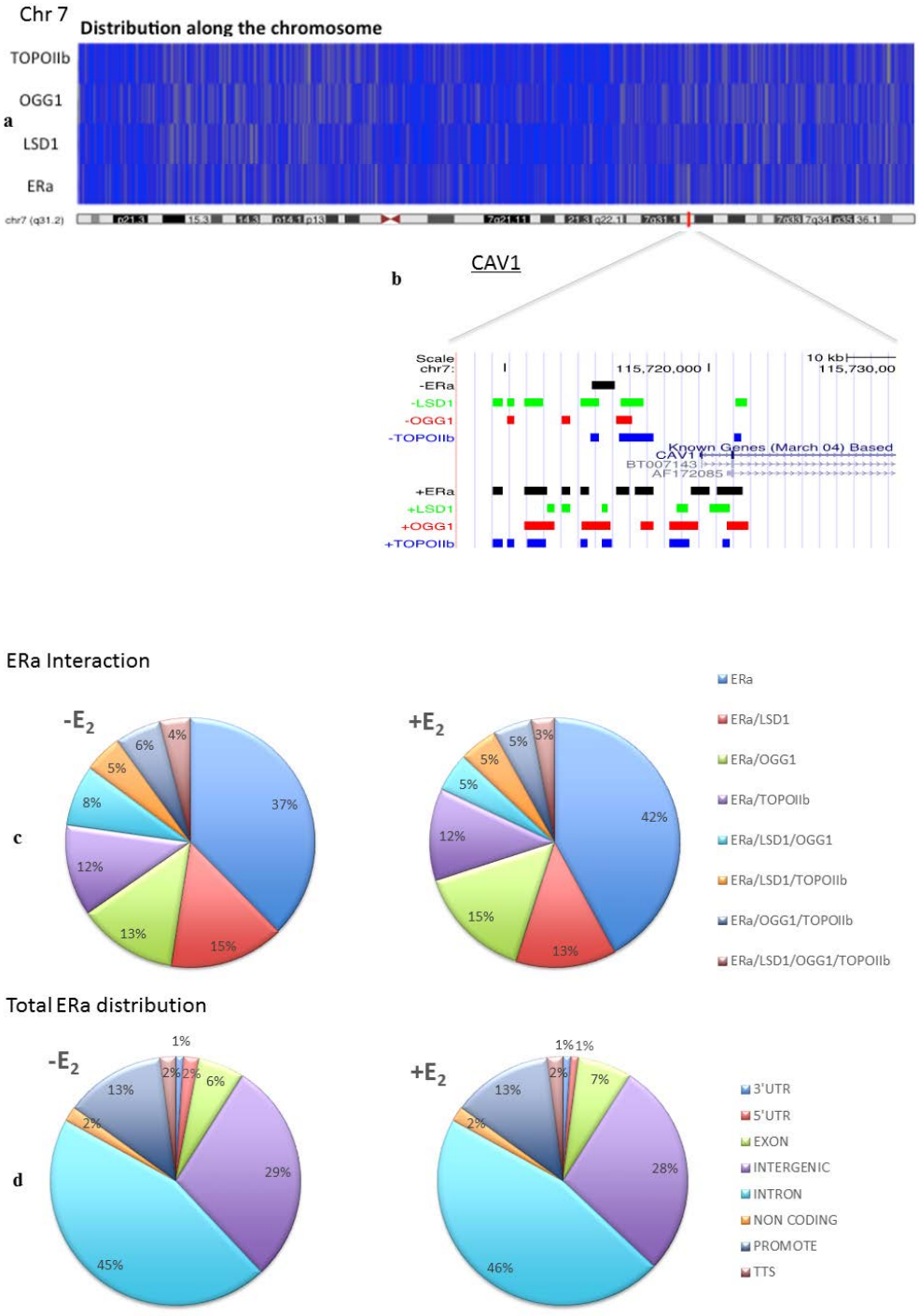


Figure 14. Screening human promoter array for ERa, OGG1, LSD1 and TOPOIIb sites induced by estrogen on chromosome 7.

a) Heatmap of ERa, LSD1, OGG1 and TOPOIIb binding relative to chromosome 18 at 45 min of E2 treatment. In yellow are the binding sites, while in blue the unbound regions. We consider only sites that fall within the 95 percentile. **b)** ERa, LSD1, OGG1 and TOPOIIb binding relative to CAV1 gene. The black blocks represent ERa binding sites, green blocks represent LSD1 sites, red blocks represent OGG1 binding sites and blue blocks represent TOPOIIb binding sites. The orientation is 3'UTR-5'UTR. The black arrows indicate the direction of the gene and predicted transcripts as reported in

Genome Browser. **c)** Percentage of ERa interaction with other factors (LSD1, OGG1 and TOPOIIb) in complex with 2, 3 or 4 elements at 0 and 45 min of E2. The percent is calculated as the ratio of common number of bonds sites to ERa-complex respect to the total number of ERa binding sites (alone or in complex). **d)** Percentage of ERa distribution alone or in complex with LSD1, OGG1 and TOPOIIb along chromosome 18 at 0 and 45 min. We consider only sites that fall within the 95 percentile.

The data obtained from this analysis shed light on the role of E2 in recruiting and re-distributing, at discrete chromatin sites, the proteins ERα, LSD1, OGG1 and Topo IIβ, alone or grouped with each other. It is noteworthy that the accumulation of these complexes, and in particular of the ERα-OGG1, occurs massively throughout the genome but notably in the EREs and near TSS of prototypical E2-regulated genes. This finding implies that during transcription induction estrogen influences the recruitment of repair enzymes.

4.2 Estrogens induce the recruitment of epigenetic modifiers at responsive regions of E2-induced genes

Oxidative agents and ROS released during the assembly of the transcription initiation complex can damage the DNA, as shown for LSD1-mediated release of H_2O_2 in actively transcribed regions of Myc- or E2- induced genes (Amente S et al., 2010; Perillo B et al., 2008). The accumulation of oxidized G in the nucleus has been correlated with estrogen induction of transcription (Figure 15).

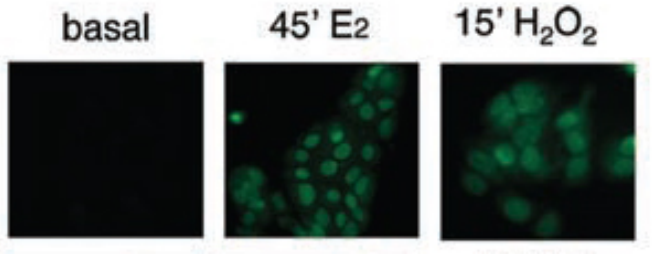


Figure 15. Estrogens induce nuclear 8-oxo-Gs foci formation in MCF7 cells as the treatment with H_2O_2. (Adapted from Perillo et al., 2008).

The BER pathway is firstly involved in the recognition and resolution of 8-oxo-G, but also the NER pathway have been implicated in the repair of oxidized Gs (Melis JPM et al., 2013) and it has been demonstrated that NER enzymes can

be recruited to regulatory sites upon hormonal stimuli (Zuchegna et al., 2014). Moreover, the *de novo* methyltransferase DNMT3a has been found at the ERE sites of several E2-induced genes, which undergo cycles of methylation (Métivier R et al., 2008). ChIP analysis with antibodies to OGG1 (BER) or XPC (NER) shows that these 2 proteins are recruited at the ERE and polyA sites of estrogen target genes (BCL2-CAV1-pS2) (Figure 16). Also DNMT3a (Figure 17) undergo an ordered and timely recruitment at distant regulatory sites of E2-induced genes.

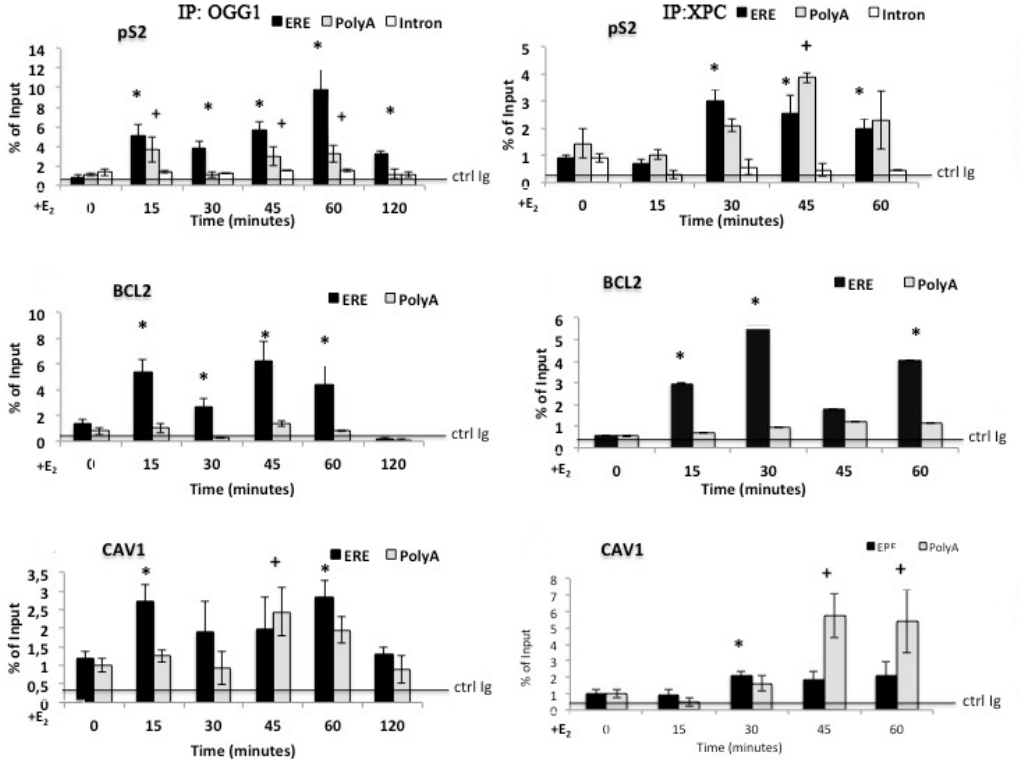


Figure 16. Estrogen induces recruitment of OGG1 and XPC at the ERE and PolyA sites of target genes. MCF7 cells were serum starved and exposed to 50 nM E2 at the indicated times (0, 15, 30, 45 and 60 min). qChIP was carried out using specific antibodies recognizing OGG1 and XPC. Is shown the occupancy at ERE and PolyA of pS2, BCl2 and Cav1 genes. The statistical analysis derived from at least 3 experiments in triplicate (n ≥9; Mean ± SD). *p <0.01 (matched pairs t test) compared to E2-unstimulated sample (Adapted from Pezone A, **De Rosa M** et al., submitted).

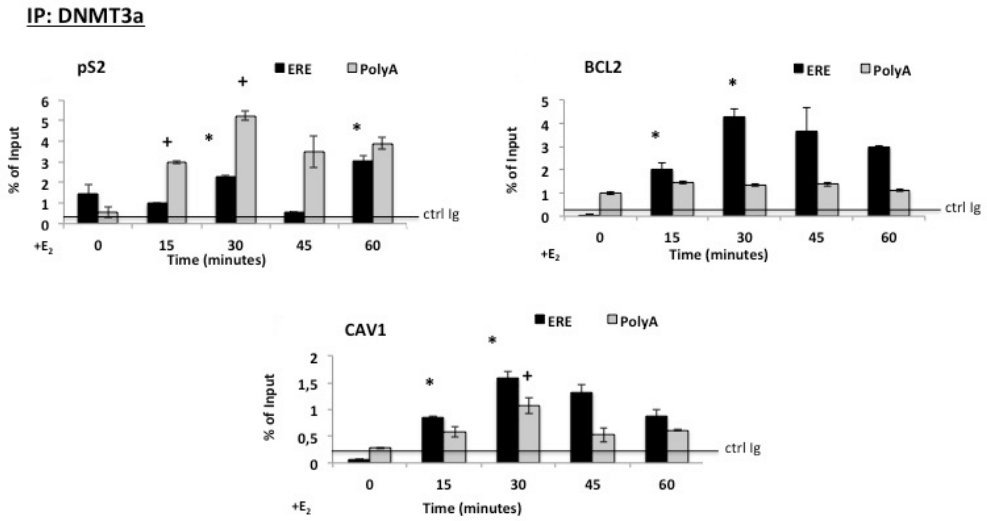


Figure 17. Estrogen induces recruitment of DNMT3a at the ERE and PolyA sites of target genes. MCF7 cells were serum starved and exposed to 50 nM E2 at the indicated times (0, 15, 30, 45 and 60 min). qChIP was carried out using the specific antibody recognizing DNMT3a. Here is shown the occupancy at ERE and PolyA sites of pS2, BCl2 and Cav1 genes. The statistical analysis derived from at least 3 experiments in triplicate (n ≥9; Mean±SD). *p <0.01 (matched pairs t test) compared to E2-unstimulated sample (Adapted from Pezone A, **De Rosa M** et al., submitted).

OGG1 is recruited to the ERE and polyA sites with 2 main peaks at 15 – 30 min and 45 – 60 min of E2; XPC accumulates selectively at the ERE at 30 minutes of E2 and at PolyA site at 45 min of E2. DNMT3a instead, is recruited to the ERE and PolyA with a period of 30 - 60 min upon E2 induction. These data suggested that the recognition and repair of DNA oxidation occur at distant sites at 5’ and 3’ end of the gene. Moreover, ERE and PolyA regions undergo the same periodic oscillation for the protein observed, suggesting that these distant regions interact upon estrogenic stimulus.

We also wondered whether an E2 challenge was able to highlight a macroscopic difference in the nuclear enrichment of some proteins previously analyzed by ChIP analysis; therefore, we tested by immunofluorescence the nuclear staining of XPC and DNMT3a upon estrogen treatment (Figure 18).

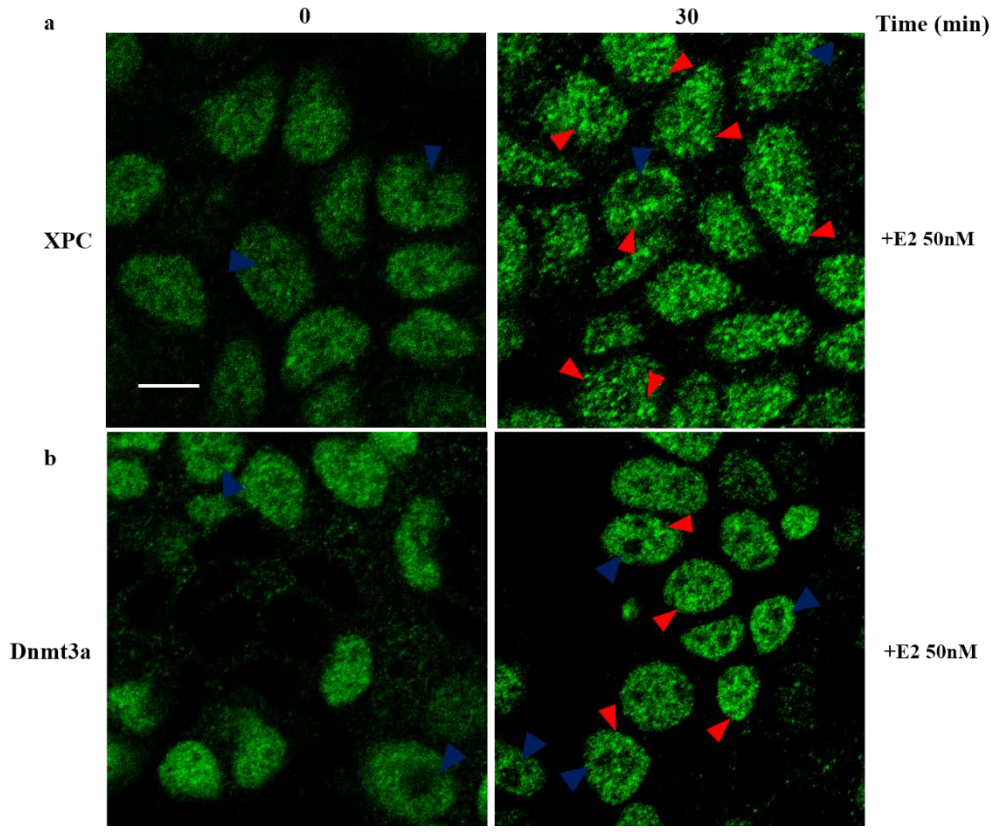


Figure 18. Estrogen induces XPC and DNMT3a foci formation in MCF7 cells. MCF7 cells in basal conditions or treated with 50nM E2 for 30 min were subjected to immunostaining with polyclonal anti-XPC **(a)** and monoclonal anti-DNMT3a **(b)** antibodies. Images were collected and analyzed by confocal microscopy. Scale bar, 10 µm. Legend: blue arrows indicate nucleoli; red arrows indicate discrete foci of XPC **(a)** and DNMT3a **(b)** significantly detectable after E2 treatment.

We noticed a discrete increase in the nuclear signal for both XPC and DNMT3a after E2 treatment and an enrichment of discrete foci for both proteins. These data confirmed the recruitment of these proteins in specific segments of the nucleus, and this event is tightly controlled by estrogens. Immunofluorescence staining shows a general enrichment in the nucleus but does not specify if the binding of the protein(s) takes place on chromatin or on naked DNA. Our data suggest that DNMT3a underwent the same regulation of repair enzymes and was recruited at chromatin by E2 induction. Since DNMT3a is a DNA methyltransferase and interacts directly with DNA, we tested with a novel method the direct binding of DNMT3a with DNA. We exploited the ability of 5-aza-2'-deoxycytidine (5-AzadC) to trap covalently DNMT enzymes to study DNA-bound DNMT3a by ChIP assay in the absence of formaldehyde. 5-AzadC is a chemical analogue of the cytosine present in DNA. 5-Aza-2′-

deoxycytidine-triphosphate is a substrate for the DNA replication machinery and can be incorporated into DNA. When DNMTs recognize 5-AzadC:G base pairs as substrate, they initiate the methylation reaction by a nucleophilic attack, which results in a stable covalent bond between the carbon-6 atom of the cytosine ring and the enzyme, permanently blocking its DNA methyltransferase activity (Stresemann C and Lyko F, 2008). Under these conditions, we can detect the direct contact of the enzyme (bound to 5-AzadC in the DNA) by performing an immunoprecipitation of the enzyme without crosslinking with formaldehyde (Figure 19).

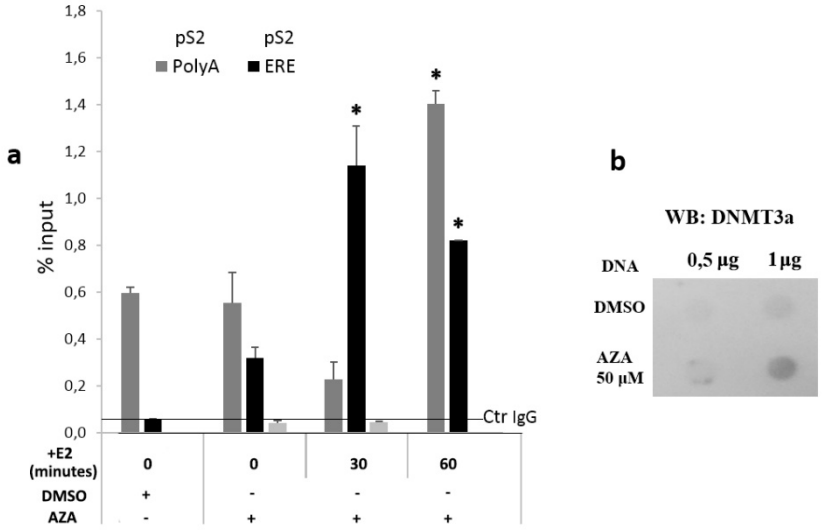
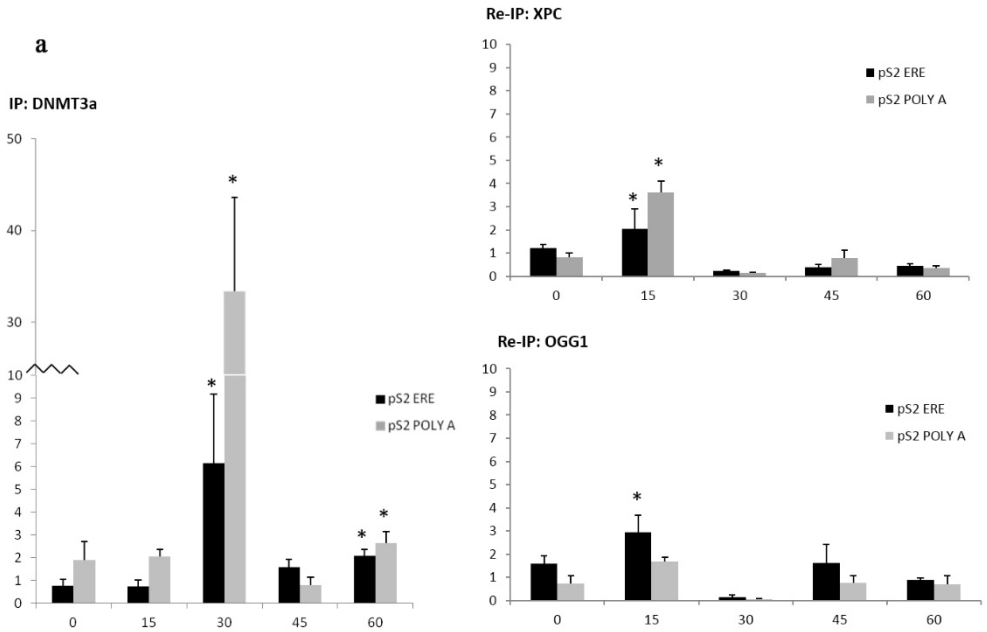


Figure 19. Aza-trapping for the detection of DNMT3a directly linked with DNA. MCF7 cells in basal conditions or stimulated with 50nM E2, were treated with DMSO as a control or pre-treated for 30 min with 5-azadC, and then the DNA extracted was immune-precipitated for DNMT3a. **(a)** The immunoprecipitated DNA was used as template to amplify ERE and PolyA sites of pS2 gene. **(b)** Dot-blot assay performed loading the listed concentrations of DNA and revealing for Western Blot assay, incubating with the monoclonal anti-DNMT3a antibody. The statistical analysis derived from at least 3 experiments in triplicate (n ≥9; Mean ± SD). * = p <0.05 (matched pairs t test) compared to E2-unstimulated sample.

It is important to stress that this procedure reveals the direct contact of DNMT3a to the DNA. Figure 19 shows that DNMT3a binding to the pS2 ERE site is induced by E2 treatment at 30 minutes and later at PolyA site. Conversely, ChIP analysis (Figure 17) shows that DNMT3a is already present in the absence of E2, suggesting that the enzyme interacts with histones or other chromatin proteins independently on its binding to the DNA.

4.3 Estrogen stimulates the association of DNMT3a with BER or NER enzymes at target sites

The promoter array data revealed that estrogen stimulates the interaction of the ERα with repair enzymes at discrete chromatin sites of responsive genes. Since we identified cyclical recruitment of the single proteins after E2 challenging, we wondered whether DNMT3a interacts in an E2-dependent fashion with BER and NER enzymes. To obtain a more precise analysis of the various protein-protein complexes recruited to the ERE and PolyA sites, we prepared single nucleosomes by micrococcal nuclease digestion of chromatin and sucrose gradient purification and performed sequential ChIP assays (re-ChIP) to reveal DNMT3a and OGG1 or XPC (Figure 20).



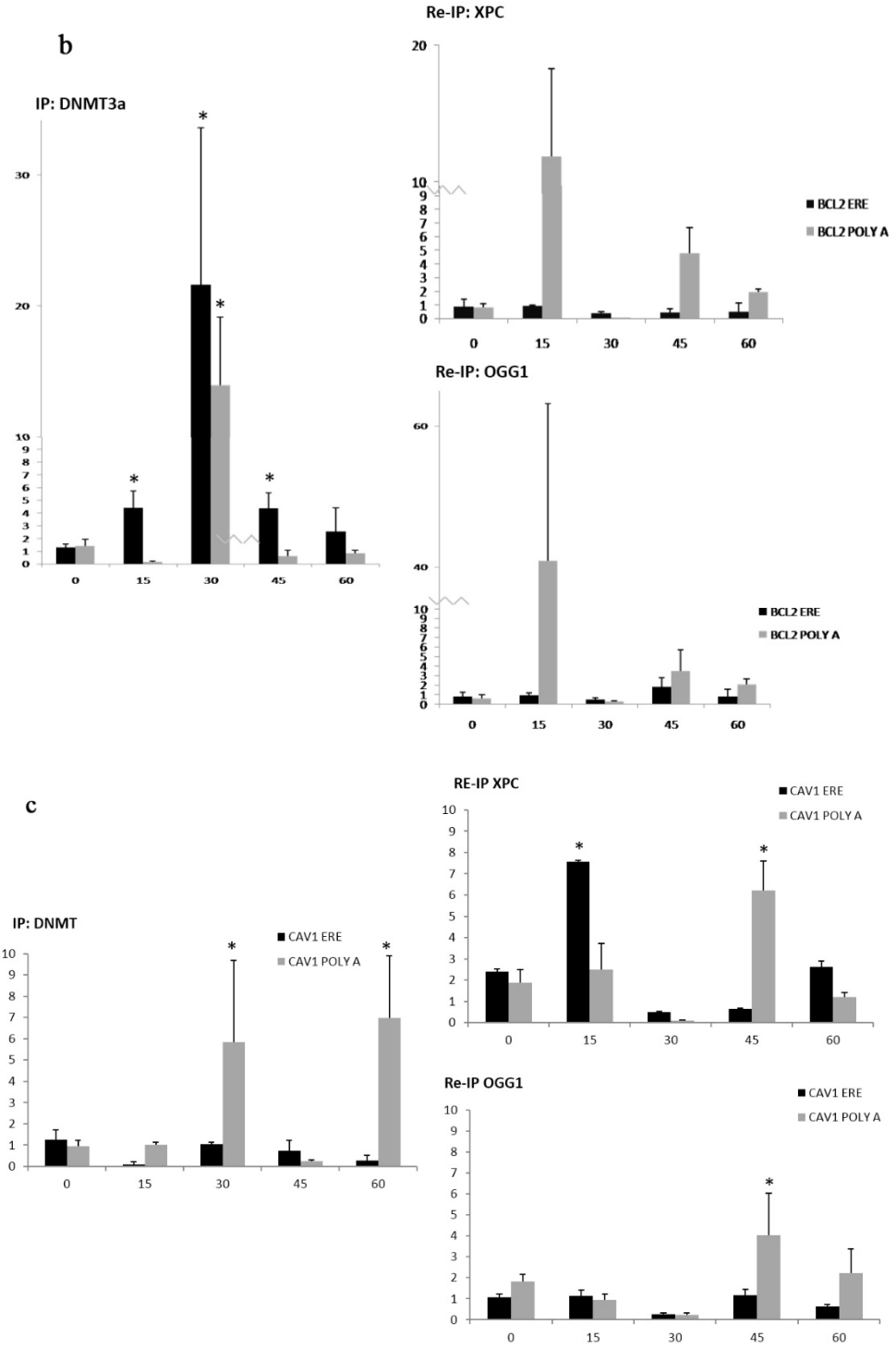


Figure 20. Re-ChIP assays performed on mononuclesomes to detect DNMT3a and its interactions with OGG1 and XPC. (a) ChIP and re-ChIP on ERE and PolyA regions of pS2 gene. (b): ChIP and re-ChIP on ERE and PolyA regions of BCl2 gene. (b): ChIP and re-ChIP on ERE and PolyA regions of Cav1 gene.

DNMT3a was firstly isolated thanks to its association with "bivalent nucleosomes" that contain both the repressive (H3-K9me3) and the activating (H3-K9ac) marks (Harikumar A and Meshorer E, 2015). Then, DNMT3a

nucleosomes were subjected to re-ChIP assays with anti OGG1 and XPC antibodies. Our data show that DNMT3a interacts with chromatin with a peak at 30 minutes, while OGG1 and XPC associated with DNMT3a peak at 15 and 45 minutes throughout the E2-induced cycle, both at ERE and PolyA regions of the analyzed genes. It is important to stress that the re-ChIP data do not show the recruitment of the single OGG1 or XPC at the chromatin sites, but the amount of the two proteins complexed with DNMT3a. Taking into account this information, it is remarkable that the highest signals for XPC and OGG1 are detected when DNMT3a alone is poorly loaded on chromatin. This suggests that most of DNMT3a molecules are associated with XPC and OGG1, which are recruited to the chromatin through DNMT3a.

4.4 DNMT3a interacts with BER and NER enzymes

Since we found that BER-NER enzymes and DNMT3a were coordinately recruited during E2-mediated transcription, we searched for a possible physical interaction between DNMT3a and repair enzymes, wondering whether this binding was influenced by estrogen induction.

A co-immunoprecipitation assay showed that the endogenous DNMT3a interacts strongly with OGG1 (BER) and is bound also to XPC (NER) and APE1 (BER): notably, the binding appears stronger after estrogen treatment (Figure 21).

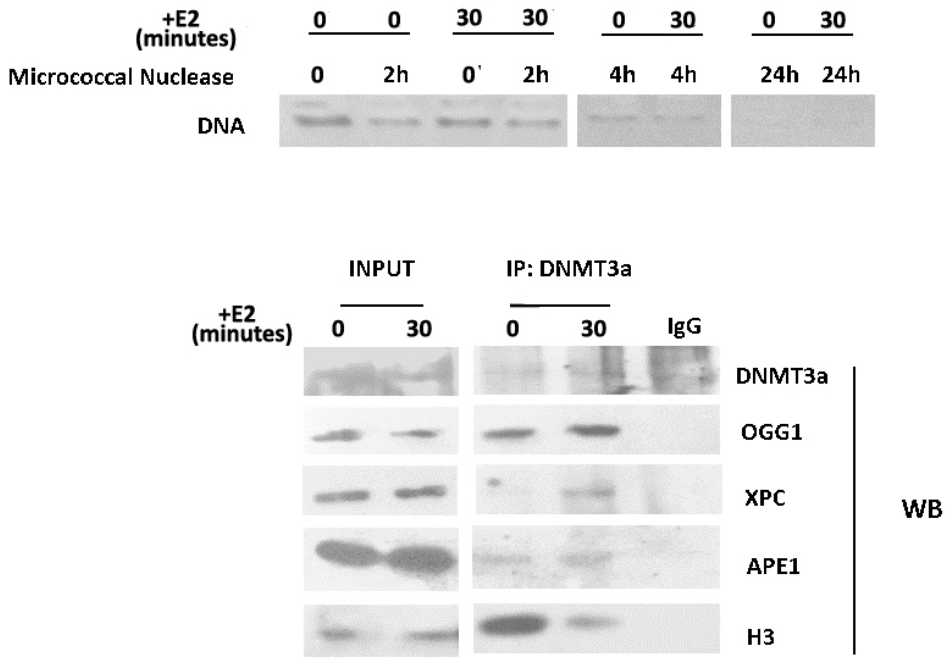


Figure 21. DNMT3a interaction with BER and NER enzymes. MCF7 cells were serum starved and exposed to 50 nM E2 at the indicated times (0 and 30 min). A. Co-IP analysis shows DNMT3a binding to OGG1, XPC and APE1. In the upper panel is shown the gradual disappearance of genomic DNA in protein samples, obtained through a treatment with micrococcal nuclease, in order to eliminate DNA and test only real interactions not influenced by DNA.

To better define which domain of the methyltransferase was mainly involved in the interactions and whether the binding with other proteins was compatible with DNMT3a catalytic function, we overexpressed in MCF7 cells a peptide containing the catalytic portion of the protein (282C). We first evaluated the effect of the peptide overexpression on the E2-dependent transcription (Figure 22).

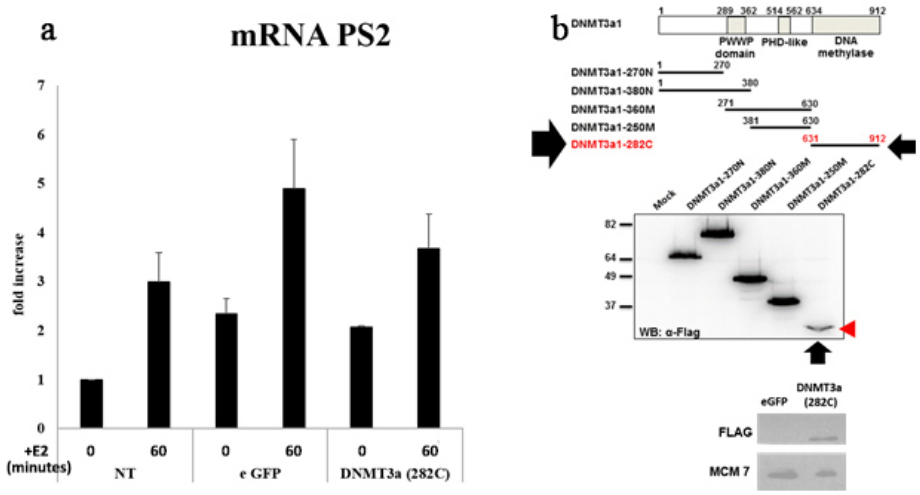


Figure 22. DNMT3a catalytic portion (282C) overexpression does not influence E2-induced transcription of pS2 gene. MCF7 cells were serum starved and exposed to 50 nM E2 at the indicated times (0 and 60 min). **(a)** Effects of DNMT3a peptides overexpression on pS2 induction of transcription. **(b)** Map of DNMT3a domains and Western blot analysis to control overexpression of 282C (the red arrow indicates the catalytic fragment).

The transfection of the catalytic peptide did not alter pS2 gene expression. Then, we performed a pull down assay after transfection of the catalytic peptide DNMT3a1 (282C) in MCF7 cells (Figure 23).

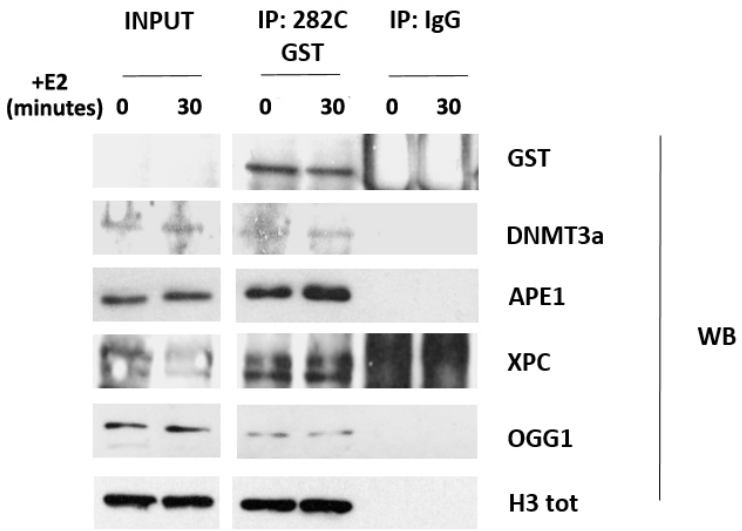


Figure 23. The catalytic domain of DNMT3a strongly interacts with Ape1 and XPC. The DNMT3a1 282C-GST was purified and a pull down assay was performed using MCF7 protein lysate treated at indicated time with E2 (0 and 30 min).

We found that DNMT3a catalytic domain binds APE1 and XPC, while the endogenous protein DNMT3a interacts with OGG1. These data suggest that the interaction with OGG1-APE1 and XPC may regulate the activity and the substrate specificity of the enzyme.

4.5 BER and NER enzymes activity is required for estrogen-induced transcription

To analyze the action of BER and NER on E2-induced transcription we altered or depleted these proteins in MCF7 cells exposed to E2 and measured transcription. First we overexpressed a dominant negative version of OGG1 (OGG1 DN) and we measured the impact on E2-mediated transcription and the influence on the other recruitment of the proteins (Figure 24 and 25).

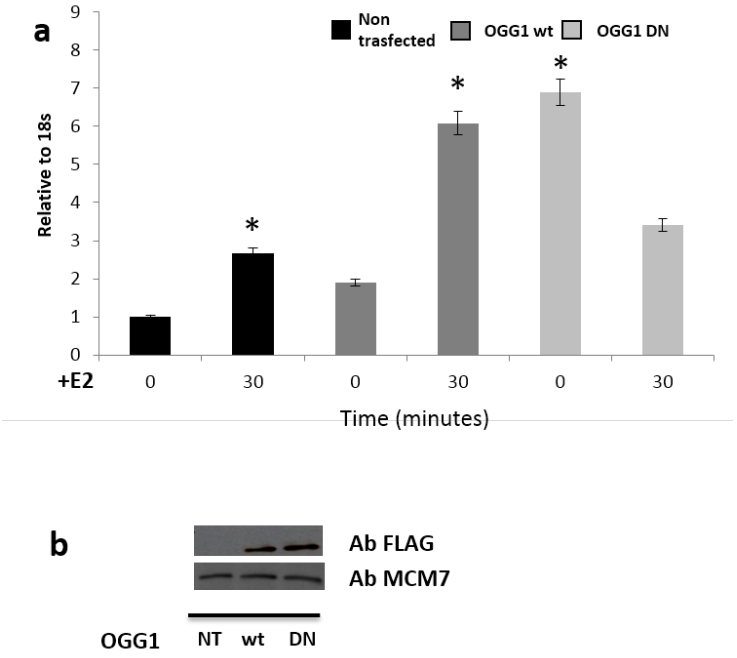


Figure 24. OGG1 DN overexpression cause the loss of E2 transcription induction. Total RNA was prepared from MCF7 cells hormone-starved or stimulated with 50nM E2 for 0 or 30 min and analyzed by qPCR with specific primers to pS2 mRNA normalized to 18S mRNA levels. **(a)** mRNA expression analysis of pS2 gene in cell transfected with OGG1 wild type (WT) or with dominant negative form K341R (DN). **(b)** Western Blot analysis as a control of OGG1 wt and DN protein expression (NT= non trasfected sample).

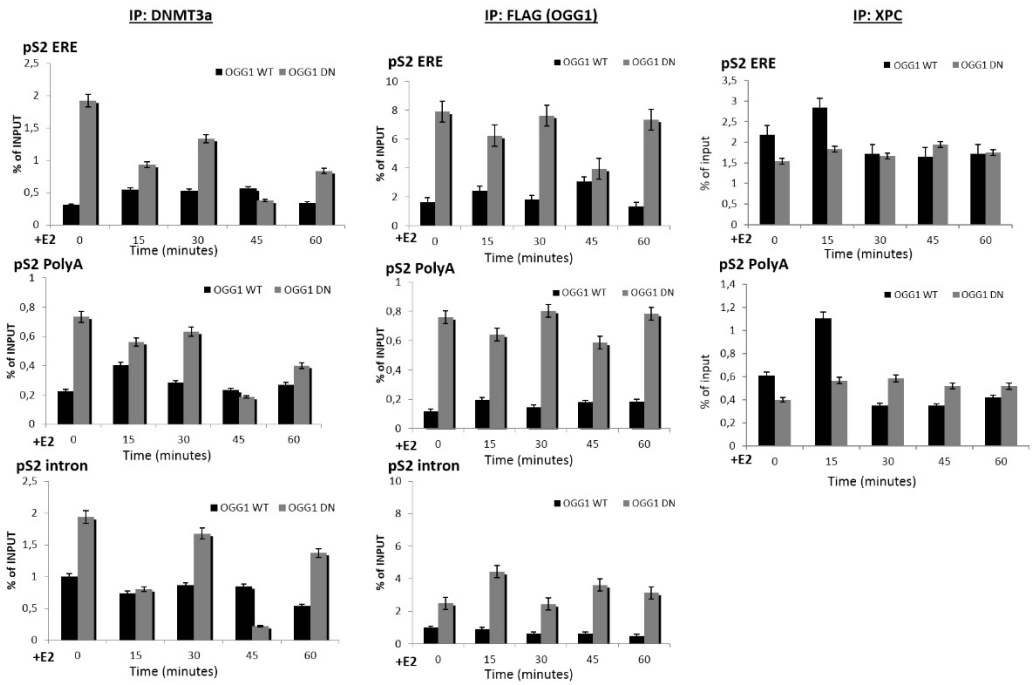


Figure 25. OGG1 DN overexpression cause the loss of cyclic recruitment of interacting proteins and high levels in basal conditions. ChIP assays performed in MCF7 cells overexpressing Flag-tagged WT or DN form of OGG1. The chromatin was immune-precipitated with antibodies against DNMT3a, XPC and FLAG (as control of OGG1) and the DNA fragments obtained were amplified with primers for the specific regions ERE and PolyA of the E2-induced genes of interest.

When the dominant negative form of OGG1 was overexpressed, the recruitment of XPC was reduced, while the recruitment of DNMT3a was increased at the basal level in the absence of E2. Moreover, depletion of DNMT3a and XPC with a siRNA inhibited E2-induced transcription (Figure 26).

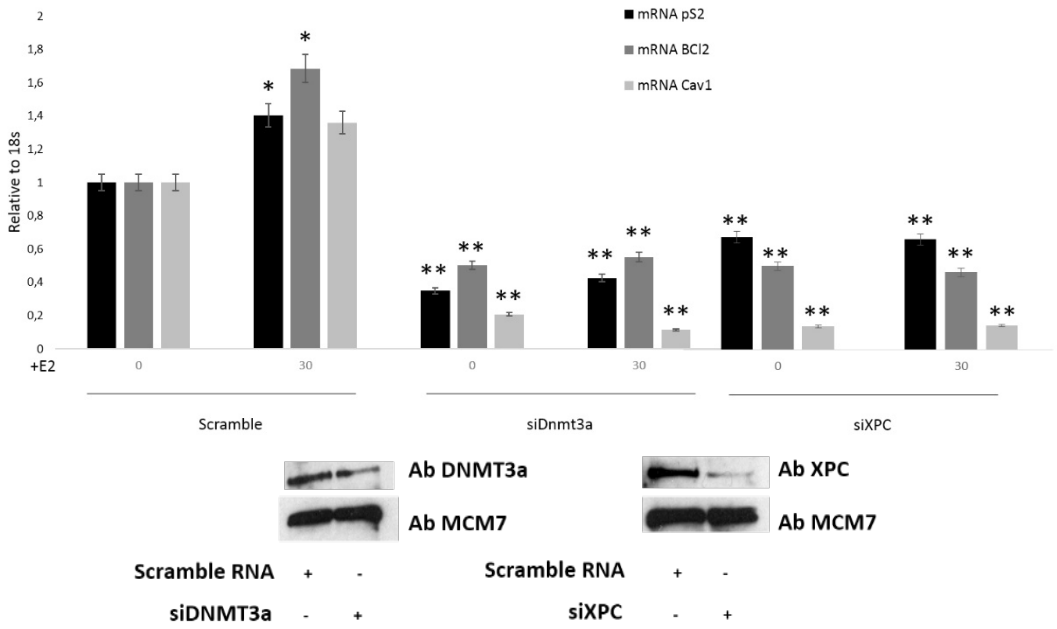


Figure 26. The depletion of DNMT3a and XPC impairs E2-induced transcription. Total RNA was prepared from MCF7 cells hormone-starved or stimulated with 50nM E2 for 0 or 30 min and analyzed by qPCR with specific primers to pS2 mRNA normalized to 18S mRNA levels. (a) mRNA expression analysis of pS2 gene in cell transfected with siDNMT3a or siXPC or with a scramble vector. (b) Western Blot analysis as a control of DNMT3a and XPC silencing. The statistical analysis derived from at least 3 experiments in triplicate (n ≥9; Mean±SD). * = p <0.005 (matched pairs t test) compared to E2-unstimulated sample; ** = p <0.005 (matched pairs t test) compared to sample expressing scramble vector.

These data suggest that the ordered recruitment of DNMT3a and BER-NER enzymes is essential for productive E2-induced transcription.

5. DISCUSSION

Transcriptional initiation in eukaryotes is a complex process that requires multiple enzymatic activities, and recent evidence suggests that there is a high frequency of DNA damage during transcriptional activation, though no clear mechanistic link between gene activation and components of the DNA repair machinery has been established (Beato M et al., 2015). Transcription induced by estrogens is associated with chromatin modification dependent on histone methylation-demethylation cycles. It has been shown that LSD1 demethylates H3K9me2 at regulatory regions of estrogen target genes causing localized DNA oxidation and recruitment of repair enzymes. The repair of oxidative lesions triggers chromatin and DNA conformational changes that dissipate transcriptional supercoiling and enhance estrogen-induced transcription (Perillo et al. 2008). The formation of loops connecting the 5' gene ends, 3' ends and enhancers has been shown necessary to obtain productive transcription cycles in many other genes induced by nuclear receptors (Li W et al., 2013; Le May N et al., 2012). In the specific case of estrogenic induction, it is still unclear how estrogens coordinate the repair of oxidative lesions linked to the initiation of transcription.

5.1 DNA repair enzyme assembly at estrogen responsive sites

Our data show that, upon estrogen induction, ERα enucleates a major fraction of complexes with LSD1 and certain repair enzymes (OGG1 and Topo IIβ) in inter-genic and intronic regions. E2 stimulation induces an accumulation of the receptor at ERE regions, which are located at the 5' end (as in the case of TFF1 and CAV1 genes) or inside the gene (BCL2), and a massive re-distribution at discrete chromatin sites of ERα-OGG1-TopoIIβ complexes. Notably, the repair protein complexes are re-distributed in EREs and closely to TSS of prototypical E2-induced genes. Focusing our attention on pS2, Cav1 and BCl2 (located respectively on chromosome 21, 18 and 7), we firstly analyzed the recruitment of BER and NER enzymes at distant sites regulated by the active receptor (ERE and PolyA regions). Our findings indicate a hierarchical recruitment of these proteins, thus suggesting different roles of BER and NER enzymes in E2-induced transcription.

5.2 DNMT3a cooperates with repair enzymes

DNA methylation is thought to coordinately regulate the chromatin status via the interaction of DNMTs with other components of the machinery (Jin B et al., 2011). The induction of CpG methylation cycles within the promoter of estrogen target genes, together with the recruitment of DNMT enzymes, has been demonstrated (Metivier et al., 2003). Therefore, with particular interest for DNMT3a, we used several different ways to uncover the precise recruitment timing and the contacts established by this protein with chromatin or DNA. We found that, upon E2 stimulus, DNMT3a contacts the DNA and the chromatin with a temporal pattern overlapping the BER-NER cycles. Moreover, our data show that the recruitment of BER and NER enzymes on distant regulatory regions of the E2-induced genes accounts for the interaction with DNMT3a. Surprisingly, searching for physical interactions between DNMT3a and BER-NER enzymes we discovered that estrogen induces the binding of the DNA methyltransferase with OGG1, APE1 and XPC. These enzymes interact with the catalytic segment of DNMT3a and likely inhibit the methyltransferase activity. These data suggest that DNMT3a activity and substrate specificity may be regulated by BER and NER enzymes.

5.3 DNA oxidation and repair: cause or effect of transcription?

Our data suggest that DNA oxidation and histone/DNA methylation-demethylation cycles (Métivier R et al., 2008) are linked: histone demethylation may be the first trigger of DNA oxidation, which is followed by site –specific repair. The oxidation of G and C and the subsequent repair of the lesions induces transient site-specific nicks, which reduce transcription-generated supercoiling (Kouzine F et al., 2013). Relaxed DNA allows the formation of DNA loops juxtaposing the distant regulatory sites of the gene that should be transcribed (Zuchegna et al., 2014). We detected a significant inhibition of E2-induced transcription after depletion of the BER-NER proteins or DNMT3a. Over-expression of a mutant OGG1, unable to dissociate from the DNA, perturbs the recruitment of DNMT3a and XPC and inhibits E2-induced transcription.

In conclusion, we suggest that the loss of DNMT3a blocks the OGG1-NER switch at the ERE sites and this seriously compromises the repair of 8-oxo-Gs.

DNMT3a and BER-NER enzymes act in a well-tuned manner in order to repair oxidative lesions and permit cycles of productive transcription.

5.4 Dysregulated transcription induction is associated with increased mutational rate in estrogen-dependent cancers

Uncontrolled and continuous estrogen exposure mainly accounts for the risk of breast cancer (Yager JD and Davidson NE, 2006) and the expression of estrogen receptor is a prognostic factor for its clinical outcome. For the majority of ER-positive breast cancers, the modulation of estrogen signaling with selective ER antagonists (SERM, such as tamoxifen) is still the most used treatment (Jordan VC et al., 2007). Despite the selectivity of these therapies, long-term treatment with SERMs is linked with resistance acquisition (Sengupta S et al., 2008).

It is well known that estrogen-induced transcriptional stress induce DNA damage and mutations. Elevated genomic instability in breast cancer cells, due to R-Loop formation, has been correlated with estrogen challenging (Stork CT et al., 2016). Considering the impact of estrogen exposure on point mutations, it is noteworthy that the 8-oxo-Gs resulting from LSD1 activity during estrogen-induced initiation of transcription can base-pair with A, instead of C, an error that, if not resolved, can induce transversion point mutations (C→A) during replication. Therefore, understanding thoroughly the transcriptional mechanisms controlled by estrogen and the role of the repair and methylating enzymes involved, could open a window of opportunity to detect novel cancer targets and therapeutic modulators.

6. CONCLUSION

Our hypothesis is that estrogens coordinate and synchronize estrogen receptor and repair enzyme complexes to interact at target regulatory regions. These regions enucleate the transcription initiation complex and define the termination sites. DNMT3a, initially positioned on the chromatin thanks to its contacts with histones H3K9/H3K4, after E2 stimulation is recruited at 8-oxo-G sites, where interacts successively with OGG1 and Ape1. These enzymes seem essential for the formation of chromatin loops and greatly reduce the possible mutations due to oxidation of Gs and Cs. The oxidation and repair may proceed in a strand specific fashion as suggested by the evidence that, when 8-oxo-G is present in a CpG doublet, DNMT3a activity on the adjacent C is inhibited (Maltseva DV et al., 2009)

Our data open a new window on transcription, oxidation, DNA repair and methylation: DNMT3a and BER-NER enzymes act in a tuned manner in order to repair oxidative lesions induced by the transcription machinery itself. The resulting CpG methylation induced by E2 marks the sites that are repaired by BER or NER enzymes. Understanding this mechanism may provide a new opportunity to target proteins involved in aberrant transcription in cancer cells. High transcription rate may induce by itself a conspicuous mutational burden, not only in the estrogen-induced network but in all hormone-regulated systems.

7. REFERENCES

Aihara E, Engevik KA, Montrose MH. *Trefoil Factor Peptides and Gastrointestinal Function.* (2016) Annu Rev Physiol. Dec 15. doi:10.1146/annurev-physiol-021115-105447. [Epub ahead of print]

Amente S, Bertoni A, Morano A, Lania L, Avvedimento EV, Majello B. *LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc-induced transcription*. (2010) Oncogene. Jun 24;29 (25):3691-702

Amente S, Lania L, Avvedimento EV and Majello B. (2010) *DNA oxidation drives Myc mediated transcription*. Cell Cycle 9, 3002-3004

Balajee AS and Bohr VA. *Genomic heterogeneity of nucleotide excision repair.* (2000) Gene. May 30;250(1-2):15-30.

Bártová E, Krejčí J, Harničarová A, Galiová G and Kozubek S. *Histone Modifications and Nuclear Architecture: A Review*. (2008) J Histochem Cytochem. Volume 56(8): 711–721

Beato M, Wright RH and Vicent GP. (2015) *DNA damage and gene transcription: accident or necessity?* Cell Research 25:769–770.

Bennett RA, Wilson DM, Wong D, Demple B. *Interaction of human apurinic endonuclease and DNA polymerase beta in the base excision repair pathway*. (1997) Proc Natl Acad Sci U S A. Jul 8;94(14):7166-9

Bird AP and Wolffe AP. *Methylation-Induced Repression - Belts, Braces, and Chromatin.* (1999) Cell Vol. 99, 451–454, November 24

Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, Eeckhoute J, Brodsky AS, Keeton EK, Fertuck KC, Hall GF, Wang Q, Bekiranov S, Sementchenko V, Fox EA, Silver PA, Gingeras TR, Liu XS, Brown M. *Genome-wide analysis of estrogen receptor binding sites*. (2006) Nat Genet. Nov 38(11):1289-97

Castillo-Aguilera O, Depreux P, Halby L, Arimondo PB and Goossens L. *DNA Methylation Targeting: The DNMT/HMT Crosstalk Challenge*. (2017) Biomolecules 7, 3

Caston RA and Demple B. *Risky repair: DNA-protein crosslinks formed by mitochondrial base excision DNA repair enzymes acting on free radical lesions.* (2016) Free Radical Biology and Medicine dx.doi.org/10.1016/ j.freeradbiomed.2016.11.025 [Epub ahead of print]

Cheng X and Blumenthal RM. *Mammalian DNA Methyltransferases: A Structural Perspective*. (2008) Structure March 16(3): 341–350

Cho YH, Yazici H, Wu HC, Terry MB, Gonzalez K, Qu M, Dalay N and Santella RM. *Aberrant promoter hypermethylation and genomic hypomethylation in tumor, adjacent normal tissues and blood from breast cancer patients.* (2010) Anticancer Res. Jul;30(7):2489-96

Cieślik M and Bekiranov S. *Combinatorial epigenetic patterns as quantitative predictors of chromatin biology*. (2014) BMC Genomics 15:76

Cooper GM. *The Cell: A Molecular Approach. 2nd edition.* (2000) Sunderland (MA): Sinauer Associates

Culhane JC, Cole PA. *LSD1 and the chemistry of histone demethylation*. (2007) Curr Opin Chem Biol. Oct; 11(5):561-8

Cuozzo C, Porcellini A, Angrisano T, Morano A, Lee B, Di Pardo A, Messina S, Iuliano R, Fusco A, Santillo MR, Muller MT, Chiariotti L, Gottesman ME and Avvedimento EV. *DNA damage, homology-directed repair, and DNA methylation*. (2007) PLoS Genet. 3, e110

Deroo BJ and Korach KS. *Estrogen receptors and human disease*. (2006) J Clinical Invest. 116(3):561-70

Du J, Johnson LM, Jacobsen SE and Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. (2015) Nat Rev Mol Cell Biol. September 16 (9): 519–532

Dyrkheeva NS, Lebedeva NA and Lavrik OI. *AP Endonuclease 1 as a Key Enzyme in Repair of Apurinic/Apyrimidinic Sites.* (2016) Biochemistry (Moscow), Vol. 81, No. 9, pp. 951-967

ESCODD (European Standards Committee on Oxidative DNA Damage). *Measurement of DNA oxidation in human cells by chromatographic and enzymic methods.* (2003) Free Radical Biology and Medicine, Volume 34, Issue 8, 15 April, Pages 1089–1099

Farooq A. *Structural and Functional Diversity of Estrogen Receptor Ligands.* (2015) Curr Top Med Chem. 15(14): 1372–1384

Forneris F, Binda C, Vanoni MA, Mattevi A, Battaglioli E. *Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process.* (2005) FEBS Lett 579, 2203-2207

Geisberg JV and Struhl K. *Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo*. (2004) Nucleic Acids Res; 32(19): e151

Grünberg S and Hahn S. *Structural insights into transcription initiation by RNA polymerase II.* (2013) Trends Biochem Sci. Volume 38, Issue 12, p603–611

Hahn S. *Structure and mechanism of the RNA Polymerase II transcription machinery.* (2004) Nat Struct Mol Biol. May ; 11(5): 394–403

Hamilton JP. *Epigenetics: Principles and Practice.* (2011) Dig Dis;29:130–135

Harikumar A and Meshorer E. *Chromatin remodeling and bivalent histone modifications in embryonic stem cells*. (2015) EMBO Rep. Dec;16(12):1609-19

Horn AE, Kugel JF and Goodrich JA. *Single molecule microscopy reveals mechanistic insight into RNA polymerase II preinitiation complex assembly and transcriptional activity.* (2016) Nucleic Acids Res. Sep 6; 44(15): 7132–7143

Jaenisch R, Beard C, Lee J, Marahrens Y and Panning B. *Mammalian X chromosome inactivation.* Novartis Found Symp. 214:200-9; discussion 209-13, 228-32

Jin B, Li Y and Robertson KD. *DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy?* (2011) Genes Cancer. Jun 2; (6): 607–617

Jordan VC and O'Malley BW. *Selective Estrogen-Receptor Modulators and Anti-hormonal Resistance in Breast Cancer*. (2007) J Clin Oncol. Dec 20;25(36):5815-24

Juven-Gershon T and Kadonaga JT. *Regulation of Gene Expression via the Core Promoter and the Basal Transcriptional Machinery.* (2010) Dev Biol. March 15; 339(2): 225–229

Kiianitsa K and Maizels N. *A rapid and sensitive assay for DNA-protein covalent complexes in living cells.* (2013) Nucleic Acids Res. May; 41(9):e104

Klungland A and Lindahl T. *Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1).* (1997) EMBO J. Jun 2;16 (11):3341-8.

Klungland A, Rosewell I, Hollenbach S, Larsen E, Daly G, Epe B, Seeberg E, Lindahl T, Barnes DE. *Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage.* (1999) Proc Natl Acad Sci USA. Nov 9;96(23):13300-5

Kouzine F, Gupta A, Baranello L, Wojtowicz D, Ben-Aissa K, Liu J, Przytycka TM, Levens D. *Transcription-dependent dynamic supercoiling is a short-range genomic force.* (2013) Nat Struct Mol Biol. Mar;20(3):396-403

Kouzarides T. *Chromatin Modifications and Their Function*. (2007) Cell 128, 693–705, February 23

Le May N, Mota-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*. (2010) Mol Cell. Apr 9; 38(1):54-66

Li B, Carey M and Workman JL. *The Role of Chromatin during Transcription.* (2007) Cell 128, 707–719

Li E, Beard C and Jaenisch R. *Role for DNA methylation in genomic imprinting.* (1993) Nature 366, 362-365, 25 December

Li E, Bestor TH and Jaenisch R. *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* (1992) Cell Jun 12;69(6):915-26

Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, Oh S, Kim HS, Glass CK, Rosenfeld MG. *Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation.* (2013) Nature Jun 27;498(7455):516-20

Li Z and Kono H. *Distinct Roles of Histone H3 and H2A Tails in Nucleosome Stability*. (2016) Scientific Reports 6:31437

Lin CY, Vega VB, Thomsen JS, Zhang T, Kong SY, Xie M, Chiu KP, Lipovich L, Barnett DH, Stossi F, Yeo A, George J, Kuznetsov VA, Lee YK, Charn TH, Palanisamy N, Miller LD, Cheung E, Katzenellenbogen BS, Ruan Y, Bourque G, Wei KL, Liu ET. *Whole-Genome Cartography of Estrogen Receptor α Binding Sites*. (2007) PLoS Genet. Jun 3(6): e87

Lindahl T and Wood RD. *Quality control by DNA repair*. (1999) Science Dec 3; 286(5446):1897-905

Melis JPM, Van Steeg H and Luijten M. *Oxidative DNA Damage and Nucleotide Excision Repair*. (2013) Antioxidants & Redox Signaling. Jun, Vol. 18, No. 18: 2409-2419

Métivier R, Gallais R, Tiffoche C, Le Péron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G. *Cyclical DNA methylation of a transcriptionally active promoter*. (2008) Nature Mar 6;452(7183):45-50

Morano A, Angrisano T, Russo G, Landi R, Pezone A, Bartollino S, Zuchegna C, Babbio F, Bonapace IM, Allen B, Muller MT, Chiariotti L, Gottesman ME, Porcellini A and Avvedimento EV. *Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene.* (2014) Nucleic Acids Res. 42, 804-821

Morera L, Lübbert M and Jung M. *Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy.* (2016) Clin Epigenetics. 8:57

Morera L, Lübbert M, Jung M. *Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy.* (2016) Clin Epigenetics. May 24;8:57

Nilsson S, Mäkelä S, Treuter E, Tujague M, Thomsen J, Andersson G, Enmark E, Pettersson K, Warner M, Gustafsson JA. *Mechanisms of estrogen action*. (2001) Physiol Rev. 81(4):1535-65

Nottke A, Colaiácovo LP and Yang Shi. *Developmental roles of the histone lysine demethylases.* (2009) Development 136, 879-889

Park JH, Lee MY, Han HJ. *A potential role for caveolin-1 in estradiol-17beta-induced proliferation of mouse embryonic stem cells: involvement of Src, PI3K/Akt, and MAPKs pathways*. Int J Biochem Cell Biol. Mar; 41(3):659-65

Perillo B, Sasso A, Abbondanza C, Palumbo G. *17b-Estradiol inhibits apoptosis in MCF-7 cells, inducing bcl-2 expression via two Estrogen Responsive Elements present in the coding sequence.* (2000) Mol Cell Biol. Apr;20(8): 2890–2901

Porrua O and Libri D. *Transcription termination and the control of the transcriptome: why, where and how to stop.* (2015) Nature Reviews Molecular Cell Biology 16, 190–202

Semlow DR, Zhang J, Budzowska M, Drohat AC, Walter JC. *Replication-Dependent Unhooking of DNA Interstrand Cross-Links by the NEIL3 Glycosylase.* (2016) Cell. Oct 6; 167(2):498-511.e14

Sengupta S and Jordan VC. *Selective estrogen modulators as an anticancer tool: mechanisms of efficiency and resistance*. (2008) Adv Exp Med Biol;630:206-19

Stork CT, Bocek M, Crossley MP, Sollier J, Sanz LA, Chédin F, Swigut T, Cimprich KA. *Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage*. (2016) Elife. Aug 23;5

Stresemann C and Lyko F. *Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine.* (2008) Int J Cancer. Jul 1;123(1):8-13

Sung JS and Demple B. *Roles of base excision repair subpathways in correcting oxidized abasic sites in DNA.* (2006) FEBS J. Apr;273(8):1620-9

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L and Rao A. *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1*. (2009) Science 324(5929):930-5

Virani S, Colacino JA, Kim JH, and Rozek LS. *Cancer Epigenetics: A Brief Review*. (2012) ILAR Journal Volume 53, Number 3/4

Wong WM, Poulsom R, Wright NA. *Trefoil peptides.* (1999) Gut. Jun; 44(6):890-5.

Yager JD and Davidson NE. *Estrogen carcinogenesis in breast cancer.* (2006) N Engl J Med. Jan 19;354(3):270-82

Yang S, Zhang J, Zhang Y, Wan X, Zhang C, Huang X, Huang W, Pu H, Pei C, Wu H, Huang Y, Huang S, Li Y. *KDM1A triggers androgen-induced miRNA transcription via H3K4me2 demethylation and DNA oxidation*. (2015) Prostate. Jun 15;75(9):936-46

Yi X, Jiang XJ, Li XY, Jiang DS. *Histone methyltransferases: novel targets for tumor and developmental defects*. (2015) Am J Transl Res; 7(11):2159-2175

Zaidi SK, Young DW, Montecino M, Lian JB, Stein JL, van Wijnen AJ, Stein GS. *Architectural epigenetics: mitotic retention of mammalian transcriptional regulatory information.* (2010) Mol Cell Biol. Oct 30(20):4758-66

Zuchegna C, Aceto F, Bertoni A, Romano A, Perillo B, Laccetti P, Gottesman ME, Avvedimento EV, Porcellini A. *Mechanism of retinoic acid-induced transcription: histone code, DNA oxidation and formation of chromatin loops*. (2014) Nucleic Acids Res. 42(17):11040-55

8. ACKNOWLEDGEMENTS

At the beginning of this amazing adventure called PhD, I never expected I would have so many people to thank.
I am grateful to each person
who began this journey with me,
who experienced with me the joy of the discoveries and the pain of the failures,
who entered suddenly in my life giving me only good advices and sincere friendship,
who stayed with me and never left me also from a long time before the beginning of the PhD program.

First, I would like to thank my mentor and supervisor Professor Vittorio Enrico Avvedimento, for the inestimable opportunity he gave me to attend the PhD in his research group and to learn how to intensely approach to science, being conscious that “you always have something more to understand!”.

All my gratitude and my respect is for Professor Bruce Demple, who hosted me in his laboratory at Stony Brook University, NY, involving me in interesting projects and showing me how important is personal initiative to be a good-minded scientist.

I would also thank Professor Lucio Nitsch for the great dedication he has for PhD students, and for having always interacted with me, when I was a PhD student representative, to improve the conditions of students and educational activities.

To Dr. Antonio Pezone goes all my admiration, for the great example he gave me of being a passionate researcher, and all my gratitude for helping me and working together.

I want to thank Dr. Rosaria Landi for all the support, the inestimable advices, the beautiful moments spent together inside and outside the biological tower; for all that times she saved me at the very last moment, wherever I was, whatever was the problem, of daily life or scientific matter.

My complete gratefulness is for Dr. Imma Castaldo, who allowed me to work on interesting new topics and, above all, she has been an indispensable and motivational presence in my life during this last year.

To all my crazy, funny, lovely friends and colleagues of VEA laboratory: without any order of importance I want to thank each of you because without your support, help and presence, these years of PhD would have been darker and even more difficult to face. So, thank you Dr. Alfonso Tramontano, Dr. Federica Barra, Dr. Jacopo Dolcini, Dr. Giuliana La Rosa, Dr. Francesca Boffo and Dr.

Giusi Russo for each single day you thought to send my “somewhere” but you decided to continue loving me and working with me side by side.

To all the amazing people of Feliciello’s Lab goes my thankfulness, for their friendship and their precious suggestions and encouragements.

Lastly, the final and far from least important acknowledgements are for my family and my friends, because without their support I would never be able to become a PhD and the woman I am today.

Thanks for everything to my parents, because they always support my choices and deserve for me all the possible existing love, even if I am far from home and I do not have the time they deserve to stay together.

Thank you Fabio for the respect you have for my choices, even if they influence also your and our lives, and for the limitless love and support you donate me every day.

Thanks to my brother and sister and my brother and sister in law, for the inspiration and encouragement they always offer to me.

Thank you to all my near and far friends, because your presence in my life helped me to achieve this goal, at least with a smile on my face and with more confidence in myself.