UNIVERSITY OF NAPLES "FEDERICO II"

Doctorate School in Biology

Ph.D. COURSE IN BIOLOGY

XXIX CYCLE

DNA looping, histones and DNA methylation are coupled by DNA oxidation to modulate transcription

Coordinator

Prof. Salvatore Cozzolino

Supervisor

Prof. Antonio Porcellini

Ph.D. Student

Dr. Antonella Romano

Academic Year 2015-2016

Table of contents

Declaration

 I declare that this thesis is my own work except where it contains work based on collaborative research my contribution was as follows:

- *Mechanism of retinoic acid-induced transcription: histone code, DNA oxidation and formation of chromatin loops*: Silencing of JMJD2A and histone code of silenced cells; ChIP analisy of H3K4me2/me3, H3K9me2/me3, H3K9Ac, LSD1, JMJD2A, NCoR1, NCoR2 and SUV39H1.

- *DNA damage and Repair Modify DNA methylation and Chromatin Domain of the Targeted Locus: Mechanism of allele methylation polymorphism*: ChIP on histone code after 24h, 48h, 7 and 14 days from DBS in Rec H and Rec L cells; 3C analisy.

- *Glycolysis controls the induction of human regulatory T cells by modulating the expression of FOXP3 exon 2 splicing variant*: Splicing variants of FOXP3 mRNA in iTreg-CTR, iTreg-2DG and iTreg-Etx cells; recruitment of enolase-1 to the promoter and CNS2 of FOXP3 in iTreg-CTR, iTreg-2DG and iTreg-Etx cells; Quantitative real-time PCR analysis of all FOXP3 transcripts and FOXP3-E2 mRNA in iTreg-CTR and iTreg-2DG cells silenced for ENO1.

All materials contained in this thesis, which have been derived from the published or unpublished work of others, are acknowledged.

Abstract

Changes of histone methylation code, DNA hydroxymethylation and formation of chromatin loops are associated with transcription induction by nuclear hormones, but it is not known if these events are the consequence or the cause of transcription initiation. We studied the effect of activation of the retinoic acid receptor, at the RARE–promoter chromatin of CASP9 and CYP26A1 genes, and estrogen receptor α, at ERE and Poly_A of PS2, CAV1 and BCL2 genes. We found that histone H3 lysines 4 and 9 are demethylated by the lysine-specific demethylase, LSD1 and by the JMJ-domain containing demethylase, D2A. The action of the oxidase (LSD1) and a dioxygenase (JMJD2A) in the presence of Fe++ elicits an oxidation wave that locally oxidate the DNA, recruits the enzymes involved in base and nucleotide excision repair (BER and NER) and bring to formation of chromatin loops that juxtapose the 5' and 3' ends of induced genes. In addition we evaluated the crucial role of DNA methyltransferase 3a that work closely to OGG1 and APE1, favoring strand specific repair of oxo-dG and methylation of the complementary C preventing the accumulation of G to T mutations at regulative regions of studied genes. We suggest that coupling transcription with methylation and repair enzymes is an evolutionary strategy to reduce the mutational burden induced by DNA oxidation necessary for chromatin looping. We characterized the changes in chromatin structure, DNA methylation, the recruitment of Base Excision Repair during and after homologous DNA repair (HR) in HeLa DR-GFP cells. Repaired genes display stable but different methylation profiles. These profiles manage the levels of expression in generated recombinant populations. Our data argue that DNA methylation and chromatin remodelling induced by HR may be a source of permanent variation of gene expression in somatic cells. Finally we studied the relationship between chromatin remodeling of FOXP3 gene and immunophenotypic differentiation of human primary $CD4^+$ T lymphocytes moreover the altered immunosuppressive activity in autoimmune desease. We found that the induction and suppressive function of iTreg cells (induced Treg cells) tightly depended on glycolysis, which controlled FOXP3 splicing variants containing exon 2 (FOXP3-E2) through the glycolytic enzyme enolase-1. The FOXP3-E2–related suppressive activity of iTreg cells was altered in human autoimmune diseases, including multiple sclerosis and type 1 diabetes, and was associated with impaired glycolysis and signaling via interleukin 2. This link between glycolysis and FOXP3-E2 variants via enolase-1 shows a previously unknown mechanism for controlling the induction and function of Treg cells in health and in autoimmunity.

1

Riassunto

I cambiamenti del codice istonico, l'idrossimetilazione del DNA e la formazione di loops cromatinici sono associati alla trascrizione dipendente da ormoni nucleari; non è ben chiaro se questi eventi siano la conseguenza o la causa dell'inizio del processo trascrizionale. Abbiamo studiato gli effetti dell'attivazione trascrizionale dipendente dal recettore dell'acido retinoico sulle regioni RARE-promotore dei geni CASP9 e CYP26A1, e del recettore degli estrogeni α sulle regioni ERE e Poly_A dei geni PS2, CAV1 and BCL2. I risultati ottenuti hanno evidenziato che dopo stimolazione ormonale la lisina 4 e 9 dell'istone H3 sono demetilate dalla demetilasi lisina specifica LSD1 e dalla demetilasi contenente il dominio JMJ: JMJD2A. L'azione dell'ossidasi (LSD1) e della diossigenasi (JMJD2A) genera specie reattive dell'ossigeno che ossidano localmente il DNA, reclutano gli enzimi dei complessi di riparo per escissione di base e nucleotidica (BER e NER) e portano alla formazione di loops cromatinici che giustappongono l'estremità 5' e 3' dei geni indotti. Abbiamo inoltre valutato il ruolo cruciale della DNA metiltransferasi 3A (DNMT3A) che, lavorando a stretto contatto con OGG1 ed APE1 favorisce il riparo filamento specifico dell'8-oxoG e la metilazione della C complementare prevenendo così la transversione della G in T quindi, all'accumulo di mutazioni nelle regioni regolative dei geni attivi. In seconda analisi abbiamo valutato i cambiamenti della struttura cromatinica, della metilazione del DNA e del reclutamento degli enzimi di riparo (BER) conseguenti la rottura a doppio filamento del DNA nel sistema pioniere DR-GFP (cellule HeLa DR-GFP) in due popolazioni ricombinanti (cloni Rec H e Rec L) generate in seguito ad homologous recombination (HR). I dati ottenuti evidenziano che la metilazione del DNA ed rimodellamento della cromatina indotta da HR possono essere una fonte permanente della variazione dell'espressione genica nelle cellule somatiche. Infine abbiamo studiato la relazione tra rimodellamento cromatinico del gene FOXP3 e differenziazione immunofenotipica di cellule primarie umane (linfociti T CD4[+]) e l'alterata attività immunosoppressiva in individui affetti da patologie autoimmuni come la sclerosi multipla. I risultati ottenuti mostrano che la funzione immunosoppressiva delle iTreg (cellule Treg indotte) dipende dalla glicolisi, che controlla attraverso l'enzima enolasi-1 l'espressione delle varianti di splicing contenenti l'esone 2 (FOXP3-E2). Il legame tra la glicolisi e trascritti FOXP3-E2 dipendente da enolasi-1 mostra un meccanismo sconosciuto che controlla l'induzione e la funzione delle cellule Treg nella autoimmunità.

Chapter 1: Introduction

The chromatin structure

DNA, deoxyribonucleic acid, is a long linear polymer in which the genetic information is contained. It consists of four monomeric units: the deoxyribonucleotides, composed by nucleotide bases, a sugar with five carbon atoms (2-deoxyribose) and a phosphate group (PO_4^{2-}). The nucleotide base is linked through a glycosidic bond to the C_1 position of sugar, while the phosphate group (PO_4^{2-}) is bound through phosphodiester bond to the C_5 position of 2-deoxyribose. There are two different types of nucleobases: the purine bases [adenine (A) and guanine (G)], and the pyrimidine bases [cytosine (C) and guanine (G)]; that are paired together (A-T and C-G) to give DNA its double helical shape.

Eukaryotic cells package their genetic information into chromatin, a nucleoprotein complex composed of the primary genetic information, the DNA, and a mix of structural and regulatory proteins; the complex is tightly bound by attraction of the negatively charged DNA to the positively charged histones (Nick Gilbert et al., 2005). In mammalian cells we can distinguish three levels of chromatin organization: the nucleosome, the 30 nm chromatin fiber and the chromosomes (the highest level of DNA compaction) (Fig.1). Chromatin is a dynamically entity that contains the regulatory signals necessary to program appropriate cellular pathways and it is believed to contribute to the control of genes expression. Genetic information encoded in DNA is largely identical in every cell of a eukaryote. However,

Figura 1 Schematic representation of chromatin structure. The chromatin fiber presents in nucleo is composed by nucleosome units, each consisting of DNA wrapped around histone proteins proteins.

cells in different tissues and organs can have widely different gene expression patterns and can exhibit specialized functions. Gene expression in different cell types needs to be appropriately induced and maintained; inappropriate expression patterns lead to disease (Greer E. L. & Shi Y., 2012).

Epigenetics

Epigenetics is defined as the study of alteration in gene expression without any changes in DNA sequence that are transmitted through both mitotis and meiosis (Wu C. & Morris J. R., 2001). The term is made of two parts: Greek prefix "epi", which means "above" or "over" and "genetics", which is the science of genes. The word "*epigenetics*" was coined by Waddington to define the changes in phenotype without involve change in genotype. Epigenetic regulation in mammals includes three major mechanisms: DNA methylation, histone variants and histone posttranslational modification (PTMs), and noncoding RNAs (Xu W. et al., 2016) (Fig.2). DNA methylation plays an important role in genomic imprinting, X-chromosome inactivation during embryonic development in a several diseases such as cancer; Histone modification and variant histones influence gene regulation through changes in nuclear chromatin structure; noncoding RNAs knock down gene

Figura 2 Mechanisms of epigenetic modification. Cell transduces the environmental changes (e.g., nutrients, toxins, etc.) into epigenetic modifications including DNA methylation, histone PTMs, and noncoding RNAs (especially the microRNAs), which eventually turn certain genes on or off and regulate the transcription process without changing DNA sequence. Xu W. et al., 2016)

expression through different mechanisms such as microRNA, siRNA pathway, RNA-mediated interference (RNAi) or post-trascriptional gene silencing (PTGS). Epigenetic mechanisms work in addition to the DNA template to stabilize gene expression programmes defining cell-type identity (Allis C. D. & Jenuwein T., 2016) and the interaction between different mechanisms controls the accessibility of genes by the transcriptional machinery.

Histone proteins

Histones are the most abundant proteins associated with DNA. They are small basic proteins, whose side chains are rich in positive charged basic amino acids, in particular lysine and arginine; this positive charge allows the histones to interact in a firm and non-specific with the negative charges of the DNA phosphate groups.

In eukaryotes we can identify five types of histones denominated H1, H2A, H2B, H3 and H4, which differ in molecular weight and content in lysine and arginine. Histones are associated with DNA to form nucleosomes, the fundamental units of compaction of chromatin, which give it an appearance of "pearl necklace" (Alberts et al., 2004). Each nucleosome consists of a protein octamer defined histone core, around which DNA is wrapped for one turn and 3/4; the quantity of DNA associated with the histone octamer is standard and consists of 147 base

pairs (bp). The histone octamer is formed by a central H3/H4 tetramer (associated to 60 pairs of central bases of nucleosomal DNA) sandwiched between two H2A/H2B dimers. (Fig. 3).

Each nucleosome is separated from the adjacent one by a region of "linker DNA", which length is highly variable among species and can go from 10 to 90 bp. This DNA sequence is linked to H1 histone; this histone consists of a central globular
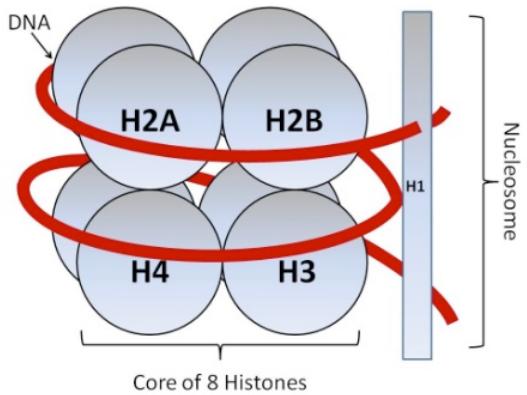
Figura 3 Schematic representation of nucleosome structure.

region linked to amino and carboxy terms. Each molecule H1 binds a single site on a nucleosome with its globular domain and its terms is in connection with adjacent nucleosomes sites, allowing the transition to a next level of organization called "the 30-nm fiber" or "solenoid structure ". DNA can be further compacted by the formation of loops locked to the base by a protein structure known as "nuclear scaffolds" which form the highest level of condensation of chromatin: the chromosome.

Histones are tripartite proteins that are composed of a globular domain (**histone fold**) and unstructured N- or C-terminal tails that are subjected to several covalent modifications such as methylation, acetylation, phosphorylation as well as addition of large groups like ubiquitin and ADP-ribose (Kouzarides T., 2007). These modifications work acting in a combinatorial or sequential manner on one or multiple histone tails, to build a real code known as "**histone code**" (Fig. 4), that contribute to the control of genes expression and influence chromatin compaction or signaling to other protein complexes (Strahl & Allis, 2000). Therefore, an appropriate balance of stability and dynamics in histone post-translational modifications (PMTs) is necessary for accurate gene expression. Chromatin structure or landscape is a composite of various domains characterized by the local enrichment of a specific combination

of histone post-translational modifications (PMTs), histone variants, nucleosome occupancy, DNA methylation patterns and nuclear localization.
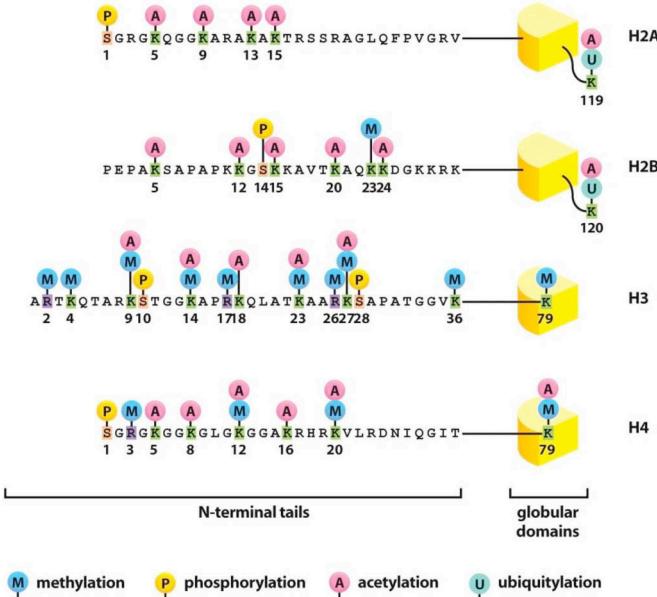
Figura 4 Histone posttranslational modifications. Histone modifications are covalent post-translational modifications that take place on histone tails. They operate in combined manner to generate a code (histone code) and influence either 'active' or 'inactive' chromatin states

The distributions of histone PTMs form a indicative signature of the chromatin state of a given loci (Turner B.M., 1993) Euchromatin is generally associated with high levels of histone acetylation. H3K9ac and H4K16ac are found at the promoters of actively transcribed genes while H3K27ac is enriched at enhancers together with another histone modification H3K4me1 (Wang Z. et al., 2008; Pasini D. et al., 2010). Di- and tri-methylations of histone H3 at K4 are modifications which are found in close proximity to active promoters (Santos-Rosa H et al., 2002).

On the other hand, constitutive heterochromatin is enriched in repressive marks such as H3K9me3 and H4K20me3 which are deposited in mammals by Suv39-H1/2 and Suv4-20H1/2 respectively (Rice J.C. et al., 2002). The characteristic mark of facultative heterochromatin is H3K27me3. Polycomb repressive complex 2 (PRC2) methylates histone H3 at K27, both at promoters and gene bodies, via its catalytic subunit EZH1/2.

Histone methylation

Methylation is a post-translational modification of histone tails which occurs on all basic residues: lysine, arginine and histidines. It is catalyzed by enzymes (histone

methyltransferases) that use S-adenosyl-L-methionine (SAM) as the methyl group donor. Lysines can be monomethylated (me1), dimethylated (me2), or trimethylated (me3) on their ε-amine group; arginines can be mono or while histidines can be only monomethylated. The most studied histone methylation sites include histone H3 lysine 4 (H3K4), K9 , K36, K79 and H4K20. Sites of arginine (R) methylation include H3R2, H3R8, H3R17, H3R26 and H4R3.

The level of histone methylation is essential for regulation of transcriptional activity, in fact methylation can influence gene expression in both positive and negative way. Repressive marks include di- and trimethylation of lysine 9 on histone H3 (H3K9 me2/me3), trimethylation of lysine 27 on histone H3 (H3K27me3) and trimethylation of lysine 20 on histone H4 (H4K20me3). Conversely, activation is associated with histone H3 di- and trimethylation of lysine 4 (H3K4me2/me3) and trimethylation of lysines 36 and 79 (H3K36me3, H3K79me3).

Methyl groups can be removed from histone lysine residues by the action of histone lysine demethylases. Two families of these enzymes have been identified that remove methyl groups from different lysine residues on histones. These are the amine oxidases (LSD1) (Shi et al., 2004) and jumonji C (JmjC)-domain-containing, iron-dependent dioxygenases (Tsukada Y. et al., 2006; Whetstine J. R. et al., 2006). Methylated histones are recognized by chromatin effector molecules ('readers'), causing the recruitment of other molecules to alter the chromatin and/or transcription states (Taverna S. D. et al., 2007). These chromatin-regulating complexes include histone acetyltransferases (HATs), DNA methyltransferases (DNMTs), histone deacetylases (HDACs), histone methyl-transferases (HMTs) and histone demethylases (HDMs) that act in a stepwise and/or combinatorial manner and engage in extensive cross talk.

DNA methylation and hydroxymethylation

DNA methylation is a covalent postreplicative modification of genomic DNA. The process consists in the addition of the methyl group at the 5-carbon of the cytosine ring resulting in 5-methylcytosine (**5-mC**), also known as the "fifth base" of DNA in the context of CpG dinucleotide, in which a cytosine nucleotide is located next to a guanidine nucleotide.

In the mammalian genome, the CpG dinucleotides occur in CG dense regions termed "CpG islands", whose length is generally between 300 and 3,000 bp; these are prevalent at

transcription start sites of housekeeping and developmental regulator genes (Deaton et al., 2011), where they are ipomethylated. The degree DNA methylation of CpG Island regulates genes expression; in fact DNA can directly interfere with transcription by preventing binding of basal transcriptional machinery or ubiquitous trancription factors that require contact with cytosine in the major groove of the double helix (Cuozzo C. et al., 2007). This covalent modification regulates different biological process such as development and differentiation (Li E. et al., 1992), imprinting (Li E. et al., 1993), X chromosome inactivation (Panning B. & Jaenisch R., 1998), and cancer (Laird P.W. et al., 1995; Baylin S.B. & Ohm J.E., 2006). Aberrant methylation has been found in cancer cells (Cho Y.H. et al., 2010) and it was shown that DNA methylation is associated with DNA damage and repair (Cuozzo C. et al., 2007) and that methylation is reduced by

Figura 5 DNA methylation pathways. A family of DNA methyltransferases (DNMTs) catalyzes the transfer of a methyl group from S-adenyl methionine (SAM) to the fifth carbon of cytosine residue to form 5-methylcytosine (5mC). DNMT1 is the maintenance DNMT and maintains DNA methylation pattern during replication. (Moore D. 2013)

transcription of the repaired regions as a mechanism of adaptation to environmental perturbations (Morano A. et al., 2014).

DNA methylation is catalyzed by a family of DNA methyltransferases (DNMTs): DNMT1, DNMT2, DNMT3A, DNMT3B, and DNMT3L (Bestor T. H. et al., 2000). DNMT1 is highly expressed in mammalian tissues including the brain (Goto K. et al., 1994) it preferentially methylates hemimethylated DNA (Pradhan S. et al., 1999; Ramsahoye B. H. et al., 2000). During DNA replication, DNMT1 is localized near the replication fork where newly synthesized hemimethylated DNA is formed (Leonhardt H. et al., 1992). DNMT1 binds the newly synthesized DNA and methylates it to mimic the original methylation pattern present before DNA replication (Hermann A. et al., 2004) (Fig. 5). For this reason, DNMT1 is called the maintenance DNMT because it maintains the original pattern of DNA methylation in a cell lineage. DNMT3A and DNMT3B, in contrast to DNMT1, have preference for unmethylated CpG dinucleotides and perform de novo methylation during development. The difference between DNMT3a and DNMT3b is DNMT3a gene expression pattern, in fact it is
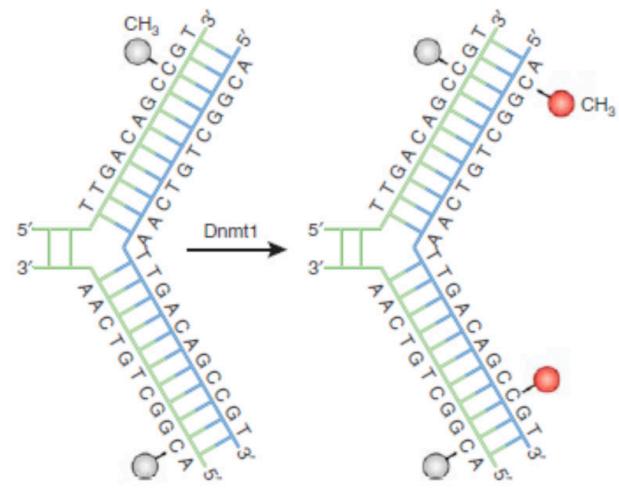
expressed relatively ubiquitously, DNMT3b is poorly expressed by the majority of differentiated tissues. However, mounting evidence indicates that DNMT1 may also be required for de novo methylation of genomic DNA (Egger G. et al., 2006) and that DNMT3A and DNMT3B contribute to maintenance methylation during replication. (Riggs A. D. et al., 2004). DNMT3L has no catalytic function, it assists de novo methyltransferases by increasing their ability to bind to the methyl group donor, S-adenosyl-Lmethionine (SAM), and stimulating their activity in vivo (Kareta M. S. et al., 2006).

Figura 6 Epigenetic crosstalk. Transcription is regulated by the interaction of multiple epigenetic mechanisms that cooperate to activate or silence gene expression. To suppress gene expression, DNMTs target CpG sites and actively methylate DNA. Their catalytic activity is enhanced by association with with DNMT3L. DNA methylation is recognized by methyl-binding proteins such as MBDs that recruit enzymes that modify the histone tails such as histone deacetylases (HDACs), which remove acetylation, and histone methyltransferases (HMTs), which methylate histones and in conjunction with DNA methylation repress gene expression. In regions of DNA with activate transcription, Tet removes DNA methylation, and histone tails in this region often contain H3K4me3 that inhibits DNMT binding to unmethylated CpG sites and maintains a permissive environment for transcription. (Moore D. 2013)

DNA methylation works with histone modifications and microRNA (miRNA) to regulate transcription (Fig. 6). These modifications influence not only DNA packaging, but also their transcriptional activity, in fact DNMTs interact with enzymes that regulate histone modification involved in gene repression.

Methylation inhibits transcription through the action of a protein, methyl-CpG-binding protein 2 (MECP2), that specifically binds methylated DNA and represses transcription (Lewis J. D. et al., 1992). Interestingly, MeCP2 functions as a complex with histone deacetylase, linking DNA methylation to alterations in histone acetylation and nucleosome structure. DNA methylation is directly connected with histone methylation, it is inversely correlated with H3K4 trimethylation, in fact this histone modification impairs the binding of DNMT3a, DNMT3b and DNMT3L to H3 histone tails and prevents methylation (Zhang Y. et
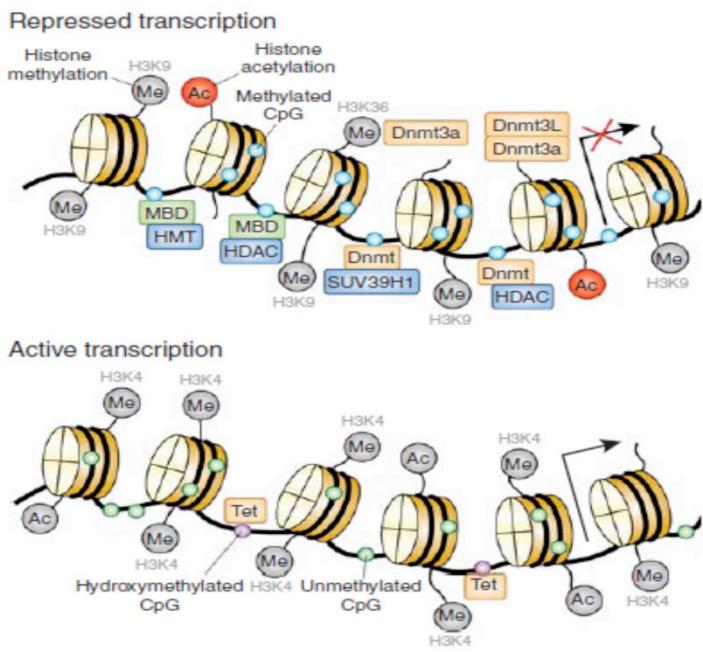
al., 2010); but it is directly correlated with repressive histone mark, in fact DNMT1 and DNMT3a bind the histone methyltransferase SUV39H1 that restricts gene expression by methylation on H3K9 (Fuks F. et al., 2003). In addition, DNMT1 and DNMT3b can bind to histone deacetylases that remove acetylation from histones to compact DNA and restrict access for transcription (Fuks F. et al., 2000; Geiman T. M. et al., 2004).

DNA methylation was assumed to be a stable, irreversible epigenetic modification associated with gene repression, but the discovery of TET proteins shows that DNA can be hydroxymethylated and demethylated. The DNA demethylation process is important for epigenetic reprogramming of genes and is also directly involved in many important disease mechanisms such as tumor progression. It can either be passive or active, or a combination of both. Passive DNA demethylation occurs during DNA replication in dividing cells, when maintenance of DNA methylation is abrogated through down regulation of DNMT1 because this DNA methyltransferase protein methylate DNA emi-methylated not emi-hidroxymethylated, (Takahashia S. et al., 2015), it involves a gradual loss of 5mC in a dilution manner. As an alternative mechanism, active DNA demethylation removes 5mC in a replication-independent manner, in particular the ten eleven translocation protein family (TETs) catalize the oxidation of 5-methylcytosine in 5-hydroxymethilcytosine. This proteins family includes TET1, TET2 and TET3. These proteins may promote DNA demethylation by binding CpG rich regions to prevent unwanted DNA methyltransferase activity, and by converting 5-mC to 5-hmC (5- hydroxymethylcytosine), 5-hmC to 5-fC (5-formylcytosine), and 5-fC to 5-caC (5-carboxylcytosine) through hydroxylase activity (Ito S. et al., 2011). The TET proteins have been shown to
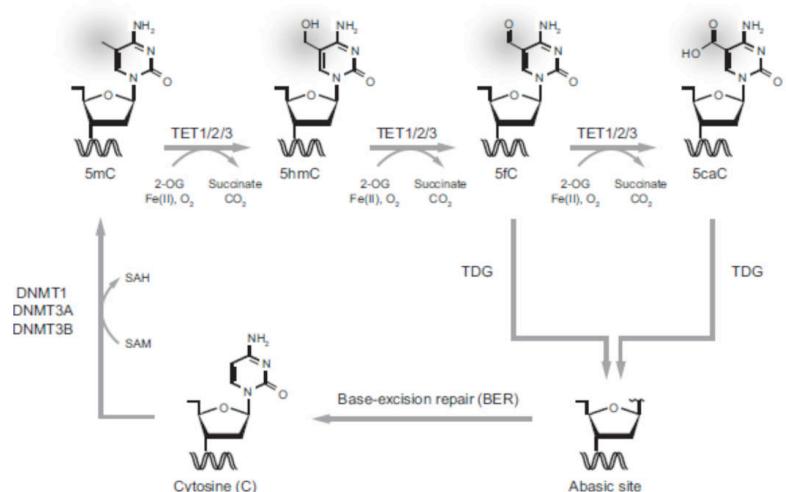
Figura 7 Pathways of DNA demethylation mediated by TET enzymes. Model of active DNA demethylation by a TET/TDG-dependent pathway. A cytosine base can be methylated by the DNA methylation machinery(DNMT1or DNMT3A/B) to form 5mC, which in turn can be oxidized by TET enzymes to produce 5hmC, 5fC, and 5caC. TDG then recognizes 5fC and 5caC, and the oxidized cytosine base is excised. The site is repaired by BER enzymes that restored the unmodified cytosine state. (Rasmussen K. D. and Helin K. 2016)

function in transcriptional activation and repression (TET1), tumor suppression (TET2), and DNA methylation reprogramming processes (TET3).

The oxidative cytosine derivates from 5-fC and 5-caC are recognized and excised by thymine DNA glycosylase (**TDG**) to allow subsequent base excision repair (**BER**) processing which converts modified cytosine back to its unmodified state (Yamaguchi S. et al., 2013) (Fig. 7)

DNA damage

DNA is the holder of all the genetic information contained in every living cell, so it is essential that the integrity and stability are constantly safeguarded to allow cells to survive. However, DNA is not an inert chemical molecular entity but it is subject to a variety of aggressions responsable of many damages. The latter may involve directly or indirectly causing a break in the DNA or mutations which, if not repaired, can be transmitted to the next cellular generation, with a consequent accumulation of damage. The process of replication of DNA during cell division, is subject to numerous errors caused essentially by the speed of DNA polymerase which incorporates incorrect nucleotides during synthesis of the new strand. Most of these errors is repaired by a repair system of its own DNA polymerase which has proof reading activity because it is equipped with exonuclease 3'-5 'able to recognize and remove nucleotides wrongly matched. Some of these errors may, however, escape the control of exonuclease, causing a incorrect pairing between the two strands that if not identified induces, in the next cycle of replication, a permanent change in the DNA sequence, determining the onset of a mutation.

Maintaining the integrity of genetic information is critical both for normal cellular functions and for suppressing mutagenic events that can lead to cancer. DNA damage can arise from external sources such as exposure to ionizing radiation, ultraviolet radiation (UV), or environmental toxins, or for endogenous sources, such as reactive oxygen species or errors during DNA replication. These events can generate DNA lesions, including modified bases, abasic sites, the formation of DNA adducts, the production of single- (SSBs) and double-strand breaks (DSBs) (Lindahl T. et al., 1993). Persistent DNA damage can induce mutagenesis, such as base substitutions and small insertions/deletions, as well as chromosomal rearrangements. Thus, constant genome maintenance is essential for the viability and longevity of a healthy organism. Cells have therefore evolved multiple DNA repair pathways to preserve genome integrity when damage arises.

Oxidative damage

DNA is constantly attacked by reactive oxygen species generated during normal cellular metabolism and exposure to ionizing or ultraviolet radiation (1-3 borh). The oxidative stress is the physiological disturbance between the reactive oxygen species (ROS), such as hydrogen peroxide (H_2O_2), superoxide (O_2^-) and hydroxyl radicals (OH^-) (and the ability to remove them it can also be defined as the disordered redox signaling and control (Jones, 2006). The reactive oxygen species play an important role in cell signaling and regulation of cytokine, growth factor and hormone action, transcription, ion transport, neuromodulation, immunemodulation, and apoptosis (Gloire G. et al., 2006; Mittler R. et al., 2011); and also play a fundamental role in normal functioning of immune system, proliferation of T cells, and immunological defence (Devadas S. et al., 2002; Hildeman D. A. et al., 2004). The formation of ROS is prevented, by an effective system of antioxidants such as ascorbic acid, glutathione, enzymes such as superoxide dismutase (SOD), catalase and peroxidase that interacts directly with ROS, neutralizing them. One of the most deleterious consequences of oxidative stress is the formation of DNA lesions as bases oxidation. Both purine and pyrimidine bases are subject to oxidation. The most common mutation is guanine oxidized to 8-oxo-7,8-dihydroguanine, resulting in the nucleotide 8-oxo-deoxy-guanosine (8-oxo-dG). When 8-oxo-dG assumes its syn conformation is able to bind deoxyadenosine, instead of pairing with deoxycytotidine as expected, because it can mimic thymine (Fig. 8).

Figura 8 The base pairing possibilities of 8-oxoG (8oG). In an *anti* conformation it forms a Watson-Crick base pair with dCTP (a); by assuming a *syn* conformation, it can form a Hoogsteen base pair with dATP (b). (Wang Y. and Schlick T., 2006)

If this error is not detected and corrected by mismatch repair enzymes, the DNA subsequently replicated will contain a C→A point mutation. When 8-oxoG is inserted during DNA replication, it can generate double-strand breaks, very deleterious lesion (Cheng K. C. et al., 1992). The mutagenic and genotoxic
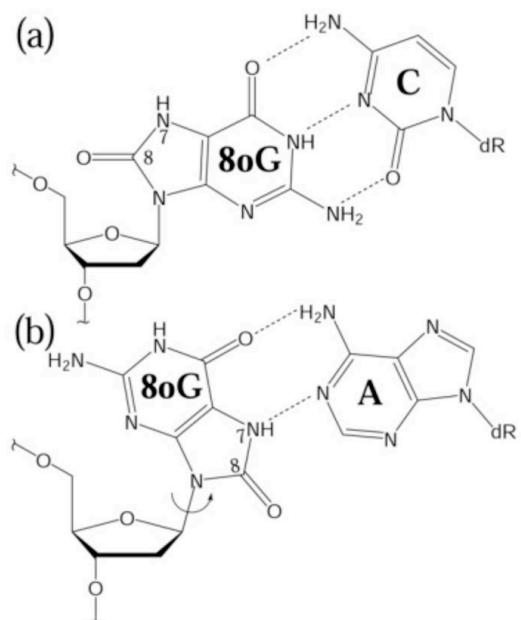
lesions are removed from DNA by mechanisms such as the base excision DNA repair (BER) and nucleotide excision repair (NER).

Base excision repair (BER)

Base excision repair (BER) is a highly conserved pathway from bacteria to humans and it is responsible of repairing the vast majority of endogenous DNA damage including alkylations, oxidations, deaminations and depurinations, as well as single-strand breaks (SSBs) (Krokan H. E. et al., 2013). Thus, the primary function of BER is to remove these frequently produced lesions and maintain genomic integrity. Oxygen radicals generate mostly non-bulky DNA lesions, most of them are substrates for Base Excision Repair (BER). The BER pathway engages various enzymes and proteins and involves the following major steps: (i) recognition and excision of an inappropriate base, (ii) incision at the resulting abasic site, (iii) replacement of the excised nucleotide, (iv) processing of the terminal end(s), and (v) sealing of the final nick (Wilson III D. M. et al., 2007). This repair system is iniziated by a lesion-specific DNA glycosylase (mono or bi-functional), which recognizes and hydrolyzes N-glycosidic bond (bond between the nucleotide base and ribose) of a substrate base, creating an apurinic or apyrimidinic (AP) site intermediate. DNA-glycosylases are the key enzymes of the BER pathway. We can distinguish monofunctional DNA glycosylases, such as uracil-DNA glycosylase (UNG) (it excises the uracil that is the product of cytosine deamination, thereby preventing the subsequent C→T point mutation) and N-methylpurine-DNA glycosylase (MPG) that possess only glycosylase activity, and bifunctional DNA glycosylases, such as 8-oxoguanine DNA glycosylase (OGG1), mutY homolog (MUTYH), endonuclease III-like 1 (NTH1) and NEIL1, exhibits both glycosylase activity and an intrinsic 3' AP lyase activity. This repair system may take place by "shortpatch" BER (also called "single-nucleotide BER"), in which a single nucleotide gap is generated and subsequently filled and ligated, or by long-patch BER in which a gap of 2–10 nucleotides is generated and filled. The short-patch BER is generally the dominant pathway, long-patch BER may be the dominant mechanism of postreplicative BER initiated by UNG2 (Otterlei M. et al. 1999) or NEIL1 (Hegde M. 5L. et al. 2008b), both expressed at highest levels during S phase.

In the short-patch BER AP sites in the DNA are repaired by the action of AP-endonuclease 1 (APE1). APE1 cleaves the phosphodiester chain 5' to the AP site. Then DNA strand contains a 3'-hydroxyl group and a 5'-abasic deoxyribose phosphate. DNA polymerase β (Polβ) fills in the gap with correct nucleotide and removes the deoxyribose phosphate through its associated AP-lyase activity. The presence of X-ray repair cross-complementing group 1 (XRCC1) is necessary to form a heterodimer with DNA ligase III (LIG3). XRCC1 acts as a scaffold protein to present a non-reactive binding site for Polβ, and brings the Polβ and LIG3 enzymes together at the site of repair (Lindahl T. and Wood R. D., 1999). Poly(ADP-ribose) polymerase (PARP-1) interacts with XRCC1 and Polβ and is a necessary component of the BER pathway (Caldecott K.W. et al., 1996; Dantzer F. et al., 2000). The final step of repairing is performed by LIG3, which connects the deoxyribose of the replacement nucleotide to the deoxyribosylphosphate backbone.In long-patch BER, APEX1 catalyzes the formation of a nick 5' to the AP site, PCNA (proliferation cell nuclear antigen) acts as a scaffold protein for the

Figure 9 Scheme of base excision repair showing the two subpathways: (A) the 'short-patch' or single-nucleotide pathway, and (B) the 'long-patch' pathway. Crossing over of the pathways can occur at points (3) and (9). There are essentially four steps in the base excision repair pathway. First, when an altered base is detected (1) the surveillance glycosylases remove that base (2). Next, the endonuclease that is specific for an apurinic or apyrimidinic site cleaves the strand on the 5' side of the abasic site (3). This is followed by filling in of the gap with a correct nucleotide by DNA pol β, and at the same time releasing the dRP (4). Finally, DNA ligase III ligates the newly introduced nucleotide with the downstream sequence (5), thereby restoring the repaired DNA (6). Sometimes, other DNA polymerases such as DNA polymerase δ or ε, along with PCNA, are involved in filling larger sized gaps, also in a strand-displacement manner (long-patch repair [B]; steps 7–12). (Rao K. S., 2007).

recruitment of Polδ and Polε that generate an oligonucleotide flap. The existing nucleotide sequence is removed by flap endonuclease-1 (FEN1). Then oligonucleotide is ligated to the DNA by DNA ligase I (LIG1), sealing the break and completing the repair (Fig. 9).

The decision to proceed via the long-patch or shortpatch BER mechanism is poorly understood. Different hypotheses exist for the switch between long-patch and short-patch
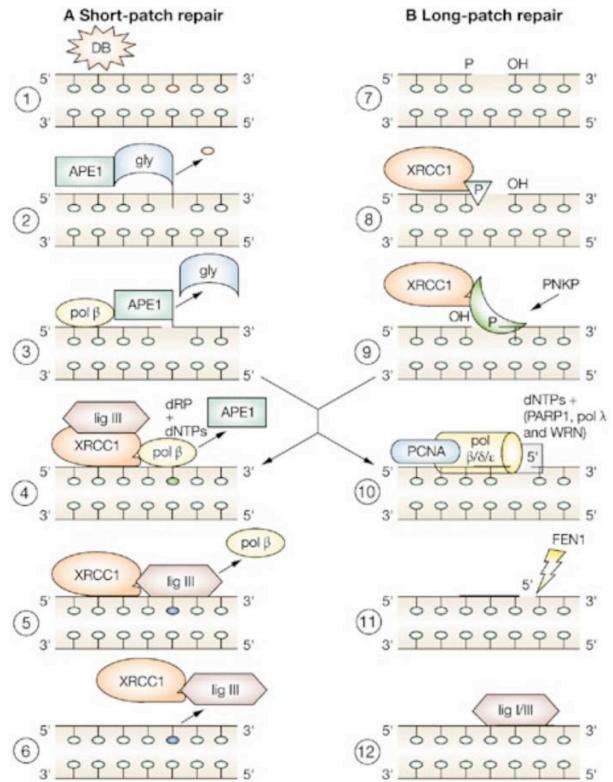
BER. One hypothesis suggests that the switch from short-patch to longpatch BER depends on the relative ATP concentration near the AP site, which is modulated by LIG3 and X-ray cross-complementing protein 1 (XRCC1) (Petermann E. et al., 2003). It was shown that long-patch BER neads low ATP concentrations, whereas short-patch BER prefers mechanism with elevated ATP concentrations. A second study has shown that the decision to proceed via long-patch BER or short-patch BER can be traced by the 5'-dRP intermediate produced by AP endonuclease activity. It was shown that if this dRP can be efficiently removed by dRP lyase activity of POLβ, BER proceeds by the short-patch mechanism. However, if the dRP can not be effectively removed BER pathway proceeds by the long-patch mechanism, not to generate a nick that is refractory to the action of a DNA ligase (Klungland, A. Et al., 1993).

Nucleotide excision repair (NER)

The nucleotide excision repair (NER) is the most flexible of all DNA repair mechanisms for its ability to resolve numerous DNA lesions, particularly base modifications that distort the normal helical structure of duplex DNA (Gillet L. C. et al., 2006). Examples of NER substrates include: cyclobutane pyrimidine dimers (CPDs) and pyrimidine-(6,4)-pyrimidone photoproducts (6-4PPs) generated by UV radiation; base adducts created by exogenous chemical agents such as cisplatin and benzopyrene; and reactive oxygen species (ROS)-induced base modifications such as the cyclopurines (Fig. 10). These "bulky" DNA alterations arrest progression of replication and transcription, resulting replication fork collapse or stalled transcription, but in some circumstances can be bypassed in an error-prone manner. The NER response involves four primary steps: (i) recognition of the damage, (ii) incision on both sides of the lesion and removal of the damage-containing oligonucleotide fragment, (iii) gap-filling synthesis to restore a DNA duplex, and (iv) ligation
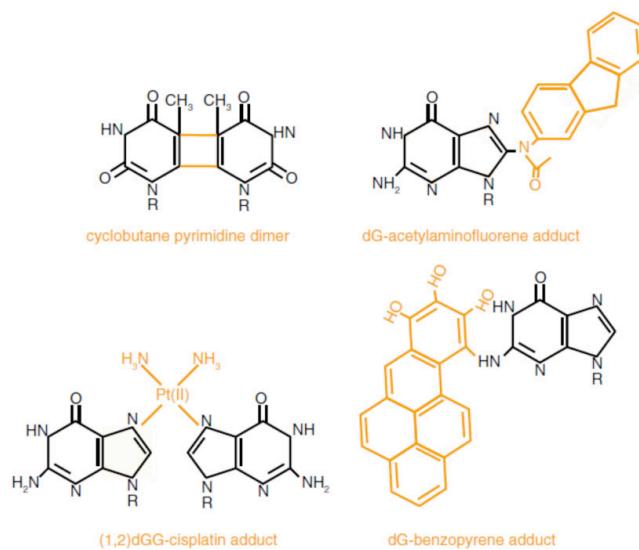
Figura 10 Structure of NER substrates. Structures of the main NER substrate cyclobutane pyrimidine dimer and the more peculiar dG-acetylaminofluorene, (1,2)-dGG-cisplatin and dG-benzopyrene adduct. (Alekseev S. 2014)

of the nick. The NER pathway involves the action of about 20-30 proteins operating in a coordinated manner. NER can recognize and remove the lesion through two sub-pathways mechanistically conserved from prokaryotic to eukaryotic cells: global genome repair (GG-NER) removes lesions anywhere in the genome (Fagbemi A. F. et al., 2011) and transcription-coupled NER (TC-NER), which removes lesions on the transcribed strand of DNA that block RNA polymerase (RNAP) progression (Diderich K. et al., 2011). These two pathways differ only at the step of recognition, but utilize common machinery to execute the final steps of the repair response (Fig. 11). Defects in NER are linked to genetic disorders characterized by extreme UV radiation sensitivity such as xeroderma pigmentosum (XP), Cockayne syndrome (CS) and a photosensitive form of trichothiodystrophy (TTD) (Kraemer K.H. et al., 2007).
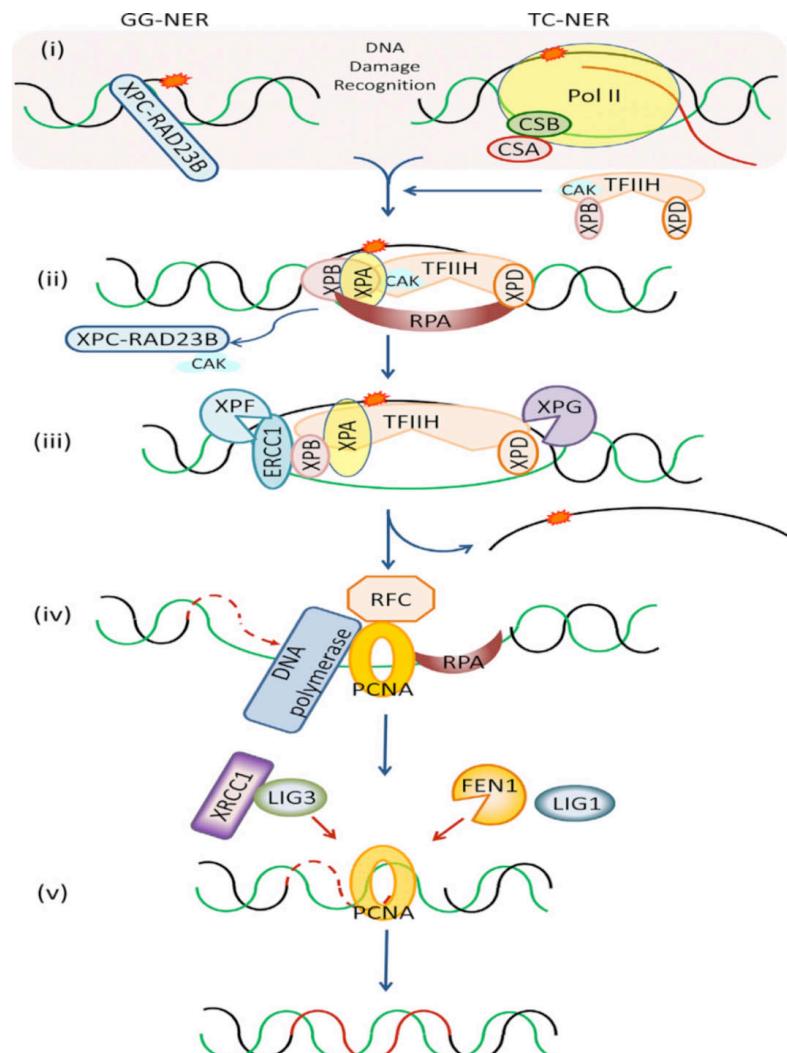
Figura 11 Nucleotide excision repair pathways. Two subpathways of mammalian NER: GG-NER and TC-NER. (i) XPC-RAD23B recognizes DNA damage-induced structural change as the initiation step of GG-NER. TC-NER is initiated by stalling of an elongating RNAP at a blocking lesion on the transcribed strand within an active gene. After these initial recognition steps, GG-NER and TC-NER pathways involve many of the same protein components. (ii) Following recognition, the TFIIH complex is recruited. Through the activity of the helicase

subunits, XPB and XPD, TFIIH promotes opening of the DNA duplex around the lesion, facilitating recruitment of XPA and RPA. (iii) The XPF–ERCC1 complex is recruited to the lesion via a direct interaction with XPA, while XPG is specifically engaged through an interaction with TFIIH. The two endonucleases, XPF–ERCC1 and XPG, are responsible for carrying out incision 5'and 3', respectively, to the DNA damage. (iv) After dual incision and removal of the damage-containing oligonucleotide fragment, a DNA polymerase carries out gap-filling repair synthesis in cooperation with RFC and PCNA. (v) Finally, the nick is sealed by either XRCC1–LIG3α or a FEN1–LIG1 complex. CAK, the cyclin-dependent kinase (CDK)-activating kinase; GG-NER, global genome-NER; RFC, replication factor C; RPA, replication protein A; TC-NER, transcription-coupled NER; TFIIH, transcription factor II H. (Iyama T. et al., 2013).

The GG-NER removes helix-distorting "blocking" lesions located throughout the genome, presumably in a cell cycle-independent manner (Fagbemi A. F. et al., 2011). Two protein complexes XPC-RAD23B recognized a damage-induced structural change in DNA, binding the strand opposite the lesion and not the chemical adduct itself (Min J. H. et al., 2007; Scharer O.D. et al., 2007), after mediate the recruitment of the transcription factor II H (TFIIH) complex, which contains ten subunits, including two helicases XPB (3'-5') (Coin F. et al., 2007) and XPD (5'-3') (Mathieu N. et al., 2010; Sugasawa K. et al., 2009). Through its helicase activity TFIIH promotes opening of the DNA duplex around the lesion, creating a "bubble" platform for recruitment of XPA and replication protein A (RPA), and assembly of the pre-incision complex. XPA promotes the release the cyclin-dependent kinase (CDK)-activating kinase (CAK) associated with TFIIH, and the association of RPA with the single stranded damaged DNA (Camenisch U. et al., 2009; Missura M. et al., 2001). The dissociation of CAK allows the recruitment of the XPF-excision repair cross complementing 1 (ERCC1) complex and XPG, XPF–ERCC1 complex is recruited to the lesion via a direct interaction with XPA (Tsodikov O.V. et al., 2007; Orelli B. et al., 2010), while XPG is recruited through TFIIH. The two endonucleases XPF-ERCC1 and XPG carry out the incision 5' and 3' to the DNA damage respectively, and an oligonucleotide of 24-32 nucleotide containing the lesion is excised. The single strand gap is filled by DNA polymerases δ, ε or κ in cooperation with replication factor C (RFC) and proliferating cellular nuclear antigen (PCNA). Finally, the nick is sealed in dividing cells by either a XRCC1-DNA ligase III (LIG3) or a flap endonuclease 1 (FEN1)–DNA ligase I (LIG1) complex (Mocquet V. et al., 2008), or in non-dividing cells by XRCC1–LIG3α (Moser J. et al., 2007).

In the transcribed strand of active genes lesions are preferentially repaired via TC-NER. In order to repair the damage, TC-NER uses many of the same protein components of GG-NER. Although this molecular mechanism is not yet known, the current model proposes that it is started by stalling of an elongating RNAP at a lesion on the transcribed strand. This arrested

RNAP serves as a critical signal via in which acts CS proteins, CSA and CSB, that facilitate the eventual removal of the damage and restart of transcription. After recruitment of the TFIIH complex the helicases XPB and XPD unwind DNA helix and after the same protein machinery described for GGNER is presumably called upon for incision, excision of the lesion containing strand, gap-filling and nick ligation. The fundamental role of transcriptional factor TFIIH in these repair mechanisms underlines a direct biochemical link between transcription and DNA repair. For example, transcriptional activators (such as Gal4–VP16 and RAR) can stimulate DNA repair (Frit P. et al., 2002). Conversely, a DNA repair complex (that contains XPC) seems also to function as a co-activator for OCT4 (octamer-binding protein 4). Finally, NER factors are recruited to active promoters and facilitate chromatin modification to regulate transcription in the absence of exogenous genotoxic attack (Le May N. et al., 2010) . Indeed, NER factors seem to be associated with the transcription machinery at the promoters of several activated nuclear receptor-dependent genes. The recruitment occurs in a sequential order following PIC assembly and it is distinct from the order that is required for a repair complex. Although NER factors are not essential for PIC formation, it is likely that NER components optimize the efficiency of transcription influencing chromatin remodelling (Compe E. & Egly J., 2012).

DNA damage: double strand break (DSB)

The double strand breaks (DSBs) are one of the most deleterious forms of DNA damage. These lesions are generated when the two complementary strands of the DNA double helix are broken simultaneously in neighboring sites (resulting in physical cleavage of the DNA backbone) so the coupling of the bases and the structure of chromatin become insufficient to maintain the ends of the juxtaposed DNA. Dissociated ends can recombine with other genomic sites leading to mutations and chromosomal aberrations. DSBs can cause problems for DNA transcription, replication and chromosome segregation and it may result in genomic instability, such as translocations, if misrepaired (Bohgaki T. et al., 2010). They can arise endogenously through the action of ROS that are produced by normal cellular metabolism, or through replication-fork collapse, during the processing of interstrand crosslinks, or following exposure to IR (radiation ionizing). The machinery involved in repair of these lesions must be able to (i) detect DNA damage in different chromatin structures; (ii) remodel the local chromatin architecture to provide access to the site of damage; (iii) reorganize the
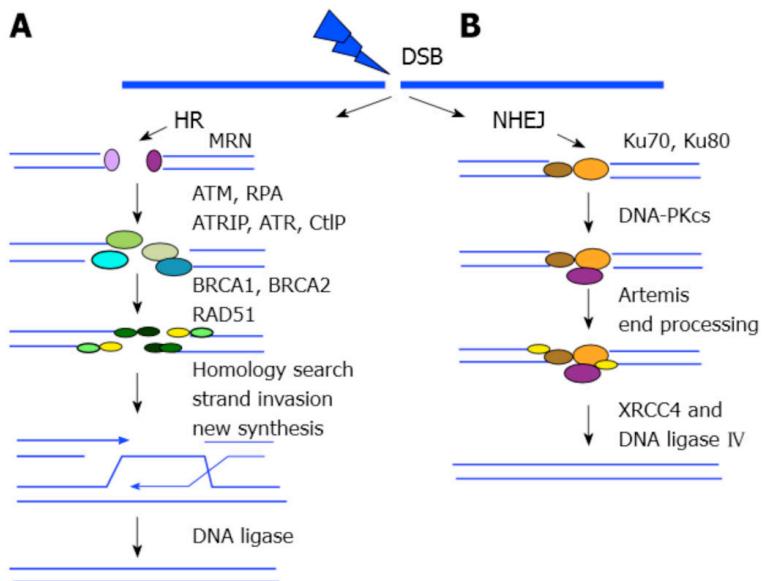
nucleosome-DNA template for processing and repair of the damage; and, importantly, (iv) restore the local chromatin organization after repair has been completed (Brendan D. et al., 2013). Eukaryotic cells have complex systems that detect and signal the presence of these lesions. The double strand break repair (DSBR) is divided into two major pathways: homologous recombination (HR) and non-homologous end joining (NHEJ). HR operates in dividing cells and in S phase or G2 phase, when the double-strand breaks are repaired using the long regions of homology (> 100bp) available from sister chromatids during separation; whereas NHEJ can function in both dividing and non-dividing cells and independently of cell cycle.

Homologous recombination (HR)

During the HR, the damaged chromosome forms a pairing complex with a not damaged DNA molecule with which it shares the homologous sequence. Under these conditions the lost genetic information on one allele can be recovered on the other, wich remained intact. The enzymatic machinery responsible for this process is identical to the crossing-over one during meiosis. The steps of the repair by homologous recombination are: 1. processing of the ends to single strand; 2. homology pairing; 3. copying the information from the non-damaged strand; 4. resolution of the complex and reconstitution of the double helix.

Homologous recombination starts with the recognition of the DSB by the MRN complex, (MRE11, RAD50, NBS1) which initiates 5'-3' end resections to create the 3' ssDNA overhang (Stracker T. H. et al., 2011) Further resection is carried out by exonuclease EXO1, and the resulting ssDNA is stabilized by binding of RPA. Subsequently, RAD52 promotes replacement of RPA with Rad51, leading to the formation of Rad51 nucleoprotein filaments. In this process BRC2 has a key role because stimulates the assembly of RAD51 onto the RPA-ssDNA complex (Jesen R. B. et al., 2010) RAD51-coated ssDNA moves into (invades) the similar or identical recipient DNA duplex in a process called *strand invasion,* which provides the genetic instruction for accurate repair, thus formed a displacement-loop (D-loop) (structure formed between the invading 3' overhang strand and the homologous chromosome). After the invasion, the DNA strand is extended by polymerase (Pol η) and so the D-loop is changed to a cross-shaped structure known as a *Holliday junction;* alternative model, designated the synthesis-dependent strand annealing (SDSA) pathway, completes DSBR via non crossover products, without Holliday junction formation (Fig. 12)

The nonhomologous end joining recombination (NHEJ) is the major DSBR system in higher eukaryotes (Lieber M. R., 2010) particularly during phases of the cell cycle when a homologous sister chromatid is absent; it can operate during all phases of the cell cycle, but it is most active during G1. This mechanism brings the direct ligation of two DNA ends in close spatial proximity, it often results in an error-prone outcome, with partial loss of genome information at the site of the DSB. NHEJ provides three steps: (i) recognition of the two-ended DSB, (ii) processing to remove non-linked termini or other forms of DNA damage near the break and to reveal short stretches of microhomology, and (iii) joining of two ends. The process starts when Ku70/Ku80 (Ku) heterodimer binds the two DSB ends and recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). This multiprotein complex both stabilizes and aligns the DNA ends (Getts R. C. et al., 1994). The interaction between two DNA-PKcs positioned at each DSB terminus activates its intrinsic protein kinase activity, bringing to DNA-

Figura 12 DSB repair pathway. The 5'-3' exonuclease activity of MRN complex creates strand invasion and ATM phosphorylation on each side of the break. RPA assists the assembly of RAD51-B to D, and XRCC2 and XRCC3. RAD52 stimulates filament assembly while RAD54 identifies the homologous sequence. Finally, the resolution of Holliday junctions is completed by resolvases and DNA ligases. NHEJ, Ku and its subunits Ku70(XRCC6) and Ku80(XRCC5) bind the broken DNA termini, recruiting PKcs and Artemis. XRCC4 allows the Ligase IV along with XLF/Cernunos factor, APLF, WRN, PNK, and Polμ to bind DNA and ligate both DSB ends. (Peg G. et al., 2011)

PKcs autophosphorylation and dissociation. Depending on the complexity of the DSB and the nature of the ends, different processing factors are then recruited. In one example, DNA-PKcs activates the endonuclease, Artemis, which cuts the 3' and 5' single-stranded overhangs at the DNA ends to reveal complementary nucleotide stretches . After generation of terminal overhangs by Artemis, and prior to ligation, PNKP is recruited via an interaction with the XRCC4–Ligase IV (LIG4) protein complex to remove 3'-P groups or add 5'-P residues (Koch C. A. et al., 2004). In instances where short nucleotide gaps remain after microhomology

mediated annealing, DNA polymerases μ and λ function to fill in the small gap segments (Andrade P. et al., 2009). Finally, once appropriate termini have been generated, XLF-Cernunnos (XLF) can interact with the XRCC4–LIG4 complex to stimulate end joining, the final step of the repair process (Fig. 12).

Mismatch repair (MMR)

The architecture of the MMR pathway is well conserved from bacteria to mammals. This system recognizes and repairs base-base mismatches and insertion–deletion loops (IDLs) that arises primarily as errors or intermediates of DNA replication or homologous recombination. This system is very important for the cell in fact leads to an ~100-fold increase in DNA replication fidelity by preventing base substitutions or repeat sequence instability. To preserve genome integrity, MMR must occur selectively on the new synthesized DNA strand that contains the mispaired nucleotide. Thus, it is critical to discriminate between the nascent and template DNA strands. In prokaryotes, the E. coli endonuclease, MutH, recognizes the new synthesized strand by its unmethylated status. In eukaryotic cells this does not appear to be the mechanism of strand discrimination. Study reveals that the MMR machinery is associated with the replication apparatus, proposing that the 5' ends of existing Okazaki fragments serve as markers for strand discrimination. MMR pathway has been extensively studied in *E.coli.,* it requires the following protein components: MutS, MutL, MutH, DNA helicase II (MutU/UvrD), four exonucleases (ExoI, ExoVII, ExoX, and RecJ), single stranded DNA binding protein (SSB), DNA polymerase III holoenzyme, and DNA ligase (Burdett V. et al., 2001) MutS, MutL, and MutH initiate MMR and play specialized biological roles in the process. MutS homodimer recognizes base-base mismatches and small nucleotide insertion/deletion (ID) mispairs (Modrich P. et al., 1996) and recruits MutL. MutL enhances mismatch recognition, and recruits and activates (in presence of ATP) the endonuclease activity of MutH. MutH specifically incises the unmethylated daughter strand of hemimethylated sequence dGATC (Junop M.S. et al., 2001) and this strand-specific nick provides a starting point for excision of the mispaired base. The UvrD helicase unwinds the ends of the nicked error-containing strand from the template and it generates single-strand DNA until mismatch that is removed through exonucleolytic degradation.

In eukaryotes enzymes of MMR have a high homology to E.coli MMR proteins *E. coli*. In human cells, it was been identified five MutS homologues (MSH) but only MSH2, MSH3 and

MSH6 participate in MMR in the form of heterodimers. The most abundant mismatch-binding factor is MutSα that initiates the repair of base–base mismatches and IDLs of one or two extrahelical nucleotides, whereas the repair of larger IDLs is initiated by MutSβ. MutS α/β recognizes the damage and recruits MutLα. hMutLα possesses an ATPase activity and regulates termination of mismatch-provoked excision (Zhang Y. et al., 2005). Recent studies show that MutLα possesses a PCNA/replication factor C (RFC)-dependent endonuclease activity which plays a critical role in 3′ nick-directed MMR involving EXO1 (Kadyrov F. A. et al., 2006). PCNA interacts with MSH2 and MLH1 and it is thought to play roles in the initiation and DNA resynthesis steps of MMR (Gu L. Et al., 1998) in fact during DNA replication, PCNA sliding clamp is loaded onto the 3′ terminus of an Okazaki fragment or onto the 3′ end of the leading strand by replication factor C (RFC), it coordinates Exo1 to excise the to excise the mismatch from the nascent strand, generating a multinucleotide gap. The removed DNA segment is then resynthesized by polymerase δ and the nicked sealed by LIG1 (Larrea A. A. et al., 2010).

Transcription

In eukaryotes gene expression is a highly complex and regulated process. The first level of complexity encountered by the cellular transcriptional machinery is the organization of the genome into chromatin. The dynamic interactions of different cellular factors such as histone modifying enzymes, ATP-dependent remodelers and histone chaperones with the chromatin components facilitate gene expression.

Transcription is the process by the information in DNA is copied into RNA. It is performed by RNA polymerase. In the nucleus of eukaryotes, transcription is carried out by three different RNA polymerases, RNA polymerase I, II and III (Pol I, II and III) that transcribe distinct classes of genes: Pol I is responsible for the transcription of the large ribosomal RNA genes (28S, 18S and 5.8S), Pol II for the transcription of the protein-coding genes and some small nuclear RNAs and Pol III transcribes some structural and catalytic RNAs, including most small nuclear RNAs, tRNAs and 5S rRNA (Sentenac A., 1985). All three nuclear RNA polymerases are complex enzymes, consisting of 8 to 14 different subunits each. Eukaryotic transcription is a sequence of event that exhibits regulation at multiple steps. It has three main steps: initiation, elongation and termination. Initiation consists in the binding of RNA polymerase to double-stranded DNA; this step involves a transition to single-strandedness in

the region of binding; in particular this process starts when an activator binds a DNA specific sequence (enhancer DNA elements) which promotes before a sequential recruitment of general transcription factors (GTFs) and after RNA polymerase II (RNAPII) to the target gene promoters (Thomas M. C. et al., 2006). In many eukaryotic organisms, the promoter contains a conserved gene sequence called the TATA box. Various other consensus sequences also exist and are recognized by the different TF families. The TATAA sequence is recognized by a general transcription factor called TFIID. TFIID is composed by multiple subunits, including the TATA-binding protein (TBP), which binds and distorts specifically to the TATAA consensus sequence, and 10-12 other polypeptides, called TBP-associated factors (TAFs). TBP then binds a second general transcription factor (TFIIB) forming a TBP-TFIIB complex at the promoter. TFIIB in turn serves as a bridge to RNA polymerase, which binds to TBP-TFIIB complex in association with a third factor, TFIIF. Following recruitment of RNA polymerase II to the promoter, the binding of two additional factors (TFIIE and TFIIH) is required for initiation of transcription (Fig. 13A). TFIIH is a multisubunit factor that appears to play at least two important roles. First, two subunits of TFIIH are helicases, which may unwind DNA around the initiation site. Another subunit of TFIIH is a protein kinase that phosphorylates Ser5 residue of the repeated sequences present in the C-terminal domain (CTD) of the largest subunit of RNA polymerase II in particular. Phosphorylation of these sequences is very important because permits the transition from abortive to productive transcription initiation (Fig. 13B). After the formation of PIC at the gene promoter the RNA polII is stalled on promoter regions, in this condition it produces transcripts of less than 5 nucleotides that resulting in abortive initiation. In order to create a condition for productive transcription initiation, specific post-translational modifications are required on the C-Terminal domain (CTD) of RNAPII and N-terminal histone tails of the nucleosome. To increase chromatin accessibility, gene-specific activators mediate the recruitment of histone modifying enzymes such as lysine acetyltransferases (KAT), acetyltransferases such as p300, CBP and PCAF that add acetyl-groups to several residues of histones H3 (K9, K14) and H4 (K16), lysine acetyltransferases (KAT), lysine/arginine methyltransferases, histone demethylases and kinases to the chromatin which work together with nucleosome remodelers to reorganize the chromatin architecture. On the other hand the CTD of RNA polII is subject to a switch of post-translational modification (there is an increase of phosphorylation status of Ser2 and gradual loss of Ser5 phosphorylation of the CTD repeats) through the positive

transcription elongation factor, P-TEFb, that plays crucial role to facilitate the entry of RNA polymerase II into the transcription elongation. After these events Pol II leaves the promoter, TFIIB and TFIIF are released, whereas other factors such as activators, TBP, Mediator, TFIIH and TFIIE remain largely promoter-associated and form what is termed a reinitiation intermediate or scaffold, to facilitate subsequent rounds of transcription (Yudkovsky N. et al., 2000). The elongating RNAPII recruits several factors which regulate the rate of transcription. The processivity of transcription is facilitated by elongation factors such as TFIIS and Spt5 as well as the histone chaperone complex, FACT. One of the key epigenetic marks associated with elongation is methylation of histone H3 at K36 by the methyltransferase SET2. Another modification of histones associated with elongation is H2B mono-ubiquitination, which is present at both promoters and within the open reading frame (ORF) (Li B. et al., 2007; Fuda N. J. et al., 2009). The remarkable processivity of RNAPII requires that the transcription elongation complexes maintain stable association of RNAPII with the DNA since any event of disengagement from the template will result in abrupt transcription termination. A stretch of 8–9 bp nascent RNA remains bound to the template DNA during transcript elongation. The maintenance of RNA:DNA hybrid is a critical determinant for the processivity of RNAPII during transcription elongation and any disruption of this heteroduplex may result in premature termination of transcription (Fig. 13C).

The dissociation of RNA polII from the DNA template marks the end of a transcription cycle. Termination is the last step in the series of regulates events required for gene transcription; this event may also serve to promote subsequent round a of transcription. Transcription termination occurs through the poly A-dependent pathway; during this process RNA polII transcribes the Poly_A signal, 5′-AAUAAA-3′ which is followed by a G/U-rich sequence towards the 3′ end, that reduce its processivity and leads to its pausing further downstream. After that, endoribonucleolytic cleavage of the nascent transcript occurs in polyadenylation sequence. In humans, there are different protein complexes that facilitate the key events of transcription termination, which include the cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulatory factor (CstF) and Poly (A) polymerase (Kuehner J. N. et al., 2009). The subunit Rpb1 of RNAPII acts as a scaffold for the recruitment of these termination complexes. The Ser2-phosphorylated CTD of RNAPII plays a key role in both 3′ processing and transcription termination (Ahn S. H. et al., 2004). Both CPSF and CstF complexes are recruited to this phosphorylated residues. The CPSF complex binds to the transcribed poly

(A) site on the pre-mRNA, induces RNAPII pausing, and eventually stimulates cleavage and release of RNAPII at the gene terminus (Nag A. et al., 2007) (Fig. 13D).
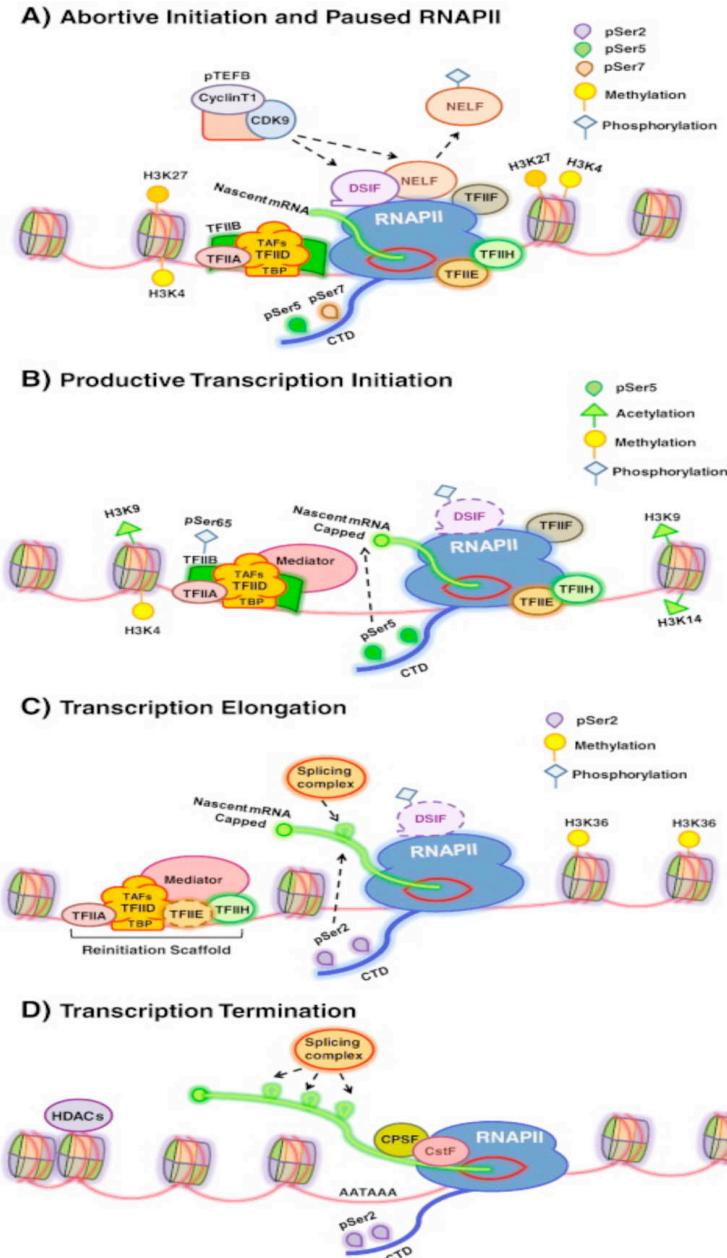
Figura 13 The transcription cycle. (A). Depending on the transcriptional competence of RNAPII, it can potentially enter a paused state. Presence of negative factors such as NELF and DSIF inhibits productive transcription initiation resulting in abortive transcription. Nucleosomes flanking such a paused promoter are characterized by the presence of a bivalent
histone mark: methylation of histone H3 at lysines 4 and 27 (H3K4/27). Paused RNAPII is also enriched with phosphor Ser7 mark (orange balloon) at its C-terminal domain (CTD). The kinase activity of the CDK9 subunit of the pTEFb complex alleviates this repression via phosphorylation of NELF and DSIF which results in dissociation of phosphorylated NELF from RNAPII, while DSIF may move along with the elongating RNAPII. (B). RNAPII can switch to productive initiation mode when all the inhibitory signals/factors are overcome.
The nucleosomes flanking the active promoters harbor the histone marks for active transcription. Active initiation is dependent on TFIIH-mediated promoter melting (red bubble) and phosphorylation of the CTD repeats at serine 5 (green balloon). Along with the phosphorylation of Ser5-CTD, productive transcription

initiation also requires the phosphorylation of TFIIB at serine 65. The phospho-Ser5-CTD recruits capping enzyme to the 5′ region of nascent mRNA (green string) which triggers RNAPII-escape from the promoter to the open reading frame (ORF) of the gene. (C). Following promoter clearance, RNAPII proceeds for elongating the transcript while a part of the PIC components remains associated at the promoter forming a reinitiation scaffold.Meanwhile, GTFs such as TFIIB, TFIIF and likely TFIIE fall off. The elongating RNAPII CTD repeat is progressively phosphorylated at serine 2 by cyclin-dependent kinase 9 (CDK9), while the SSU72 phosphatase removes the phospho-Ser5 mark. Phospho-Ser2-CTD (purple balloon) recruits mRNA splicing complex for co-transcriptional splicing of nascentmRNA. The nucleosomes occupying the ORF are enrichedwithmethylation of histone H3 at lysine 36which is a mark of active elongation (D). Once the RNAPII reaches a pause signal (poly A) at the gene terminal, 3′ end processing and termination specific complexes such as CPSF and CstF are recruited. The CTD repeat is hyperphosphorylated at serine 2 at the gene terminus. Phospho-Ser7 (orange balloon) regulates the termination of snRNA genes by recruiting the Integrator complex. The region already transcribed by RNAPII is efficiently reassembled into chromatin with the aid of histone chaperones and deacetylases (HDACs). (Shandilya J. and Roberts S. G., 2012)

This mechanism of transcript termination is a general norm for most RNAPII transcribed protein-coding genes.

Transcription by RNA polymerase II is coupled to RNA processing, including capping, splicing and cleavage/ polyadenylation. The C-terminal repeat (CTD) of RNA pol II orchestrates both processes by recruiting RNA processing factors. Indeed, CTD directly binds polyadenylation factors and its truncation inhibits transcript cleavage *in vivo*. Ssu72 is a phosphatase protein that catalyzes the dephosphorylation of the C-terminal domain of RNA polymerase II. Genetic and physical interactions between Ssu72 and RNAP II have been demonstrated and it has been hypothesized a role for Ssu72 in basal (noninduced) transcription by RNAP II (Pappas D. L. Jr & Hampsey M., 2000). Ssu72 was initially identified in a screen for suppressors of sua7-1, a cold-sensitive mutation in yeast TFIIB, hence Ssu72 (Suppressor of sua7-1 clone 2). Recent analysis revealed that Ssu72 dephosphorylates Ser-5 in the CTD of RNA pol II and regenerates initiation competent hypo-phosphorylated RNA pol II (Krishnamurthy S. et al., 2004). Large-scale analysis of protein complexes in yeast identified Ssu72 as a component of a cleavage and polyadenylation factor (CPF) complex, it interacts directly with the Pta1 subunit of CPF and is implicated in transcript cleavage and termination (Dichtl B. et al., 2002). Alternatively, Ssu72 may exert two independent functions: transcription/RNA processing in the nucleus and as yet to be defined activity in the cytoplasm. Interestingly, there are no apparent Ssu72 homologs in bacterial or archaeal genomes, implying that Ssu72 function is specific for eukaryotes.

Transcription is a cyclic process in which a same sequence can be transcribed more times. Some studies revealed that the terminal and promoter regions of active genes can interact via a phenomenon known as gene looping (Fig. 14). In yeast and humans, transcription initiation factor TFIIB is known to interact with the cleavage and polyadenylation factor (yCPF and hCPSF) and CstF complexes; this association with termination-specific complexes is regulated by the phosphorylation status in fact, the phosphorylation of TFIIB at Ser65 not only helps productive transcription initiation, but also facilitates gene looping via through its interaction with CstF components (Wang Y. et al., 2010). Also TFIIH kinase subunit (CDK7) have a crucial role in the formation of gene looping, in fact CDK7 is the major kinase involved in phosphorylation of Ser5 and Ser7 residues of RNAPII; in addition TFIIH may be involved in reinforcing the

Figura 14 Transcription recycling via gene looping. The terminator and promoter regions of a gene juxtapose during active transcription in order to facilitate RNAPII recycling and multiple rounds of transcription. The prebound GTFs and mediator in the reinitiation scaffold stabilized by the activator along with phospho-TFIIB interact with RNAPII and the termination complexes such as CPSF, CstF, and mediate such promoter–terminator contacts known as gene looping and thereby increase the efficiency of reinitiation by RNAPII. (Shandilya J. and Roberts S. G., 2012)

promoter–terminator contacts for active RNAPII recycling. Another promising candidate for facilitating gene loop formation is human positive cofactor 4 (PC4), a multifunctional protein involved in transcriptional regulation (Banerjee S. et al., 2004).

Chromatin state can limit access of transcription factors and RNA polymerase to DNA promoters, contributing to the restrictive ground state of gene expression. In order for gene transcription to occur, the chromatin structure must be unwound the chromatin structure and the presence of specific transcription factors there are other control activities in the cell, such as epigenetic mechanisms, including DNA methylation and imprinting, noncoding RNA and histone post-translational modifications (Phillips T., 2008).
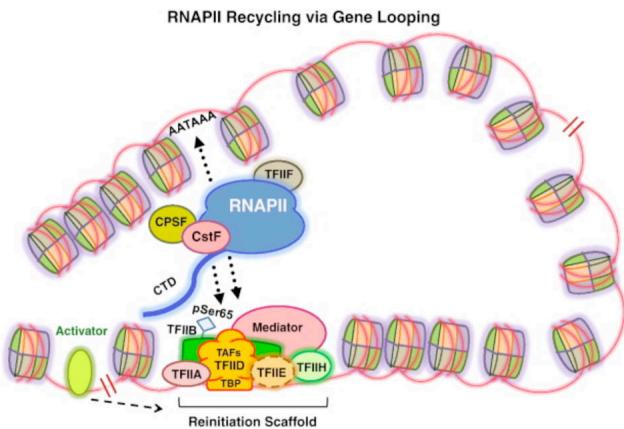
Chapter 2: The aim of study

The maintenance of the genome integrity and its accurate replication are prerequisites for life. The processing of genetic information by transcription, which include RNA templates for protein synthesis and regulatory information to coordinate a chemical reactions, is crucial for cell function and survival. These processes can be threatened by alterations in the DNA structure that interfere with the progression of DNA and RNA polymerases, or compromise the fidelity of replication and transcription. The perturbation of genome integrity can be induced by DNA damage. Damage to DNA arises from the exposition to exogenous genotoxic agents such as ionizing radiation, ultraviolet radiation (UV) or environmental toxins; or for endogenous sources, such as reactive oxygen species (byproduct of cellular metabolism) or errors during DNA replication. These events can generate DNA lesions, including modified bases, abasic sites, the formation of DNA adducts, the production of single- (SSBs) and double-strand breaks (DSBs) (Lindahl T. et al., 1993).

The cellular 'toolbox' of responses to genomic stress includes checkpoints that result in arrested cell cycle progression, stabilization of DNA at blocked replication forks and recruitment of DNA repair enzymes, to facilitate progression of DNA and RNA synthesis. (Harper J. W et al., 2007; Ljungman M., 2005). DNA lesions can lead to a prolonged transcription arrest followed by a strong signal for cellular apoptosis. For some types of lesions, the RNAPII might be able to carry out translesion transcription but with transcriptional mutagenesis.

Cells have evolved multiple, often overlapping mechanisms to sense and demarcate the DNA lesions, the four major pathways are: base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR) and recombinational repair. Transcription and DNA repair have the same action target: the DNA; therefore it is not surprising that these two processes are often coupled, interdependent and crossfunctional. A connection between transcription and repair was first suggested by experiments showing that transcribed strands of DNA are repaired more rapidly than not transcribed strands in both *E. coli* and mammalian cells (Mellon I. et al., 1987), so transcription complex and DNA repair system are biochemical linked. Since DNA damage blocks transcription, this transcription-repair coupling is thought to be advantageous by allowing the cell to preferentially repair damage to actively expressed genes.

NER factors seem to be associated with the transcription machinery at the promoters of several activated nuclear receptor-dependent genes. The recruitment occurs in a sequential

order following PIC assembly and is distinct from the order that is required for a repair complex. Although NER factors are not essential for PIC formation, it is likely that NER components optimize the efficiency of transcription influencing chromatin remodelling (Compe E. et Egly J., 2012). An evidence of this biochemical connection is given by the multi-protein complex transcription initiation factor TFIIH. TFIIH is indispensable during basal transcription, but it has a crucial role in NER, in fact the XPB and XPD helicases associated with it are required to initiate the transcription of eukaryotic genes; but these helicases appear to be required for the unwinding of DNA during nucleotide-excision repair and the correct repair of lesion.

Active DNA demethylation play a key role in transcriptional control and resetting of epigenetic memory during embryonic development and cellular reprogramming (Wu S, C. et Zhang Y., 2010). There are multiple mechanisms by which 5-methylcytosine (5mC) is removed, the common endpoint is the excision of the deaminated and/or oxidized derivatives of 5mC by the BER enzyme thymine DNA glycosylase (TDG) (Cortazar et al., 2007; Franchini et al., 2012). Dynamic DNA demethylation occurs at promoters and enhancer of genes in mouse ESCs; TDG depletion compromised the reactivation of poised genes during ESC differentiation and embryogenesis, it underline the crucial role of DNA glycosylase in embryogenesis.

Independent of its catalytic activity, TDG acts as a scaffold to bridge the transcriptional coactivator CBP/p300 to transcription factors like c-JUN, RAR/RXR (Cortellino et al., 2011; Um et al., 1998), and estrogen receptor α (ERα) (Chen et al., 2003) to facilitate histone modification.

Transcription process exposes the DNA template to damage by genotoxic agents and generates potentially harmful DNA structures that are prone to mutagenesis and recombination; in fact it requires the separation of the two strands of a DNA double helix by RNA polymerase II (RNAPII), so single-stranded regions can be exposed to chemical modifications such as the deamination of cytosine that if not repaired by BER could lead to C-to-T substitution (Conticello, 2008).

In addition to nucleotide substitutions, ssDNA regions are prone to recombination as well as the formation of local secondary structures (e.g., hairpins and loops) formed at DNA sequence repeats (Kim & Jinks-Robertson, 2012) and the harmful duplex DNA-RNA, formed when

nascent messenger RNA hybridizes with template DNA, particularly prone to mutagenesis and recombination (Aguilera & Garcia-muse, 2012).

In the genes actively expressed spontaneous rates of mutagenesis and recombination events lead phenomena known as transcriptional-associated mutation (TAM) and transcription-associated recombination (TAR).

While global DNA damage is generally associated with gene silencing, DSB formation and localized DNA base damage are implicated in transcriptional activation by nuclear hormone receptors such as estrogen receptor (Perillo B. et al., 2008). The hormone 17β-estradiol (E_2) regulates transcription of target genes by binding to estrogen receptors (ERα and ERβ), which then bind with high affinity to estrogen-responsive DNA elements (EREs). After hormone activation, ERα interacts with transcription co-activators that covalently modify histone proteins within nucleosomes, in particular it recruits and binds LSD1, flavin-dependent monoamine oxidase, which can demethylate lysines 4 and 9 mono- and di-methylated of histone 3 in via redox process. Demethylation by LSD1 is an oxidative process that results in the production of reactive oxygen species (ROS) in particular hydrogen peroxide (H_2O_2) responsible of a burst of DNA oxidation, specifically this process can lead the oxidation of Guanine to produces 8-oxoGuanine (8-oxo-dG).

DNA oxidation involves not only G but also C. In fact C can be oxidized only when methylated in vivo. It has been reported that estrogens induce cycles of the CpG methylation/idroximethylation of cytosine which is settled on the opposite strand by the DNA methyltransferase enzyme (Metivier et al., 2003), DNMT3a and TET1. These modified bases are removed by repair system such as base excision repair (BER) and nucleotide excision repair (NER), that generate a transient DNA nick, relax chromatin tensions and it allows the physical connections between distant intergenic regions and so induces a permissive chromatin architecture for transcriptional activation. DNA damage can have opposite effects on transcription, infact exist a mechanism enable to distinguish between DNA damage event linked to gene activation from those that cause undesirable consequences.

If damage is primarily induced which is the destiny of transcriptional unit after repair?

DSBs are produced when cells are exposed to DNA damaging agents including ionizing radiation (IR), chemical agents and UV light that create replication blocking lesions (alkyl adducts, pyrimidine dimers, and crosslinks), they can be also generated naturally when replication forks encounter blocking lesions such as those produced by metabolic byproducts

of cellular respiration (ROS) leading to fork collapse. Eukaryotic cells repair DSBs primarily by two mechanisms: nonhomologous end-joining (NHEJ) and homologous recombination (HR). DSBs produced by nucleases and IR, can be repaired by either pathway while DSBs produced by replication fork collapse are repaired primarily (or exclusively) by HR (Shen Z. et al., 2007).

Mutants in the *Saccharomyces cerevisiae RAD52* epistasis group display hypersensitivity to ionizing radiation and are defective in DSB repair through homologous recombination (Nicholas, 1998), in addition the depletion of RAD54 in mice dramatically aggravates the ionizing radiation sensitivity, at the embryonic and not adult stage, of severe combined immune deficiency that are impaired in DNA double-strand break repair through DNA end-joining (Essers J. et al., 2000). A common feature of cellular response to DNA damage induced by UV, X-ray irradiation or genotoxic drugs is hyperphosphorilation of CTD of Rpb1 subunit of RNA polymerase II (Rpb1-CTD). Rpb1-CTDis the docking site for a factors that interact with it and induce its phosphorylation/dephosphorylation at Ser/Thr residues, regulate its activity and transcription progression along the gene. P-TEFb is one of the kinases that are involved in the regulation of RNA polymerase II through the phosphorilation of CTD tail; it activity is regulated by a dynamic equilibrium between small active complex and large inactive complex, the equilibrium between this complexes can be perturbed by transcription arrest. Site-specific DSBs trigger disruption of the P-TEFb LC leading to accumulation of active SC P-TEFb, as consequence Rpb1-CTD is hyperphosphorylated as consequence there are the activation of the p53-dependent transcriptional pathway and cell cycle arrest in G2/M phase (Napolitano G. et al., 2013).

The exposure of ERα-expressing breast cancer cells to E_2 induces γH2AX foci, which are markers of DNA DSBs; γH2AX foci require ERα, transcription and the catalytic activity of topoIIβ and occur at the promoter of the E_2 inducible gene PS2. Furthermore, E_2-induced γH2AX foci only appear following progression through S phase and colocalize with Rad51, suggesting that HRR is utilized for repair of E_2-induced DSBs. Transient DSBs generated at ERα-responsive genes as a result of the activity of topoIIβ can become stabilized and that the topoIIβ cleavage complex can create a barrier to DNA replication and lead to replication stress and DSB formation (Williamson L. M. et al., 2010).

Based on this background I can introduced the aim of my project targeted to the study of relationships between transcription, chromatin remodeling and DNA repair mechanisms used

three different models: (i) Gene expression regulated by nuclear receptors, (ii) Damage and direct site repair, (iii) Human primary cell lines in differentiation (CD4$^+$ T lymphocyte). In particular I analyzed:

(i)   The molecular mechanisms involved in the transcription of genes regulated by nuclear receptors (retinoic and estrogen receptor) in *"committed"* cellular model for the expression of these receptors: MCF7. To identify the link between transcription, post-translational modifications of histone H3 and genomic damage. I focused my attention on: a) methylation profiles of lysines 4 (K4) and 9 (K9) in histone H3 after hormonal stimulation; b) recruitment of demethylating enzymes, LSD1 (KDM1A) and JMJD2A (KDM4A) in association with the enzymes of repair systems BER, NER, DNMT3A, TET1; c) the role of these enzymes in 3D organization of genes PS2, BCL2, CAV1, CASP9 and CYP26 in the early phase of transcriptional activation; d) the kinetics oxidation/repair of oxidative damage upon strand plus and strand minus (in PS2 gene).

(ii)   The molecular mechanisms involved in DNA damage repair, the link between DNA damage and gene silencing in a pioneer model of DSB and repair with homologous recombination (HR). In particular my studies focused on a) the relationship between DNA methylation of repair site and changes of chromatin methylation; b) the link between *de novo* methylation at the site of DSB repair and chromatin looping; c) the connection among the recruitment enzymes of repair systems (BER) and DNA methylation after HR.

(iii)   The relationship between chromatin remodeling of FOXP3 gene and immunophenotypic differentiation of human primary CD4$^+$ T lymphocytes; I have study a) the formation of intergenic interaction and the histone code of FOXP3 gene during activation *in vitro* of nTreg (CD4$^+$CD25$^+$ T cells) and Tconv; b) FOXP3 splicing variants that induce suppressive function of iTreg cells tightly depended on glycolysis; c) the role of glycolytic enzyme enolase (ENO1) in the modulation of FOXP3 splicing variants expression.

Chapter 3: *Gene expression regulated by nuclear receptors*

Nuclear hormone receptors are ligand-activated transcription factors that bind lipophilic molecules and regulate gene expression by interacting with specific DNA sequences of their target genes. Nuclear receptor ligands are chemically diverse, including hydrophobic molecules such as steroid hormones (e.g. estrogens, glucocorticoids, progesterone, mineralocorticoids, androgens, vitamin D3, ecdysone, oxysterols and bile acids), retinoic acids, thyroid hormones. The receptor linked to the specific ligand binds as homodimer or heterodimer specific hormone response elements (HRE), which are located in the enhancer region of the gene promoter/regulatory regions promoting transcriptional activation.

I focused my attention on the study of transcriptional activation of genes regulated by retinoic and estrogen receptor. I have studied the effect of activation of the retinoic acid receptor, at the RARE-promoter chromatin of *CASP9* and *CYP26A1* genes, at 15 and 45 min following RA exposure, and we found that histone H3 lysine 4 and 9 are demethylated by the lysine specific demethylase, LSD1 and by the JMJ-domain containing demethylase, D2A. The action of the oxidase (LSD1) and a dioxygenase (JMJD2A) in the presence of Fe^{++} elicits an oxidation wave that modifies the DNA locally and recruits the enzymes involved in base and nucleotide excision repair (BER and NER). These events are essential for the formation of chromatin loop(s) that juxtapose the RARE element with the 5' transcription start site and the 3' end of the genes. The RARE bound-receptor governs the 5' and 3' end selection and directs the productive transcription cycle of RNA polymerase.

The same evaluations were be made for regulative elements ERE-promoter-Poly_A of *PS2*, *BCL2*, *CAV1* estrogen-responsive genes; in addition I have analyzed the kinetic of oxidation/repair of two strands of different regions of PS2 gene to do that we used Ligation Proximity Amplification (LPA) technique. This method consists in a first stage of annealing of oligonucleotides above oxidized G; during the following step there is a ligation of the oligonucleotides through a SALSA 65 ligase which generates a template strand subject of a subsequent amplification. The method use the sensitivity of thermostable ligase SALSA 65 to recognize mismatches between oligonucleotides and the template caused by the presence of a modified bases. To confirm that the mismatches is a consequence of 8oxoG we exploit the glycosylase and AP-lyase activity of FPG enzyme (8-oxoGuanine DNA Glycosylase from *E.coli*) that specifically recognizes the 8-oxoG. The method is sensible to 8-oxoG and not to all modified bases inasmuch the sensitivity evaluated using the same technique on filaments

synthesized introducing modified bases: 8-oxoG and FormilC underlined that after FPG treatment is altered the amplification efficiency of filaments containing oxidized G.

The results obtained show that the repair and oxidation on two stands of regulative regions of *PS2* is asynchronous; in the transcription direction is repaired before the strand plus (TATA region) because it is not the template for mRNA synthesis and so it's free to protein factors; in not codogenic direction is repaired before the strand minus (CnG region). Our data point out that DNMT3a is essential for productive transcription induced by estrogens and for the correct repair of oxidize G; in fact DNMT3A protect the C point in front to 8-oxoG promoting the correct repair of damage base and preventing the simultaneous repair of two 8-oxoG put on the different strand which can cause double strand break. The changes in methylation and hydroxymethylation of CpG are also associated with the recruitment of DNMT3A and TET1 and the changes of hydroxymethylation are associated with the timing and the location of oxidized Gs, suggesting that the 2 events (G and C methylation-oxidation) are tightly associated.

NAR Breakthrough Article

3.1

Mechanism of retinoic acid-induced transcription: histone code, DNA oxidation and formation of chromatin loops

Candida Zuchegna[1], Fabiana Aceto[2], Alessandra Bertoni[3], Antonella Romano[1], Bruno Perillo[4], Paolo Laccetti[1], Max E. Gottesman[5], Enrico V. Avvedimento[3,*] and Antonio Porcellini[1,*]

[1]Dipartimento di Biologia, Università Federico II, 80126 Napoli, Italy, [2]Dipartimento di Medicina e di Scienze della Salute, Università del Molise, 86100 Campobasso, Italy, [3]Dipartimento di Medicina Molecolare e Biotecnologie mediche, Istituto di Endocrinologia ed Oncologia Sperimentale del C.N.R., Università Federico II, 80131 Napoli, Italy, [4]Istituto di Scienze dell'Alimentazione, C.N.R., 83100 Avellino, Italy and [5]Institute of Cancer Research, Columbia University Medical Center, New York, NY 10032, USA

ABSTRACT

Histone methylation changes and formation of chromatin loops involving enhancers, promoters and 3′ end regions of genes have been variously associated with active transcription in eukaryotes. We have studied the effect of activation of the retinoic A receptor, at the RARE–promoter chromatin of *CASP9* and *CYP26A1* genes, 15 and 45 min following RA exposure, and we found that histone H3 lysines 4 and 9 are demethylated by the lysine-specific demethylase, LSD1 and by the JMJ-domain containing demethylase, D2A. The action of the oxidase (LSD1) and a dioxygenase (JMJD2A) in the presence of Fe++ elicits an oxidation wave that locally modifies the DNA and recruits the enzymes involved in base and nucleotide excision repair (BER and NER). These events are essential for the formation of chromatin loop(s) that juxtapose the RARE element with the 5′ transcription start site and the 3′ end of the genes. The RARE bound-receptor governs the 5′ and 3′ end selection and directs the productive transcription cycle of RNA polymerase. These data mechanistically link chromatin loops, histone methylation changes and localized DNA repair with transcription.

INTRODUCTION

Retinoic acid (RA), an active derivative of vitamin A, plays a role in the regulation of embryonic development, homeostasis and differentiation of adult tissues. RA metabolites, collectively known as retinoids, are well-characterized inhibitors of cancer cell proliferation or inducers of stem cell differentiation. The biological activity of RA is mediated by its binding to RA receptors (RARα, RARβ and RARγ) that function as hetero-dimers with retinoid X receptors (RXRs), targeting DNA at specific sites, known as RA responsive elements (RAREs). Following hormone binding, an active receptor complex induces covalent modifications at the N-terminal tails of nucleosomal histones and assembles an active transcription complex on chromatin (1).

Despite extensive studies on RA-induced transcription, it is not known if there is a common set of histone modifications or how the initiation transcription complex is assembled on regulatory regions. The H3 methylation changes reported so far associated with activation of the receptor(s) by RA, may be secondary to repression of transcription (2) or induced by cell differentiation (1,3). Although large DNA domains and histone modifications have been studied during development, the mechanism used by RA to activate transcription still remains elusive.

To address this issue, we studied two prototypic genes induced by RA: caspase 9 (*CASP9*) and Cyp26A1 (*CYP26A1*). *CASP9* contains a functional RARE located 9.5 kb downstream of the transcription start site (TSS) (4),

*To whom correspondence should be addressed. Tel: +39081679047; Fax: +39081679233; Email: antonio.porcellini@unina.it
Correspondence may also be addressed to Enrico V. Avvedimento. Tel: +390817463251; Fax: +390817463308; Email: avvedim@unina.it

whereas the RA-induced *CYP26A1* expression is driven by a compact RARE–promoter (5). We first analyzed the recruitment of RA receptors and RNA polymerase II to the promoter and RARE sites. Second, we studied the changes of methylation of lysine 4 (K4) and lysine 9 (K9), following the recruitment on the chromatin sites of two demethylating enzymes, LSD1 (KDM1A) and JMJD2A (KDM4A). It has been reported that LSD1 demethylates H3K4me2 or H3K9me2 and JMJD2A demethylates H3K4-K9 me3 (6). Third, we analyzed the formation after RA exposure of specific chromatin-DNA domains that connect the 5′ end-promoter-RARE and the 3′ end site of the RA-target gene. Our data indicate that histone demethylation, DNA oxidation and chromatin looping induced by RA are temporally and causally associated with the onset of productive transcription induced by RA, and inhibition of any of these three events abolishes RA-induced transcription.

MATERIALS AND METHODS

Cells and transfections

Human breast cancer MCF-7 cells were grown at 37°C in 5% CO_2 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) (Invitrogen). Non-tumorigenic breast MCF10-2A cells were grown in a 1:1 mixture of DMEM and Ham's F12 medium supplemented with 20-ng/ml epidermal growth factor, 100-ng/ml cholera toxin, 0.01-mg/ml insulin and 500-ng/ml hydrocortisone, 95%; horse serum, 5%. Cells were provided with fresh medium every 3 days. To evaluate the effect of RA, cells were grown in phenol red-free medium containing 10% dextran-charcoal-stripped FBS for 1–3 days, before being challenged with 300-nM RA for different times according to the experimental protocol.

To obtain *LSD1* knock down with siRNA, cells were transiently transfected, using a Neon® Transfection System, with siRNA in medium without serum to a final concentration of 10 nM and incubation was continued for 48 h. Scrambled RNA, at the same concentration, was used as negative control. The same procedure was used to obtain *JMJD2A, OGG1* and *APE1* knock down with the specific siRNAs (see Supplementary Information). To rescue LSD1 activity in knock down experiments with siRNAs, LSD1 full-length cDNA was inserted into the CMV 3xFLAG expression vector (Sigma-Aldrich). LSD1ALA mutant has been described elsewhere (7,8). To rescue JMJD2A activity, cells were transfected with pCMV6-AC-GFP plasmid containing JMJD2A full-length (RG200574, OriGene Technologies, Inc.). To assess the transfection efficiency at single cell level, all transfections were traced with pEGFP Vector (Clontech) or with BLOCK-iT Alexa Fluor® Red Fluorescent Control and analyzed by fluorescence-activated cell sorting (FACS).

RNA extraction and quantitative reverse transcription polymerase chain reaction and quantitative polymerase chain reaction

Total RNA was extracted using TRIzol (Gibco/Invitrogen). cDNA was synthesized in a 20-μl reaction containing 1 μg of total RNA, 100 U of Superscript III Reverse Transcriptase (Invitrogen) and 2 μl random hexamer (20 ng/μl) (Invitrogen). mRNA was reverse-transcribed for 1 h at 50°C, and the reaction was heat inactivated for 15 min at 70°C. The products were stored at −20°C. Quantitative reverse transcription polymerase chain reactions (PCRs) and quantitative PCRs (qPCRs) were performed on a 7500 Real Time PCR System (Applied Biosystems) using the SYBR Green-detection system (FS Universal SYBR Green MasterRox/Roche Applied Science). The complete list of oligonucleotides used is reported in Supplementary Table S1.

Chromatin immunoprecipitation

Cells were transfected and/or treated as indicated in the legends of the figures. The cells (∼2.5 x 10⁶ for each antibody) were fixed for 10 min at room temperature by adding 1 volume of 2% formaldehyde to a final concentration of 1%; the reaction was quenched by 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 (see Supplemental Information) containing 1X protease inhibitor cocktail (Roche Applied Science). The lysates were sonicated to have DNA fragments 300–600 bp. Sonicated samples were centrifuged and supernatants diluted 2-fold in the chromatin immunoprecipitation (ChIP) buffer (Supplemental Information). An aliquot (1/10) of sheared chromatin was further treated with proteinase K, extracted with phenol/chloroform and precipitated 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) and 20 μl of Protein A/G PLUS-Agarose (Santa Cruz Biotechnology) saturated with salmon sperm (1 mg/ml). Precleared chromatin was divided in aliquots and incubated at 4°C for 16 h with 1 μg of the specific antibody (Supplemental Information) and non-immune IgG, respectively. The immunocomplexes were recovered by incubation for 3 h at 4°C with 20 μl of protein-A/G agarose, beads were washed with wash buffers according to the manufacturer's instructions and immunoprecipitated DNA was recovered and subjected to qPCR using the primers indicated in the legend of the specific figures, primers sequences and qPCR protocols are described in Supplemental Information and in Supplementary Table S1.

8-Oxo-7, 8-dihydro-2′-deoxyguanosine DNA assay

For 8-Oxo-7, 8-dihydro-2′-deoxyguanosine (8-oxo-dG) detection, 10⁶ MCF-7 cells were seeded onto glass slides and treated with 200 or 500-nM RA for 15–30 min. Control cultures were treated with equivalent vehicle volumes and concentrations. After treatments, the cells were fixed 15 min with 4% paraformaldehyde in phosphate buffered saline (PBS). The slides were then washed three times with Tris-buffered saline (TBS)/Tween-20 and permeabilized by serial washes in methanol solutions, prior to be washed with TBS/Tween-20, blocked for 1 h at 37°C and incubated with fluorescein isothiocyanate-labeled protein, that binds

8-oxo-dG, for 15 h at 4°C (Biotrin OxyDNA Test, Biotrin, UK). Cover slips were mounted in Moviol and viewed by fluorescence. To obtain LSD1 knock down, cells were transfected with specific or control siRNAs. After 48 h, cells were subjected to different treatments, according to experimental needs, and processed for fluorescence microscopy. For single-cell transfection assays, cells were co-transfected with BLOCK-iT Alexa Fluor® Red Fluorescent Control. The efficiency of transfection was 65%+10. All images were captured with Axiocam microscopy (Zeiss) with a 63x objective in the same conditions of brightness and contrast.

Chromosome conformation capture

The chromosome conformation capture (3C) assay was performed as described (9) with minor adaptations. Briefly, the NcoI restriction enzyme was used. The cells (2.5×10^6) were crosslinked in 12 ml of PBS with 1% formaldehyde for 10 min at room temperature. The reaction was quenched by addition of glycine to a final concentration of 125 mM. Fixed cells were harvested and the pellet resuspended in 1 ml of ice-cold lysis buffer (the same used for ChIP experiments). Nuclei were washed with 0.5 ml of restriction enzyme buffer, centrifuged and resuspended in 100 μl of restriction enzyme buffer. Sodium dodecyl sulphate (SDS) was added to a final concentration of 0.1%, and nuclei were incubated at 37°C for 15 min. Triton X-100 was added to a final concentration of 1% to sequester SDS. Digestion was performed with 100 U of restriction enzyme at 37°C for 16 h. The restriction enzyme was inactivated by addition of SDS to 2% and incubation at 65°C for 30 min. The reaction was diluted into 1-ml ligation reaction buffer (see Supplemental Methods) and incubated at 16°C for 18 h with 50 U of T4 DNA Ligase (Roche Applied Science). Ethylenediaminetetraacetic acid (10 mM) was added to stop the reactions. Samples were treated with Proteinase K (200 μg/ml) and incubated for 5 h at 55°C and then overnight at 65°C to reverse the formaldehyde crosslinks. The following day, the DNA was purified by phenol/chloroform extraction and ethanol precipitation. Samples were redissolved in 20 μl of Tris-EDTA (TE) buffer. To prepare a control template, we used a pool of plasmids containing an equimolar amount of the *CASP9* or *CYP26A1* inserts spanning the genomic regions of interest. Five micrograms of plasmid DNA were digested with NcoI in 50 μl of 1x buffer for 8 h at 37°C and then ligated in 20 μl with 5 U of T4 Ligase at 16°C for 4 h. The efficiency of digestion at the end of 3C treatment was quantified by real-time PCR, amplifying a fragment spanning two NcoI (uncut) in different 3C DNA preparations. Primer sequences and positions and details of PCR and qPCR protocols are described in Supplemental Information and Supplementary Table S1. PCR products were separated on 1.2% agarose gels, stained with ethidium bromide and quantified with the imageJ program (Rasband WS, ImageJ, National Institutes of Health, Bethesda, MD, USA; http: //rsb.info.nih.gov/ij/). The amplified fragments at the end of the procedure were verified by DNA sequence analysis.

Statistical analysis

All data are presented as mean ± standard deviation in at least three experiments in triplicate ($n \geq 9$). Statistical significance between groups was determined using Student's *t*-test (matched pairs test or unmatched test was used as indicated in figure legends). All tests were performed using the JMP Statistical Discovery™ software by SAS.

RESULTS

Recruitment of RA receptor and activation of RNA polymerase II at RA-target promoters

The biological activity of RA is mediated by its binding to RA receptors (RARα, RARβ and RARγ), which function as hetero-dimers with RXRs, targeting specific sites (RAREs). We have used breast cancer cells MCF-7, because they express RARs and respond to RA (10). We studied induction by RA of *CASP9* and *CYP2626A1* (Figure 1a and Supplementary Figure S1a) by exposing MCF-7 cells to RA and measuring mRNA levels at different times following stimulation. Figure 1b and Supplementary Figure S1b show that both mRNAs accumulate 30 min after RA exposure. After a transient reduction at 60 min (a slight, but reproducible, decrease in *CYP26A1* mRNA), their levels reached a maximum 4 h after RA exposure. In *CASP9*, a RARE element is localized in intron II, 9.5 kb downstream of the TSS (Figure 1a). *CYP26A1*, instead, contains a RARE contiguous with the promoter (−150 bp from the TSS) and another site at the 3′ end close to the polyA (Supplementary Figure S1a; 11). These sites are essential for RA induction of transcription (4,12,13), which depends on RARα, because under the same conditions, RA is not able to induce *CASP9* and *CYP26A1* mRNAs in MCF10 cells, devoid of RA receptor (Supplementary Figure S1e and f) (14).

To monitor recruitment of the RA receptor to the promoter and RARE elements after RA stimulation, we assessed the timing of association of RA receptor with the chromatin of *CASP9* and *CYP26A1* by ChIP. We included in our analysis very early times after RA induction (min) to detect the earliest chromatin changes induced by the hormone. Figure 1c shows that RARα is rapidly (15 min) recruited to the RARE element and to the upstream promoter of *CASP9*. We noticed that the levels of RARα recruited to the promoter and RARE chromatin were not stable, but oscillated between 15 and 60 min after RA exposure. RARα and RXRα were also recruited to the RARE/promoter of *CYP26A1* (Supplementary Figure S1c and d). RARα did not accumulate to the RARE or promoter of *CASP9* gene in unresponsive MCF10 cells (Supplementary Figure S1f). Recruitment of RARα and RXRα was associated with accumulation of total and serine 5-phosphorylated RNA polymerase II (Pol II and p-Pol II, respectively) at the RARE and promoter regions of *CASP9* and *CYP26A1* (Figure 1d and e, and Supplementary Figure S1d). As expected, Pol II and p-Pol II accumulated preferentially at the promoter relative to the RARE following RA induction. Note that recruitment of total Pol II and p-Pol II to RARE oscillated synchronously with the recruitment of RARα. p-Pol II progressively accumulated at the *CASP9* promoter with the time of RA stimulation (Figure 1d and e, and Sup-
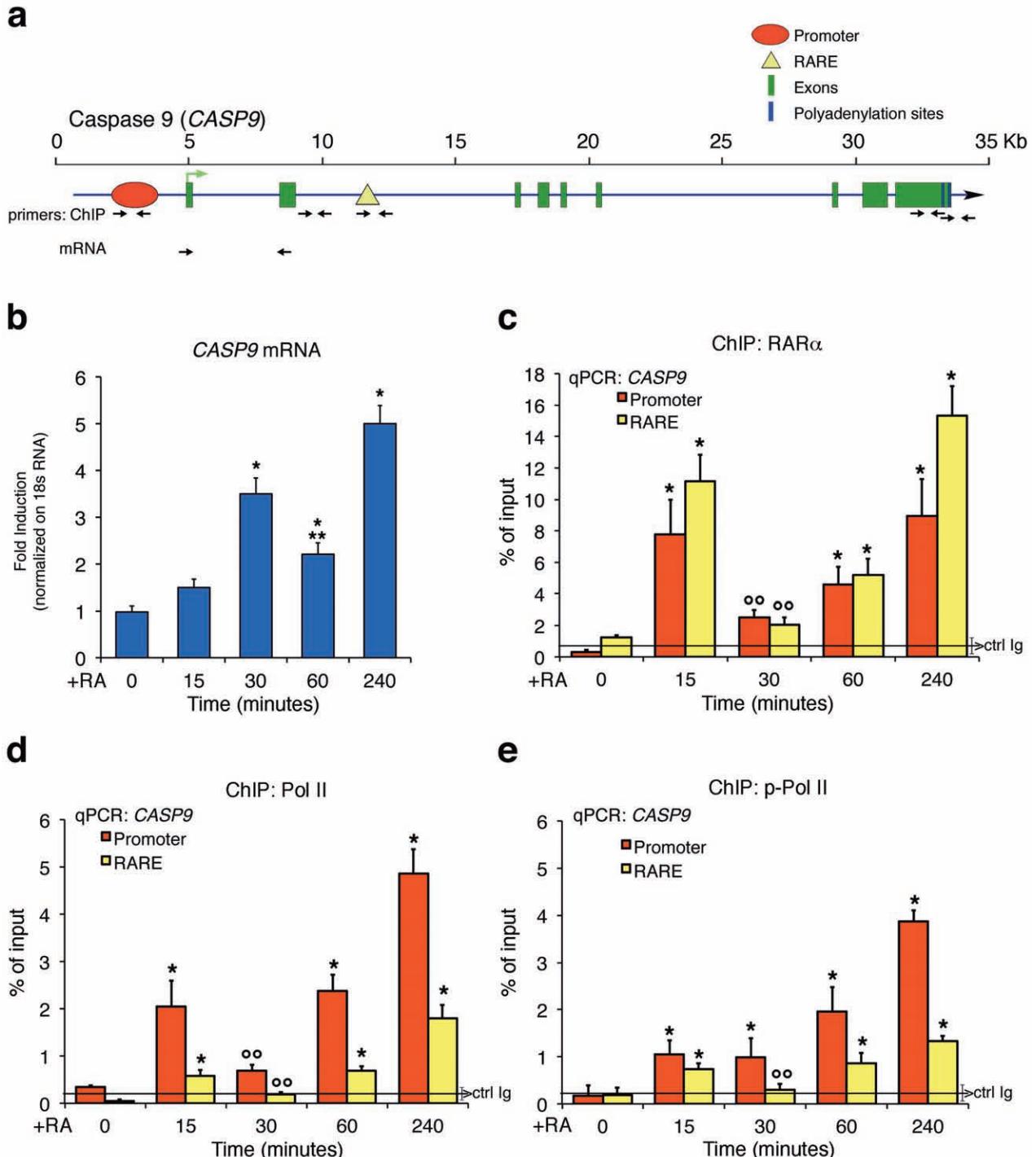
Figure 1. Retinoic acid (RA) induction of *CASP9* mRNA and recruitment of retinoic acid receptor alpha (RARα) and phosphorylated RNA polymerase II to retinoic responsive element (RARE) and promoter of *CASP9* gene. (**a**) Structure of *CASP9* gene. The TSS and the direction of transcription are indicated by a green arrow; the exons, promoter, polyA addition sites and RARE are shown by different colors indicated at the upper right corner. The black arrows indicate the primers used for ChIP and mRNA analysis. (**b**) Total RNA was prepared from MCF-7 hormone-starved or stimulated with 300-nM RA for 15, 30, 60 and 240 min and analyzed by qPCR with specific primers (panel (a)) to *CASP9* mRNA normalized to 18S RNA levels. The statistical analysis derived from at least three experiments in triplicate ($n \geq 9$; mean $\pm$ SD); *$P < 0.01$ (matched pairs *t*-test) compared to RA-unstimulated sample, **$P < 0.01$ (matched pairs *t*-test) comparing 30–60 min of RA exposure. (**c, d, e**) qChip analysis of RA-dependent occupancy of RARα, RNA polymerase II (Pol II) and phosphorylated RNA polymerase II (p-Pol II) at the promoter and RARE. MCF7 cells were stimulated with 300-nM RA for 15, 30, 60 and 240 min. The chromatin was immunoprecipitated with antibodies directed against RARα, Pol II, p-Pol II. Panel (c) shows the recruitment of RARα to the promoter and RARE sequences of *CASP9* gene. Panels (d) and (e) show the recruitment of Pol II (d) and p-Pol II (e) at the promoter and RARE. The black, horizontal line (brackets $\pm$ SD) in each plot indicates the percent of input from a control ChIP (Ab: non-immune serum). The statistical analysis derives from at least three experiments in triplicate ($n \geq 9$; mean $\pm$ SD); *$P < 0.01$ (matched pairs *t*-test) compared to RA-unstimulated sample; °°$P < 0.01$ (matched pairs *t*-test) compared 15–30-min stimulated samples.

41

plementary Figure S1d). The levels of p-Pol II at the *CASP9* RARE dipped at 30 min and then increased over the next 240 min, preceding by 30 min the oscillations of the mRNA levels (Figure 1b). The oscillation of RARα recruitment to the *CYP26* RARE promoter–RARE is delayed by 30 min compared to that of *CASP9* and mirrors the smaller and higher peaks of mRNA accumulation at 30 and 240 min following RA exposure (Supplementary Figure S1c and d). mRNA levels of four other genes induced by RA display a similar pattern with a small peak at 30–60 min and a large peak at 240 min after RA (not shown). In *CASP9* normalization of p-Pol II with total Pol II shows that this ratio increases dramatically 30 min after RA at the promoter, while it is stable at the RARE (Figure 1d and e) mirroring the rise and reduction of RNA at 30 and 60-min RA, respectively (Figure 1b and Supplementary Figure S1b). However, we note that at 30 min the change of the ratio is due to a significant depletion of total PolII from the promoter (Figure 1d) which may explain the reduction of mRNA 30 min later (Figure 1b). We believe that the loss of total Pol II at the promoter 30 min after RA does not reflect an increased Pol II phosphorylation, but instead, a change in chromatin structure leading to depletion of Pol II with other proteins bound to the promoter, including RARα (Figure 1c).

Histone H3K4 and H3K9 methylation marks induced by RA

Methylation of lysine 4 in histone H3 (H3K4) marks transcribed loci, whereas dimethyl-lysine 9 in the same histone (H3K9me2) is associated with transcription silencing (15). To find the histone marks modified by RA exposure, we analyzed the methylation profiles of H3K4 and H3K9 in cells after treatment with RA. ChIP analysis was performed with specific antibodies against methylated H3K4 and H3K9. The regions analyzed were the promoter-TSS, the RARE element and two polyA addition sites located at the 3′ end of the *CASP9*. Two segments of the gene 2 kb distant from the TSS were also included in the analysis. As a result of methylation and demethylation events induced by RA, the levels of methylated K9 or K4, normalized to the input chromatin, were selectively modified shortly after RA exposure. Figure 2 shows that promoter-associated H3K4 me2 and me3 were transiently demethylated 15 min after RA challenge and then progressively remethylated (Figure 2a and b). Rapid demethylation of H3K4 at RARE was also observed, but remethylation was slower than at the promoter. RA induced a transient loss of H3K9me2 and H3K9me3, followed by accumulation of H3K9me2 but not of K3K9me3 (Figure 2c and d). The methylation changes observed were not due to loss of H3 levels at these sites (Figure 2e and f). These cyclical methylation–demethylation events were strikingly synchronous and specific to the RARE and promoter regions of *CASP9*. In contrast, H3 methylation was unaffected by RA at a site 2 kb downstream of RARE in intron II or in a non-RA-induced gene, TGFBI (Figure 2g). A similar methylation–demethylation cycle was observed at the *CYP26A1* promoter–RARE chromatin (Supplementary Figure S2a and b). MCF10 cells, which do not respond to RA (14; Supplementary Figure S1e and f), do not show changes in methylation of H3K4, K9 me2/3 at 15-min RA (Supplementary Figure S2c). We also probed

the 3′ end of *CASP9* gene, where two major polyA addition sites are located. H3K4me2 and me3 and H3K9me2 at the polyA1 and polyA2 sites also underwent transient and early demethylation. H3K9me3 was permanently demethylated at the polyA2 site, but was essentially unchanged at the polyA1 site (Supplementary Figure S2d–i). We conclude that the polyA1 and polyA2 sites undergo methylation changes similar to those seen at the promoter and RARE, raising the possibility that these regions are functionally and physically associated in a unique chromatin domain.

To complete the description of chromatin factors directly or indirectly linked to the modification of histone marks following RA induction of transcription, we measured the recruitment of a major histone methyltransferase (SUV39H1, which methylates H3K9) (16,17) (Supplementary Figure S3a and b) or an alternative modification of H3K9, such as H3K9 acetylated (H3K9Ac) (Supplementary Figure S3c and d), or the recruitment of the co-repressors, NCoR1 and NCoR2 (18,19) (Supplementary Figures S3e and f and S3g and h, respectively) to the RARE–promoter and polyA addition sites of *CASP9* and *CYP26A1* genes. Supplementary Figure S3 shows that the recruitment of these proteins to these sites following RA induction mirrors the demethylation–methylation cycles described above. SUV39H1 accumulates at the promoter, RARE and polyA addition sites with a sharp peak 30 min following RA exposure (Supplementary Figure S3a and b). The SUV39H1 recruitment overlaps with the peak at 30 min of bi- and tri-methylated H3K9 in Figure 2c and d. On the other hand, H3K9Ac oscillates symmetrically with the demethylation wave of H3K9me2 and H3K9me3, suggesting that H3K9 nucleosomes undergo mutually exclusive demethylation–acetylation cycles (Supplementary Figure S3c and d). NCoR1 is transiently recruited at 15-min RA, while NCoR2 is present in the absence of RA and progressively disappears from the RARE chromatin following RA induction (Supplementary Figure S3e–h). Collectively, these data describe a series of intertwined oscillating cyclic events driven by RA and targeted to the RARE–polyA addition sites of *CASP9* and *CYP26A1* genes. The earliest temporal event we are able to detect following RA exposure is the massive demethylation of H3K4 and H3K9 induced by RA at 15 min. This suggests that K4 and K9 demethylation enzymes are also recruited to *CASP9* and *CYP26A1*. H3K9me3 can be demethylated by enzymes of the Jumonji C (JMJC) family (20,21), whereas H3K4me2 is demethylated *in vitro* (22) and H3K9me2 *in vivo* (7,23–25) by LSD1 (KDM1). We decided to monitor the recruitment at the RARE promoter chromatin of *CASP9* and *CYP26A1* of LSD1 and JMJD2A, which demethylate H3K4me2 or H3K9me2 and H3K9me3, respectively. Figure 3a, b and Supplementary Figure S4a show recruitment of both LSD1 and JMJD2A histone demethylases to the RARE elements and promoters of *CASP9* and *CYP26A1* following RA treatment. Notably, the kinetics of recruitment of LSD1 and JMJD2A parallels the kinetics of loss of the H3K4me2 and H3K9me2/3. H3K9me2/3 accumulation at 30-min RA is associated with the recruitment of SUV39H1 (16) to the promoter, RARE and polyA addi-
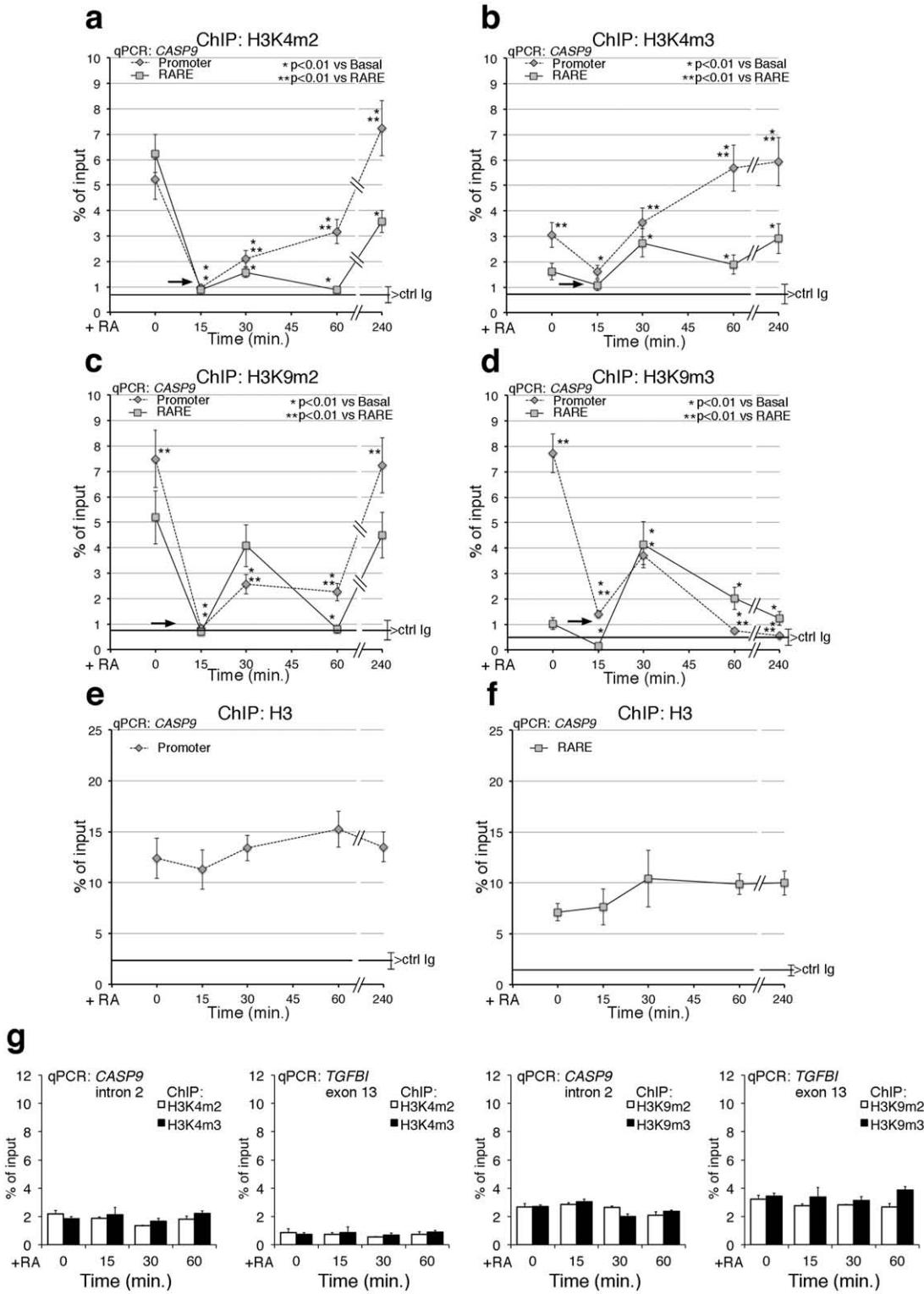
Figure 2. Methylation–demethylation cycles of histone H3K4/K9 induced by RA at *CASP9* promoter–RARE chromatin. MCF7 cells were serum starved and exposed to 300-nM RA at the indicated times (0, 15, 30, 60 and 240 min). qChIP was carried out using specific antibodies recognizing H3K4me3, H3K4me2, H3K9me3 and H3K9me2. The specificity of the antibodies was tested by competition with the specific methylated peptide(s). (**a, b**) H3K4me2 and H3K4me3 occupancy at the *CASP9* promoter and RARE. (**c, d**) H3K9me2 and H3K9me3 occupancy at the *CASP9* promoter and RARE. The black arrows indicate the loss of all methylated H3K4 and H3K9 (except H3K4me2 at the RARE) at 15-min RA. (**e, f**) ChIP Histone H3 at the promoter and RARE after RA induction. (**g**) ChIP analysis of *CASP9* II intron and of TGFBI exon 13 (non-RA-induced gene), in cells exposed to RA for 15, 30 and 60 min. For each antibody, the control IgG signal is shown (brackets ± SD). The statistical analysis was derives from at least three experiments in triplicate ($n \geq 9$; mean ± SD); *$P < 0.01$ (matched pairs *t*-test): compared to the RA-unstimulated sample; **$P < 0.01$ (student *t*-test): compared to the same time point at the promoter and RARE.

43

Figure 3. LSD1 and JMJD2A upon RA exposure are recruited to the *CASP9* promoter and RARE sites and are essential for *CASP9* induction by RA and methylation–demethylation cycle. (**a, b**) Recruitment of LSD1 and JMJD2A to the promoter and RARE of *CASP9* gene. MCF7 cells were serum starved and exposed to 300-nM RA at the indicated times (15, 30, 60 and 240 min). qChIP was carried out using specific antibo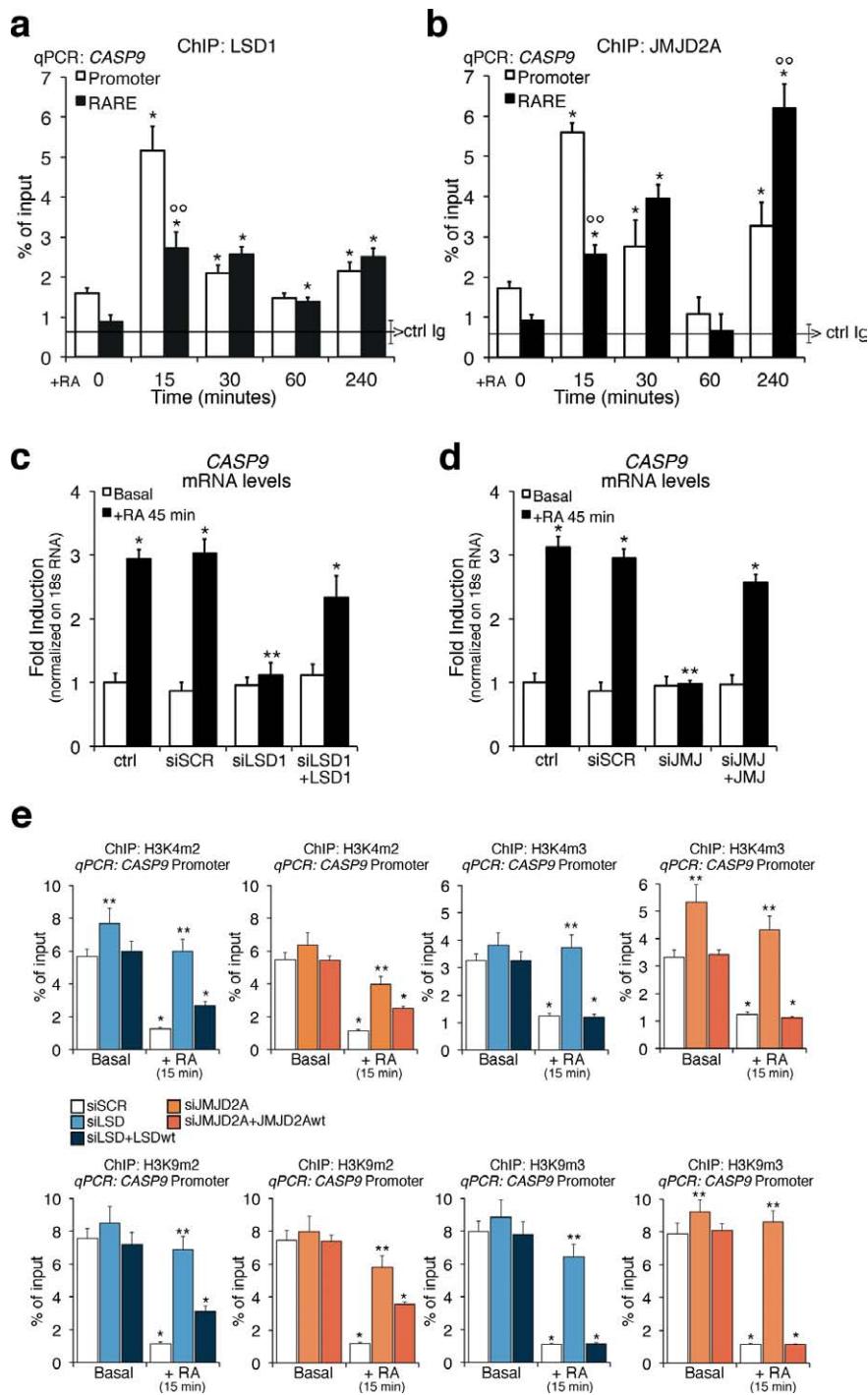dies recognizing LSD1 and JMJD2A. The panels show the time course of the recruitment of LSD1 (a) and JMJD2A (b) to the RARE and promoter sequences of *CASP9* analyzed by qPCR. The black, horizontal line in each plot indicates the percent of input from a control ChIP (Ab: non-immune serum). (**c, d**) LSD1 and JMJD2A depletion inhibits RA-induced transcription of *CASP9*. MCF7 were transiently transfected with LSD1 siRNA or JMJD2A siRNA. After 48 h, total RNA was prepared from control cells (starved) or RA-induced cells (300-nM RA for 45 min) and analyzed by qPCR with specific primers to *CASP9* mRNA. The statistical analysis derives from at least three experiments in triplicate ($n \geq 9$; mean $\pm$ SD); *$P < 0.01$ (matched pairs t-test) compared to RA-unstimulated sample; **$P < 0.01$ (student t-test): comparison between SCR and specific siRNAs; °°$P < 0.01$ (student t-test): comparison between promoter and RARE regions. (**e**) Depletion of LSD1 or JMJ2DA inhibits the methylation changes induced by RA. qChIP was performed on cells transfected with LSD1 siRNA or JMJD2A siRNA and induced with 300-nM RA for 15 min. qChIP was carried out using specific antibodies recognizing H3K4me3, H3K4me2, H3K9me3 and H3K9me2. The statistical analysis derives from at least three experiments in triplicate ($n \geq 9$; mean $\pm$ SD); *$P < 0.01$ (matched pairs t-test) compared to RA-unstimulated sample; **$P < 0.01$ (student t-test): comparison between SCR and specific siRNAs. Transfection efficiency was monitored by FACS (Alexa Fluor or co-transfected pEGFP Vector).

tion sites (Supplementary Figure S3a and b). A new wave of demethylation of H3K9me3 occurs at 60–260-min RA as indicated by reduction of H3K9me3 and accumulation of H3K9me2 (Figure 2c and d). The methylation changes (H3K9me2/3) and the kinetics of LSD1 and JMJD2A recruitment to *CYP26A1* chromatin are very similar to those seen at RARE (Figure 2b and Supplementary Figures S2b, 3a and b and S4a). We believe that this similarity is due to the fact that the promoter and RARE are physically contiguous in *CYP26A1* but are separated in *CASP9*.

To demonstrate that both lysine demethylases were necessary for RA-induced transcription, we knocked down *LSD1* and *JMJD2A* with specific siRNAs and induced the cells with RA (Supplementary Figure S4b and c). Knock down of either demethylase, as shown by retention of di- and tri-methylated H3K4 and H3K9, significantly reduced RA-induced expression of *CASP9* (Figure 3c and d) and *CYP26A1* (Supplementary Figure S4d and e). Re-expression of the wild-type LSD1 or JMJD2A and not a mutant form of these enzymes restored RA-induced *CASP9* RNA levels (Figure 3c and d and data not shown) and reduced the levels of H3K4 and H3K9 both me2 and me3 as in control cells (Figure 3e). These data were unexpected because LSD1 and JMJD2A do not demethylate H3K4me3, although JMJD2A has been shown to bind H3K4me3 *in vitro* (26,27). We suggest that JMJD2A targets to H3K4me3 another enzyme able to demethylate H3K4me3. Also, silencing of either *LSD1* or *JMJD2A* increases H3K4me3 levels, as a result of increased methylation of the me2 forms, which accumulate when LSD1 is not functional.

To explore further the relationship between H3K9 and H3K4 methylation and LSD1, we overexpressed an N-terminal dominant-negative mutant (T110A) of LSD1 (LSD1ALA). This mutant is still enzymatically active, but is unable to target transcription factors (7,8). The LSD1ALA mutant protein was defective in binding to the promoter or RARE elements of *CASP9* or *CYP26A1* following RA induction (Figure 4a and Supplementary Figure S5a, respectively) and inhibited activation of *CASP9* or *CYP26A1* transcription upon RA exposure (Figure 4b and Supplementary Figure S5b). The methylation levels both H3K4 and H3K9 me2 were already low in the absence of RA (basal in Figure 4c–f) because LSD1ALA was constitutively active and not inducible by RA (Supplementary Figure S5c and d). LSD1ALA was constitutively recruited to chromatin because not only it inhibited RA induction of *CASP9* or *CYP26A1* expression but also lowered the levels of methylated H3K4me2 or H3K9me3 at promoters of non-RA-induced genes (such as *TGFBI*; Figure 4g), which methylation status does not change upon RA induction. However, the methylation–demethylation cycle at the RARE, promoter and polyA addition sites (Figure 4 and Supplementary Figure S56) was abolished both in LSD1 and JMJD2A-depleted cells (Figure 3e) and in LSD1ALA expressing cells (Figure 4c–f) and was not dependent upon the basal methylation levels of H3K4 and H3K9 (me2/3), which were high in the former and low in the latter cells, respectively (Figures 3e and 4c–f).

Collectively, these data indicate that RA induces the recruitment of both LSD1 and JMJD2A to the chromatin of *CASP9* and *CYP26A1* and demethylation of H3K4me2 and H3K9 me2/me3 at the RARE, promoter and polyA2 sites. These localized demethylation events by both demethylases are essential for the induction of transcription of *CASP9* and *CYP26A1* by RA (Figure 4b and Supplementary Figure S5b).

Recruitment of base or nucleotide excision repair enzymes to the RARE/promoter chromatin following RA induction

It was recently reported that nucleotide excision repair (NER) enzymes are recruited to promoter and RARE elements. NER is essential for the formation of discrete chromatin loops and induction of transcription by RA (28). Similarly, transcription-induced recruitment of base excision repair (BER) enzymes (such as 8-oxoguanine-glycosylase, OGG1) to *MYC*-Ebox DNA or to estrogen responsive elements has been described (25,29). Activation of a Fe^{++} dioxygenase (JMJD2A) and a FAD oxidase (LSD1) at the same chromatin sites (ERE or Ebox-promoters) triggers local oxidation. Oxidized guanine (8-oxo-dG) is recognized by OGG1 (8,25). That oxidation of guanine also occurs after RA induction of transcription is shown in Supplementary Figure S7a. We observed a rapid (15 min) nuclear accumulation of 8-oxo-dG in discrete foci in MCF7 cells exposed to RA. As predicted, production of 8-oxo-dG foci was inhibited by LSD1 knockdown (Supplementary Figure S7b). ChIP analysis showed that OGG1 was recruited to the promoter and RARE elements of *CASP9* (Figure 5a) 15 min following RA treatment. At 30 min, occupancy at these sites by OGG1 markedly decreased. At 240 min, OGG1 was detected at the RARE element but not at the promoter. Similar oscillation of OGG1 binding to the RARE/promoter of *CYP26A1* was also seen (Supplementary Figure S8a). RA induces the binding of OGG1 to the ERE and promoter sites with the same timing of nuclear 8-oxo-dG accumulation (Supplementary Figure S7) indicating that 8-oxo-dG accumulates at specific segments of RARE, promoter and polyA addition sites. In fact, OGG1 recruitment was not seen in non RA-induced genes or in segments distant (1–2 kb) from the TSSs of *CASP9* or *CYP26A1* (Supplementary Figure S8g).

Complexes enucleated by OGG1 seem important not only for the repair of oxidized lesions but also for assembly of transcription initiation complexes at estrogen- or Myc-dependent promoters (8,25,30,31). To dissect the components of the OGG1 complex that could link repair and RA-induced transcription, we probed for other BER enzymes that associate with a RARE element or its cognate promoter after RA induction. Specifically, we monitored the recruitment of: (i) the **AP**urinic site **E**ndonuclease1, APE1, which recognizes the apurinic site generated by OGG1 and cleaves the phosphodiester backbone, immediately 5′ to the site; (ii) thymine DNA glycosylase (TDG), which is required for BER of deaminated methylcytosine, a frequent product of base oxidation; and (iii) uracil glycosylase (UNG), which removes uracil or oxidized cytosine. Figure 5b, c and d (Supplementary Figure S8a for *CYP26A1*) show that all these enzymes are recruited to the promoter and to RARE chromatin 15 min following RA induction similar to the recruitment of RARα and Pol II (compare Figure 5 with Figure 1). We knocked down two of these enzymes (OGG1 and
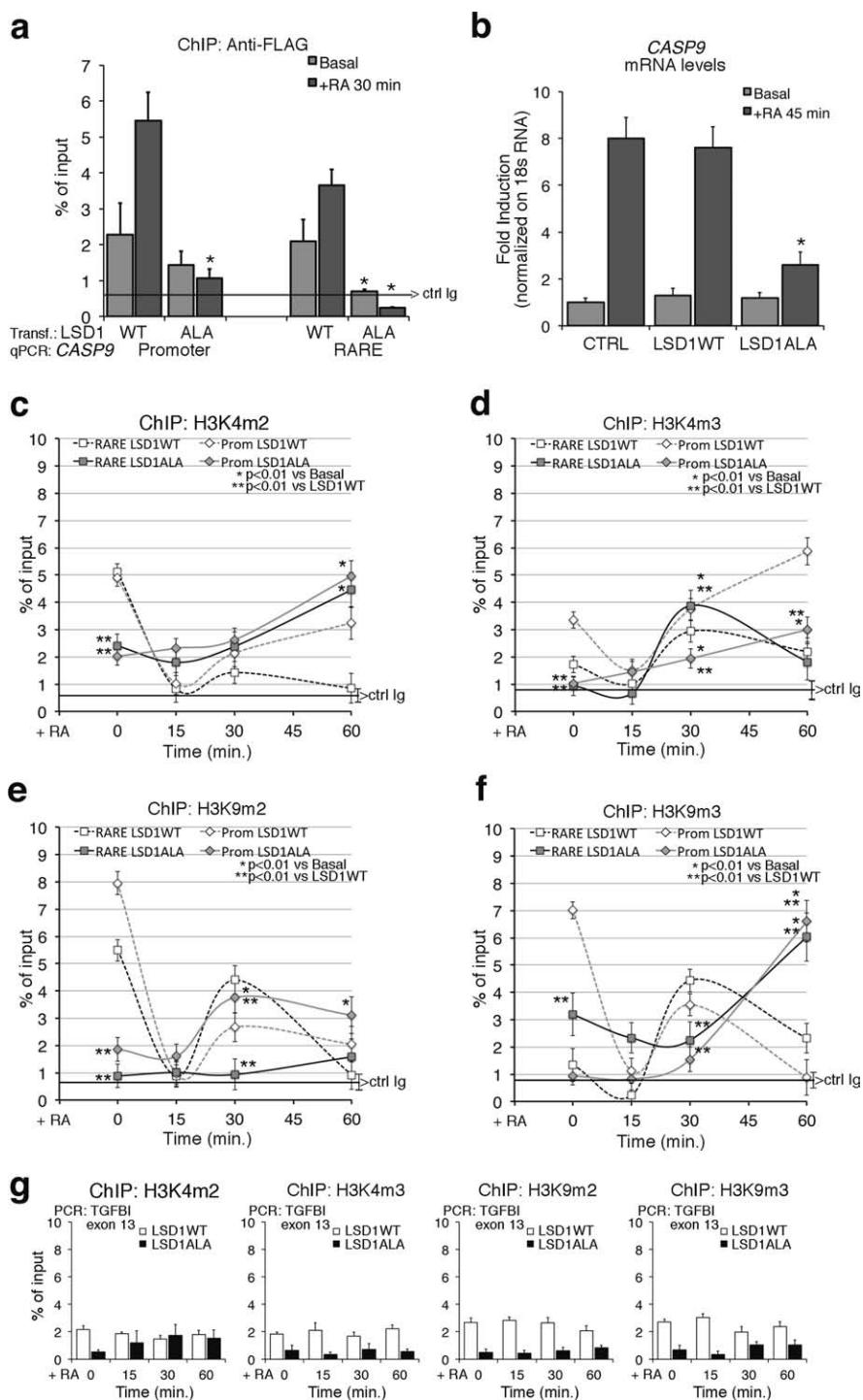
Figure 4. Expression of a dominant negative LSD1 mutant inhibits RA-induced methylation changes and *CASP9* induction by RA. (**a**) Recruitment of wild type (WT) and mutant LSD1 (LSD1ALA) to the *CASP9* RARE–promoter. The LSD1ALA mutant contains a substitution of threonine with alanine in position 110 and has been described elsewhere (17,19). MCF7 were transiently transfected with LSD1 vectors (WT or mutant), starved (Basal) or treated with 300-nM RA for 30 min and were analyzed by qChIP using Anti-FLAG antibodies to recognize the recombinant LSD1. The panel shows the recruitment of the flagged LSD1WT and LSD1ALA to the RARE and to the promoter sequences of *CASP9*. The black, horizontal line indicates the percent of input from a control ChIP (Ab: non-immune serum). (**b**) LSD1ALA inhibits RA-induced *CASP9* transcription. Total RNA was prepared from MCF7 transiently transfected with LSD1WT or LSD1ALA. After 48 h, mRNAs from control cells (starved) or RA-induced cells (300-nM RA for 45 min) were analyzed by qPCR with specific primers to *CASP9* mRNA. (**c, d**) LSD1ALA inhibits RA-induced H3K4 demethylation. H3K4me2 and H3K4me3 occupancy at the promoter and RARE of *CASP9* in LSD1ALA expressing cells as in (b). (**e, f**) LSD1ALA inhibits RA-induced H3K9 demethylation. H3K9me2 and H3K9me3 occupancy at the promoter and RARE of *CASP9* in LSD1ALA expressing cells as in (b). (**g**) ChIP analysis of TGFBI exon 13 (non-RA-induced gene), in transfected cells exposed to RA for 15, 30 and 60 min. For each antibody, the control IgG signal is shown (brackets ± SD). The statistical analysis derives from at least three experiments in triplicate ($n \geq 9$; mean ± SD); *$P < 0.01$ (matched pairs t-test): compared to the RA-unstimulated sample; **$P < 0.01$ (student t-test): comparison between control and the LSD1ALA expressing cells at the same time.
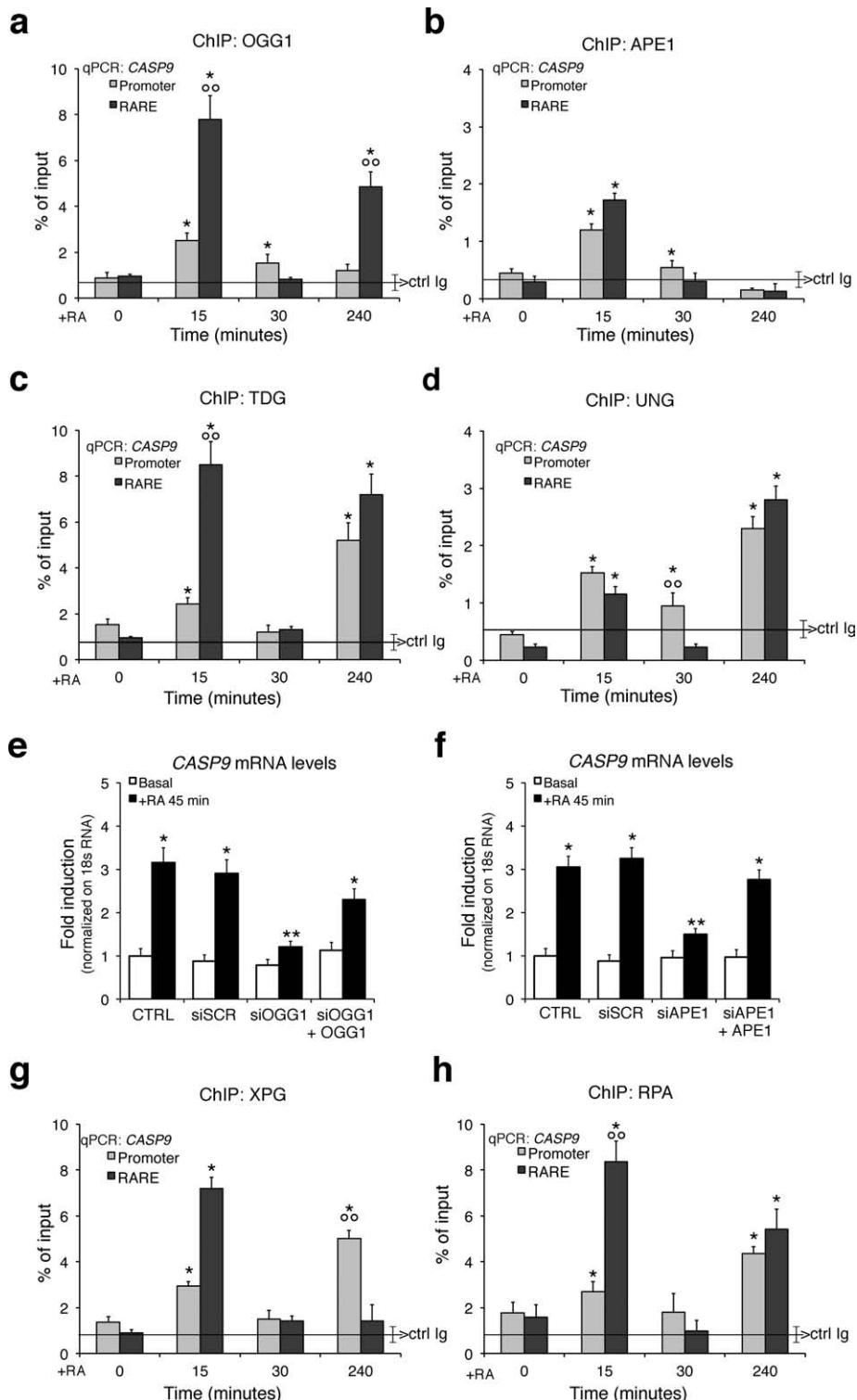
Figure 5. Recruitment of BER and NER enzymes to *CASP9* chromatin following RA induction. MCF7 cells, starved or treated with RA for 15, 30 and 240 min, were analyzed by qChIP using specific antibodies recognizing the 8-oxoguanine-DNA glycosylase-1 (OGG1), AP endonuclease (APE1), Thymine-DNA glycosylase (TDG), Uracil-DNA glycosylase (UNG), XPG and RPA. (**a, b**) The recruitment of OGG1 and APE1 to the *CASP9* promoter, and RARE sequences. (**c, d**) The recruitment of TDG and UNG to the same regions of *CASP9*. The black, horizontal line in each plot indicates the percent of input from a control ChIP (Ab: non-immune serum). (**e, f**) BER knockdown impairs the expression of RA-induced *CASP9*. Serum-deprived MCF7 cells were treated for 45 min with 300-nM RA and specific siRNA targeting OGG1 or APE1; *CASP9* expression levels were quantified by qPCR. To assess the transfection efficiency cells were co-transfected with pEGFP Vector (Clontech) and analyzed by FACS. (**g, h**) The recruitment of XPG and RPA to the *CASP9* promoter and RARE. The statistical analysis derives from at least three experiments in triplicate ($n \geq 9$; mean $\pm$ SD); *$P < 0.01$ (matched pairs t-test) compared to RA-unstimulated sample; **$P < 0.01$ (student t-test): comparison between SCR and specific siRNA; °°$P < 0.01$ (student t-test): comparison between the two amplicons.

47

APE1; Supplementary Figure S8b and c) and asked if this impacted on RA-induced transcription. Figure 5e and f and Supplementary Figure S8d–f summarize the results of these experiments. The activity of OGG1 and APE1 was essential for RA-induced transcription, since the wild type, but not a OGG1 dominant negative or a catalytically defective mutant APE1, reactivate *CASP9* transcription induced by RA in depleted cells (Figure 5e and f; Supplementary Figure S8d and e and data not shown).

Our results clearly indicate that depletion of these BER enzymes significantly reduced RA-induced transcription. We also probed for an NER enzyme (XPG) and for RPA, a single strand binding protein (Replication Protein A) involved in NER, on RARE by ChIP. XPG and RPA selectively accumulated at RARE chromatin following 15 min of RA stimulation (Figure 5g and h and Supplementary Figure S8a). It has been reported recruitment of NER enzymes to the RARE or other inducible promoters 3–4 h after hormonal induction, a period corresponding to maximal accumulation of specific mRNA levels (28). As expected, depletion of XPG and RPA reduced RA-induced transcription (not shown; 28). Our results show that BER and NER accumulate early at the RARE and promoter, shortly before we could detect mRNA accumulation (Figure 1b and Supplementary Figure S1b). The modifications we describe mark the first transcription cycle of *CASP9* and *CYP26A1* induced by RA.

Formation of dynamic chromatin loops governing the selection of 5′ and 3′ borders of RA-induced transcription units

The data shown above indicate that the *CASP9* and *CYP26A1* promoter, RARE and polyA addition sites undergo similar changes in histone H3K4 and H3K9 methylation and accumulate BER and NER enzymes after RA treatment. This coordination is consistent with the idea that these regions are physically associated after RA induction. Note that the *CASP9* RARE and 5′ start site are 9.5 kb apart and the polyA site is 22 kb to the 3′ end of RARE (Figure 1a). Recall that the methylation status H3K4 and H3K9 was not modified at chromatin neighboring these sites (2 kb at the 5′ and 3′ ends) and BER–NER enzymes were not recruited to these sites in RA-treated cells (Figure 2e and Supplementary Figure S8g).

These data suggest that RA induces early changes of specific chromatin domains that bring the TSS, the RARE and the polyA addition sites into close proximity. To find the relevant chromatin domains assembled in response to RA, we systematically analyzed the structure of *CASP9* (Figure 6a) and *CYP26A1* (Supplementary Figure S9a) chromatin by the 3C technique (see the Materials and Methods section). Briefly, fixed chromatin DNA was cleaved with a restriction enzyme (NcoI) and ligated after dilution. Real-time qPCR was then used to detect the ligated DNA segments. Figure 6a shows the summary of such analysis by using several probes and 'baits' centered on the TSS, RARE and 3′ end of *CASP9* (see the legend of Figure 6 for the detailed band analysis).

RA enhanced formation of two loops connecting the 5′ and 3′ ends of the gene with the RARE element. Extensive quantitative analysis of these loops revealed: (i) a 5′ end loop connecting the RARE to the promoter (A-F1, A-F2) was induced by RA (Figure 6b); (ii) a loop (F1-L, F2-L) connecting the RARE region to the 3′ end of *CASP9*, where two different polyA addition sites generate two mRNA ends (AB209147; AK303743). Assembly of this loop was almost entirely dependent on RA treatment (Figure 6c); (iii) a loop connecting the 5′ and the 3′ ends of the gene, bridging the above-mentioned loops (A-L) (Figure 6d). We wish to note that transcription of *CASP9* is promoted by many different stimuli. Thus, the loops formed in the absence of RA may nevertheless reflect basal transcription.

Strikingly, formation of all of these loops is cyclical. They first appear 15 min following RA exposure, disappear at 30 min and reform by 60 min. This oscillation resembles that displayed in previous figures showing *CASP9* mRNA synthesis, promoter and RARE occupancy by protein factors and histone modifications (Figures 1–5). The *CYP26A1* gene also formed chromatin loops upon RA treatment. Since RARE and promoter are contiguous in *CYP26A1* we detected essentially one major loop connecting the 5′ (promoter-RARE) with the 3′ end of the gene (polyA addition sites). This loop peaked 15 min after RA and slowly disappeared (Supplementary Figure S9b and c) similar to the early loop induced by RA on *CASP9* chromatin (Figure 6e). The physical association of the 5′ and 3′ ends of *CASP9* and *CYP26A1* genes induced by RA implies that the same proteins are present at the chromatin of the promoter, RARE and 3′ end sites. The physical contiguity (600 bp) of the two polyA addition sites (1 and 2) does not discriminate which polyA1 or 2 is included in the RA-induced loop of *CASP9* gene.

To find a 3′-end-specific RA-dependent marker of *CASP9*, we investigated the localization on *CASP9* chromatin of Ssu72, a protein which marks the 3′ end of genes and interacts with the general transcription initiation factor, TFIIB (32). Supplementary Figure S10a, b and c shows that Ssu72 binds the promoter and RARE with the same kinetics seen with RARα and Pol II following RA exposure, i.e. a peak at 15 min corresponding to the early RA-induced loop (Figure 6e). At 15 min after RA, Ssu72 disappeared from the polyA2 and concentrated at the promoter and RARE (Supplementary Figure S10a and b). Apparently, Ssu72 was present at the polyA2 site of *CASP9* gene in the absence of RA induction, except at 15′ min after RA, when the receptor and the promoter were recruiting Pol II and RARα (Supplementary Figure S10a–c).

How relevant are these loops to RA-induced transcription and how they are linked to the demethylation cycles induced by RA? To address this question we measured the loops involving RARE in cells expressing the LSD1ALA mutant. Expression of LSD1ALA inhibited RA-induced demethylation (Figure 4c–f and Supplementary Figure S6) and RA-induced transcription (Figure 4b). Supplementary Figure S10d shows that the formation of the 15-min loops connecting RARE to the polyA1/2 site or to the promoter upon RA exposure was inhibited: some loops were delayed (RARE-polyA1/2) and some others were completely eliminated (RARE–promoter).

We conclude that the demethylation cycles triggered by RA govern the ordered formation of the loops essential for RA-induced transcription.
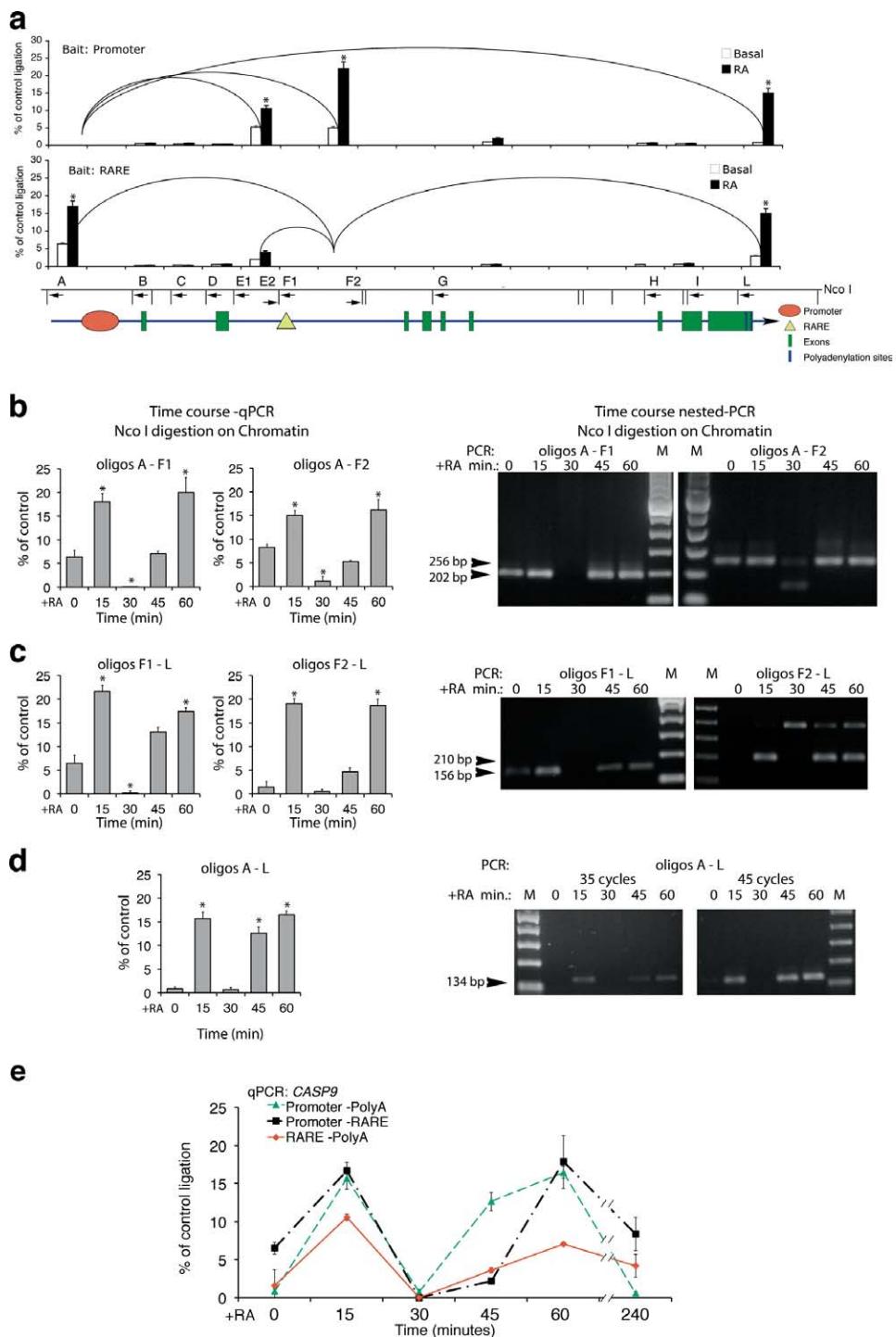
Figure 6. Formation of dynamic chromatin loops during early RA-induced transcription. 3C analysis of *CASP9* chromatin in MCF7 cells exposed to 300 nM of RA for various periods of time. (**a**) The histograms show the frequency of ligation of the *CASP9* NcoI fragments amplified with primers indicated below the NcoI restriction map. All the combinations of primers indicated were performed on ligated chromatin; the histogram shows qPCR amplifications above 1%, relative to the control. Each loop was detected with different primers pairs and the two histograms show the analysis by using several probes and 'baits' centered at the TSS, RARE and 3′ end of the *CASP9*. Differences between recombinant, Basal and RA-treated cells derived-chromatin were tested for statistical significance using Student's *t*-test: *$P < 0.01$ as compared to untreated control. (**b, c, d**) Time course of chromatin looping following RA induction. 3C analysis was carried out as described in the Materials and Methods section and the loops shown in panels (b, c, d) were quantified by qPCR (left panels) and verified by gel electrophoresis (right panels) and DNA sequencing (data not shown). The results shown derive from at least three experiments in triplicate ($n \geq 9$; mean ± SD). *$P < 0.01$ as compared to untreated samples. (**e**) The panel shows the time course of loop formation. Data were collected from real-time qPCR and from semi-quantitative, nested PCR, experiments. The extra band seen in panel (c) is a skipped restriction site that adds 187 bp to the expected fragment (210 bp). This band is specific because the same loop (RARE-polyA) analyzed with another primer (F1) does not produce extra bands. Also, following the treatment with RA, this longer fragment accumulates while the shorter one (cleaved by the enzyme in the chromatin) almost disappears. This indicates that this site is protected *in vivo* and is not cleaved efficiently by NcoI restriction enzyme.

49

DISCUSSION

The data reported here show that the methylation changes of K4 and K9 of histone H3 are linked to the recruitment of repair enzymes and, most importantly, to the formation of chromatin-DNA loops that juxtapose the 5′ end TSS, the enhancer (RARE) and the 3′ end of the transcribed gene (Figure 6 and Supplementary Figures S9 and S10). Histone methylation–demethylation cycles (33) and the formation of loops connecting the 5′ gene ends, 3′ ends and enhancers have been described extensively in many genes induced by nuclear receptors (28,33). However, these aspects of chromatin movement, although required for transcription induction by nuclear hormones, have not been mechanistically and temporally linked. Likewise, the association of methylation–demethylation cycles with NER enzyme recruitment to RA-induced promoter(s) and the formation of loops upon exposure to RA has not been described, although depletion of these enzymes seriously compromises transcription and chromatin looping induced by RA (28). However, notwithstanding the plethora of data, the mechanism used by RA or other inducers to trigger the recruitment of NER enzymes and formation of chromatin loops is still not known. Our experiments show that demethylation–methylation cycles of H3K4 and K9 at the RARE, promoter (Figure 2) and 3′ gene end (Supplementary Figure S2d–g) are the initiating and earliest events induced by the recruitment of the RA–RARα complex to the RARE. The timing and the kinetics of demethylation of H3K4 and K9 are similar at the promoter, RARE (Figure 6d) and the 3′ end of the gene, suggesting that these sites, although not contiguous in the DNA, are included in the same complex, driven by active RA–RARα. Inhibition of demethylation of H3K4 and H3K9 by depleting the demethylating enzymes (LSD1 or JMJD2A) or by expressing a dominant negative LSD1 variant (7,8) inhibits RA-induced transcription (Figure 3c and d and Supplementary Figure S4d and e), nuclear dG oxidation (Supplementary Figure S7b), recruitment of BER (25) and NER enzymes and formation of loops induced by RA (Supplementary Figure S10d). Formation of the chromatin loops with discrete 5′ and 3′ borders is facilitated by local DNA oxidation following demethylation by LSD1, which presumably releases supercoiling and rigidity of the helix and targets BER and NER enzymes to oxidized bases (34; Antonio Pezone & Antonio Porcellini, manuscript in preparation). In all cases, BER and NER enzymes are essential to localize and repair DNA nicks and altered bases produced by dG and 5mdC oxidation (25,31,35). We wish to stress that inhibition of H3K4-K9 methylation–demethylation cycles or depletion of BER or NER enzymes or overexpression of LSD1ALA does not influence the expression or modify the chromatin of RA-independent genes (Figures 2g and 4g and Supplementary Figures S4f and g and S8f). Locally, LSD1-triggered demethylation produces hydrogen peroxide, which oxidizes dG (Supplementary Figure S7) and both H3K9me2 and H3K4me2, as LSD1 substrates (7,8,25) may favor dG oxidation. We do not know if H3K9m2/3 coexist with H3K4me2/3 nucleosomes, but our data show a rapid demethylation 15 min after RA, involving both H3K4 and H3K9 N-terminals at the RARE–promoter and polyA addition sites (Figure 2 and Supplementary Figure S2d–g). At these sites, 4 h after RA, the content of H3K9me3 and H3K4me3 is low and high, respectively, as expected in transcriptionally active genes (36). The presence of H3K4me2/3 and H3K9me2/3 at time 0 at the promoter or RARE may indicate that these sites are poised between repression and activation depending on the strength of the specific stimulus. Also, RARE nucleosomes containing H3K9me2 may express less CASP9 and this may contribute to the setting of discrete expression states in cell populations.

Further complicating comparative analysis of the data published thus far are the different temporal frames used in various studies to describe the molecular events induced by transcriptional activators. These studies have been carried out at 1 h (37) to several hours or days (28) after hormonal induction. At these times, synchronization is lost. Each cell is starting and restarting the transcription cycle and only the mature RNA accumulates exponentially and can be easily detected even in asynchronous transcribing cell populations. The H3 (K4-K9) methylation code, hours and days after the initial RA induction, is not informative since H3K4me2/3 and H3K9me2/3 content is high and low both in control and chronically stimulated cells, respectively (33).

An important feature of RA-induced transcription reported here is the timing and synchronous oscillation with a period of ca. 30 min of chromatin bound RARα, Pol II, H3K4-H3K9 demethylation, H3K9Ac and looping involving the RARE and the 5′ and 3′ of *CASP9* and *CYP26A1* genes (Figure 6). RA-induced mRNA levels of *CASP9*, *CYP26A1* and five other genes (A. Pezone, unpublished observations) oscillate with a period of 60 min from the initial induction with RA. With the time (2–4 h) this periodic oscillation is lost (Figure 6). We do not know if this loss of synchrony with the time is due to our inability to track these chromatin changes in non-synchronized cell populations or whether the oscillation we observe is limited to the first transcription cycle. An early and similar cycle of transcription induced by estrogens has been reported to be unproductive in terms of RNA accumulation. This cycle is suggested to prepare the promoter for subsequent transcription followed by two different transcriptionally productive cycles (38–40). However, our data suggest that (i) the first cycle (unproductive in 38–40) is indeed the cycle (15–30 min) that sets and defines the physical borders of the transcription unit induced by the hormone (in our case RA) and (ii) the oscillations in RARα recruitment (Figure 1) and chromatin looping (Figure 6) are driven by methylation–demethylation cycles of histone H3K4 and H3K9, caused by alternate recruitment of demethylating (LSD1-JMJD2A in Figure 3a and b and Supplementary Figure S4a) and methylating (Supplementary Figure S3a and b) enzymes. We believe that these oscillations are important for BER and NER enzymes to repair oxidized DNA (Figure 5 and Supplementary Figure S8a) to restart the cycle.

We wish to note that although the structural organization of the RARE is different comparing *CASP9* (RARE 4 kb downstream the promoter) with *CYP26A1* (RARE contiguous with the promoter) genes, the timing of methylation–demethylation, the oscillation of RARα and

50

LSD1-JMJD2A recruitment to the RARE promoter appear very similar

Why should the definition of the physical borders of a transcription unit be important? Each gene in eukaryotes is indeed the target of many stimuli, which can independently induce transcription. The protein Ssu72 has been identified in yeast as an element required for marking the 3′ end of the chromatin loops and transcription directionality (41,42). The human protein Ssu72, although has not been shown to be involved in DNA-chromatin loops, may have a similar function in human cells as suggested by the crystal structure of the complex of RNA polymerase C terminal domain, a scaffold protein symplekin, and human Ssu72 (43).

The accumulation of Ssu72 in the various sites of *CASP9* gene, before and after RA induction, may indicate the changes of chromatin engaged in RA-induced and RA-independent transcription. This protein is present at 3′ end of *CASP9* 3′ end before RA induction, disappears from this site at 15 min and reappears 30 min after RA treatment (Supplementary Figure S10a–c). Fifteen minutes after RA stimulation, Ssu72 moves from the 3′ end to the *CASP9* TSS and RARE (Supplementary Figure S10a–c), where its concentration oscillates synchronously with methylation–demethylation cycles and recruitment of BER and NER enzymes (Figures 2 and 5 and Supplementary Figure S10). These data indicate that Ssu72 specifies the 3′ end of both RA-dependent and -independent transcription units (Supplementary Figure S10c), which are also marked by end- and time-selective-specific chromatin loops. Importantly, after RA induction, RA-independent transcription is shut off to allow the formation of RA-dependent loops. This is shown by the shift of Ssu72 from the polyA2 addition site (in the absence of RA) to the RARE (15 min after RA) (Supplementary Figure S10a–c) and by the time synchrony of the loops induced by RA (Figure 6e). LSD1 is essential for this event, because the synchrony of loop formation is lost in cells expressing the LSD1ALA mutant: some loops disappear (promoter-RARE) or are delayed (promoter-polyA or RARE-polyA) and resemble loops formed in the absence of RA (Supplementary Figure S10d). It is worth noting that LSD1 has been recently shown to control the rhythmicity and the circadian clock and that a mutant in the residue adjacent (aa 111) to that of LSD1ALA (aa 110) is unable to reset clock oscillations (44). We suggest that RA-induced synchronous demethylation–methylation cycles inhibit ongoing local transcription and trigger recruitment of BER and NER enzymes to the chromatin of *CASP9* promoter, RARE and 3′ end. The simultaneous DNA and histone changes bring in close proximity the 5′, enhancer and 3′ ends of *CASP9* and induce the formation of loops (Figure 6 and Supplementary Figure S10). Figure 7 shows a summary of temporal changes induced by RA on the target chromatin, including histone H3K9 acetylation, methylation, demethylation, recruitment of co-repressors NCoR1-2, of BER and NER enzymes and the formation of the loop connecting RARE and promoter in *CASP9* gene. We propose that these cycles are intertwined and triggered by histone methylation–demethylation cycles.

We believe that RA-induced synchronization of DNA chromatin loops is required for calibration and rapid rein-
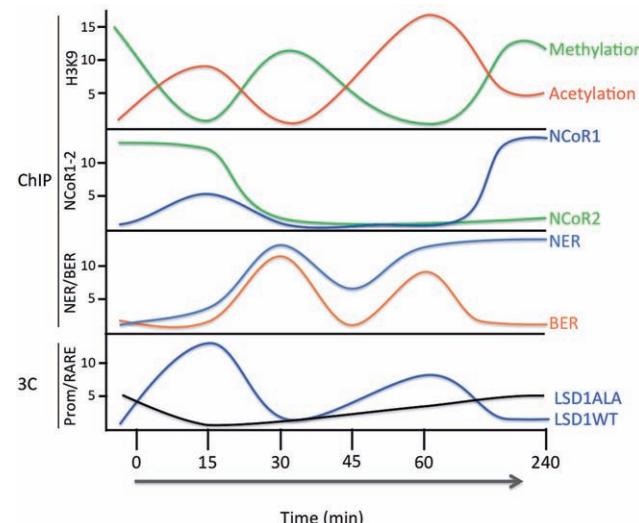
Figure 7. RA-induced cycles of histone H3K9 methylation, acetylation, recruitment of NCoR1–2, BER, NER and RARE–promoter chromatin loop. Time course of the major modifications of histone H3K9 at the RARE of *CASP9* and *CYP26A1* genes. The original data are in Figures 2, 5 and 6 and Supplementary Figures S2, S3, S8, S9 and S10. On the ordinate are indicated: i. fold-induction (arbitrary units); ii. the type of assay (ChIP or 3C). The time of RA induction is shown on the abscissa. The bottom panel shows the frequency of loop formation between RARE–promoter of *CASP9* gene in cells expressing wild type or LSD1ALA (Supplementary Figure S10d).

duction of transcription rather than for high levels of transcription. The massive accumulation of RA-specific mRNAs, in fact, occurs hours after the initial RA exposure (45). This efficient reinduction of transcription may represent a 'transcription memory', which has been noted before and requires gene looping and Ssu72 in yeast (46,47).

The periodic variation of recruited receptor at the enhancer represents a simple mechanism to titrate and calibrate the concentration of RA molecules present in the environment. Any drop in the inducer levels reduces the concentration of active receptor at the enhancer leading to dissolution of the chromatin DNA loop enucleated initially by the active receptor (RARE–promoter). Conversely, a rise in levels of the inducer rapidly reactivates transcription by stabilizing the loops (transcription memory) (see the Supplementary Movie).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENT

FUNDING

REFERENCES

1. Shahhoseini,M., Taghizadeh,Z., Hatami,M. and Baharvand,H. (2013) Retinoic acid dependent histone 3 demethylation of the clustered HOX genes during neural differentiation of human embryonic stem cells. *Biochem. Cell Biol.* **91**, 116–122.
2. Angrisano,T., Sacchetti,S., Natale,F., Cerrato,A., Pero,R., Keller,S., Peluso,S., Perillo,B., Avvedimento,V.E., Fusco,A. *et al.* (2011) Chromatin and DNA methylation dynamics during retinoic acid-induced RET gene transcriptional activation in neuroblastoma cells. *Nucleic Acids Res.* **39**, 1993–2006.
3. Compe,E. and Egly,J.M. (2012) TFIIH: when transcription met DNA repair. *Nat. Rev. Mol. Cell Biol.* **13**, 343–354 (Erratum in: *Nat. Rev. Mol. Cell Biol.* **13**, 476).
4. Donato,L.J. and Noy,N. (2005) Suppression of mammary carcinoma growth by retinoic acid: proapoptotic genes are targets for retinoic acid receptor and cellular retinoic acid-binding protein II signaling. *Cancer Res.* **65**, 8193–8199.
5. Balmer,J.E. and Blomhoff,R. (2002) Gene expression regulation by retinoic acid. *J. Lipid Res.* **43**, 1773–1808.
6. Kooistra,S.M. and Helin,K. (2012) Molecular mechanisms and potential functions of histone demethylases. *Nat. Rev. Mol. Cell Biol.* **13**, 297–311.
7. Ambrosio,R., Damiano,V., Sibilio,A., De Stefano,M.A., Avvedimento,V.E., Salvatore,D. and Dentice,M. (2013) Epigenetic control of type 2 and 3 deiodinases in myogenesis: role of Lysine-specific Demethylase enzyme and FoxO3. *Nucleic Acids Res.* **41**, 3551–3562.
8. Amente,S., Bertoni,A., Morano,A., Lania,L., Avvedimento,E.V. and Majello,B. (2010) LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc-induced transcription. *Oncogene* **29**, 3691–3702.
9. Dekker,J., Rippe,K., Dekker,M. and Kleckner,N. (2002) Capturing chromosome conformation. *Science* **295**, 1306–1311.
10. Teixeira,C. and Pratt,M.A. (1997) CDK2 is a target for retinoic acid-mediated growth inhibition in MCF-7 human breast cancer cells. *Mol. Endocrinol.* **11**, 1191–1202.
11. Mahony,S., Mazzoni,E.O., McCuine,S., Young,R.A., Wichterle,H. and Gifford,D.K. (2011) Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol.* **12**, R2.1–R2.15.
12. Loudig,O., Babichuk,C., White,J., Abu-Abed,S., Mueller,C and Petkovich,M. (2000) Cytochrome P450RAI(CYP26) promoter: a distinct composite retinoic acid response element underlies the complex regulation of retinoic acid metabolism. *Mol. Endocrinol.* **14**, 1483–1497.
13. Ray,W.J., Bain,G., Yao,M. and Gottlieb,D.I. (1997) CYP26, a novel mammalian cytochrome P450, is induced by retinoic acid and defines a new family. *J. Biol. Chem.* **272**, 18702–18708.
14. Papi,A., Guarnieri,T., Storci,G., Santini,D., Ceccarelli,C., Taffurelli,M., De Carolis,S., Avenia,N., Sanguinetti,A., Sidoni,A. *et al.* (2012) Nuclear receptors agonists exert opposing effects on the inflammation dependent survival of breast cancer stem cells. *Cell Death Differ.* **19**, 1208–1219.
15. Black,J.C. and Whetstine,J.R. (2011) Chromatin landscape: methylation beyond transcription. *Epigenetics* **6**, 9–15.
16. Wu,H., Min,J., Lunin,V.V., Antoshenko,T., Dombrovski,L., Zeng,H., Allali-Hassani,A., Campagna-Slater,V., Vedadi,M., Arrowsmith,C.H. *et al.* (2010) Structural biology of human H3K9 methyltransferases. *PLoS ONE* **5**, e8570.
17. Vaute,O., Nicolas,E., Vandel,L. and Trouche,D. (2002) Functional and physical interaction between the histone methyl transferase Suv39H1 and histone deacetylases. *Nucleic Acids Res.* **30**, 475–481.
18. Mottis,A., Mouchiroud,L. and Auwerx,J. (2013) Emerging roles of the corepressors NCoR1 and SMRT in homeostasis. *Genes Dev.* **27**, 819–835.
19. Glass,C.K. and Rosenfeld,M.G. (2000) The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev.* **14**, 121–141.
20. Loh,Y.H., Zhang,W., Chen,X., George,J. and Ng,H.H. (2007) Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* **21**, 2545–2557.
21. Wissmann,M., Yin,N., Müller,J.M., Greschik,H., Fodor,B.D., Jenuwein,T., Vogler,C., Schneider,R., Günther,T., Buettner,R. *et al.* (2007) Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat. Cell Biol.* **9**, 347–353.
22. Forneris,F., Binda,C., Vanoni,M.A., Battaglioli,E. and Mattevi,A. (2005) Human histone demethylase LSD1 reads the histone code. *J. Biol. Chem.* **280**, 41360–41365.
23. Metzger,E., Wissmann,M., Yin,N., Müller,J.M., Schneider,R., Peters,A.H., Günther,T., Buettner,R. and Schüle,R. (2005) LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* **437**, 436–439.
24. Musri,M.M., Carmona,M.C., Hanzu,F.A., Kaliman,P., Gomis,R. and Párrizas,M. (2010) Histone demethylase LSD1 regulates adipogenesis. *J. Biol. Chem.* **285**, 30034–30041.
25. Perillo,B., Ombra,M.N., Bertoni,A., Cuozzo,C., Sacchetti,S., Sasso,A., Chiariotti,L., Malorni,A., Abbondanza,C. and Avvedimento,E.V. (2008) DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. *Science* **319**, 202–206.
26. Huang,Y., Fang,J., Bedford,M.T., Zhang,Y. and Xu,R.M. (2006) Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* **312**, 748–751.
27. Cloos,P.A., Christensen,J., Agger,K. and Helin,K. (2008) Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes Dev.* **22**, 1115–1140.
28. Le May,N., Fradin,D., Iltis,I., Bougnères,P. and Egly,J.M. (2012) XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol. Cell* **47**, 622–632.
29. Amente,S., Lania,L., Avvedimento,E.V. and Majello,B. (2010) DNA oxidation drives Myc mediated transcription. *Cell Cycle* **9**, 3002–3004.
30. Huffman,J.L., Sundheim,O. and Tainer,J.A. (2005) DNA base damage recognition and removal: new twists and grooves. *Mutat. Res.* **577**, 55–76.
31. Fong,Y.W., Cattoglio,C. and Tjian,R. (2013) The intertwined roles of transcription and repair proteins. *Mol. Cell* **52**, 291–302.
32. He,X., Khan,A.U., Cheng,H., Pappas,D.L., Jr, Hampsey,M. and Moore,C.L. (2003) Functional interactions between the transcription and mRNA 3′ end processing machineries mediated by Ssu72 and Sub1. *Genes Dev.* **17**, 1030–1042.
33. Shi,L., Sun,L., Li,Q., Liang,J., Yu,W., Yi,X., Yang,X., Li,Y., Han,X., Zhang,Y. *et al.* (2011) Histone demethylase JMJD2B coordinates H3K4/H3K9 methylation and promotes hormonally responsive breast carcinogenesis. *Proc. Natl Acad. Sci. U.S.A.* **108**, 7541–7546.
34. Li,W., Notani,D., Ma,Q., Tanasa,B., Nunez,E., Chen,A.Y., Merkurjev,D., Zhang,J., Ohgi,K., Song,X. *et al.* (2013) Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520.
35. Parlanti,E., D'Errico,M., Degan,P., Calcagnile,A., Zijno,A., van der Pluijm,I., van der Horst,G.T., Biard,D.S. and Dogliotti,E. (2012) The cross talk between pathways in the repair of 8-oxo-7,8-dihydroguanine in mouse and human cells. *Free Radic. Biol. Med.* **53**, 2171–2177.
36. Lin,C., Yang,L., Tanasa,B., Hutt,K., Ju,B.G., Ohgi,K., Zhang,J., Rose,D.W., Fu,X.D., Glass,C.K. *et al.* (2009) Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* **139**, 1069–1083.
37. Liu,C.L., Kaplan,T., Kim,M., Buratowski,S., Schreiber,S.L., Friedman,N. and Rando,O.J. (2005) Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS Biol.* **3**, e328.
38. Métivier,R., Penot,G., Hübner,M.R., Reid,G., Brand,H., Kos,M. and Gannon,F. (2003) Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751–763.

39. Métivier,R., Reid,G. and Gannon,F. (2006) Transcription in four dimensions: nuclear receptor-directed initiation of gene expression. *EMBO Rep.* **7**, 161–167.

40. Gao,H. and Dahlman-Wright,K. (2011) The gene regulatory networks controlled by estrogens. *Mol. Cell. Endocrinol.* **334**, 83–90.

41. Hampsey,M., Singh,B.N., Ansari,A., LainŽ,J.P. and Krishnamurthy,S. (2011) Control of eukaryotic gene expression: gene loops and transcriptional memory. *Adv. Enzyme Regul.* **51**, 118–125.

42. Tan-Wong,S.M., Zaugg,J.B., Camblong,J., Xu,Z., Zhang,D.W., Mischo,H.E., Ansari,A.Z., Luscombe,N.M., Steinmetz,L.M. and Proudfoot,N.J. (2012) Gene loops enhance transcriptional directionality. *Science* **338**, 671–675.

43. Xiang,K., Nagaike,T., Xiang,S., Kilic,T., Beh,M.M., Manley,J.L. and Tong,L. (2010) Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature* **467**, 729–733.

44. Nam,H.J., Boo,K., Kim,D., Han,D.H., Choe,H.K., Kim,C.R., Sun,W., Kim,H., Kim,K., Lee,H. *et al.* (2014) Phosphorylation of LSD1 by PKCα is crucial for circadian rhythmicity and phase resetting. *Mol. Cell* **53**, 791–805.

45. Okuno,M., Caraveo,V.E., Goodman,D.S. and Blaner,W.S. (1995) Regulation of adipocyte gene expression by retinoic acid and hormones: effects on the gene encoding cellular retinol-binding protein. *J. Lipid Res.* **36**, 137–147.

46. Brickner,J.H. (2010) Transcriptional memory: staying in the loop. *Curr. Biol.* **20**, R20–R21.

47. Lainé,J.P., Singh,B.N., Krishnamurthy,S. and Hampsey,M. (2009) A physiological role for gene loops in yeast. *Genes Dev.* **23**, 2604–2609.

53

3.2 Histone and DNA methylation cycles are coupled by DNA oxidation to induce transcription by nuclear receptors

Abstract

Changes of histone methylation code, DNA hydroxymethylation and formation of chromatin loops are associated with transcription induction by nuclear hormones, but it is not known if these events are the consequence or the cause of transcription initiation. We have located genomic regions that synchronously recruit estrogen receptor (ERα) and enzymes that demethylate histone H3, (e.g. LSD1) and recognize oxidized G (e.g. OGG1). We have found in three prototypic estrogen-induced genes that the periodic recruitment of these enzymes and their products at the enhancers and polyA sites are functionally linked to generate localized DNA damage and bring to formation of chromatin loops that juxtapose the 5' and 3' ends of induced genes. DNA methyltransferase 3a (DNMT3a) recruits BER and NER enzymes at the ERE and Poly_A sites, favoring strand specific repair of oxo-dG and methylation of the complementary C. Loss of DNMT3a results in accumulation of G to T mutations at these sites. We suggest that coupling transcription with methylation and repair enzymes is an evolutionary strategy to reduce the mutational burden induced by DNA oxidation necessary for chromatin looping.

Introduction

Histone methylation-demethylation, targeted DNA damage and formation of chromatin loops that juxtapose physically distant DNA regions have been variously associated with transcription in general and, specifically, with transcription induction by nuclear hormones [1]. Studies of synchronized transcription, in the time window of the observation, have documented several oscillatory cycles of histone methylation-demethylation, DNA oxidation and mRNA accumulation [2]. The interplay between histone modification writers and readers on chromatin has been extensively investigated, yielding a long list of histone methyl transferases and demethylases that variously associate with transcription induction or repression [3]. However, although the detailed structure and activity of these proteins, it is not known if these enzymes modify the accessibility of RNA polymerase complex to critical chromatin-DNA sites or directly participate to the assembly of the initiation complex at the transcription start site.

In recent years, evidence of transcription associated with DNA damage has been reported in neurons [4] or in nuclear hormones sensitive cells [5,6]. DNA repair enzymes have been found associated with the transcription initiation complex in the absence of exogenous genotoxic stimuli [7] and the phosphorylated histone H2AX, a specific signature of DNA damage, is enriched at transcription start sites (TSS) [8]. High levels of transcription increase dramatically mutation rates in *S.cerevisiae* and this is tightly associated with BER and NER pathways [9,10,11].

However, the following questions still pend: 1. What are the cause and the nature of this DNA damage and how is it induced? 2. Is the DNA damage the consequence of, or does it drive the formation of the transcription initiation complex?

We have proposed that DNA damage at the TSS is caused by DNA oxidation. All the histone demethylation enzymes are mono or di-oxygenases, which generate formaldehyde and modify the local chromatin redox state [12]. We have reported that oxidation of DNA G residues was a consequence of histone demethylation by the mono-flavin oxidase LSD1 and was important for estrogen-induced transcription. In addiction not only G can be oxidized but also C can be methylated by DNA methyltransferases and oxidized by a new class of di-oxygenases protein, TET enzymes [13].

According to these informations, we have hypothesized that both histone and DNA methylation-demethylation [14] cycles are causally connected to DNA oxidation. Histone demethylation induced direct oxidation of Gs residues through ROS production, while the Cs oxidation is an effect of DNA demethylation. Both G and C oxidation is finalized to reduce transcription-generated supercoiling and to allow the formation of DNA loops that juxtapose the transcription start site, the enhancer and the 3' end of the gene [2].

Here, we have presented data on estrogen-induced genes that support this model. First, promoter array analysis have showed that oxidation, histone demethylation and DNA repair enzymes accumulate at the transcription start sites and estrogen responsive elements (ERE) upon hormone challenge. Second, on 3 prototypic estrogen-induced genes, we have described the oscillatory recruiting of histone methylation and demethylation enzymes with BER, NER enzymes and DNMT3a. Third, we have showed that these recruitment cycles are synchronized with the cycles of chromatin loops of the induced genes and that the inhibition of histone demethylation and/or BER or NER cycles prevents formation of loops and the induction of transcription. Finally, we have uncovered the role of DNA methyl transferase3a

(DNMT3a), which links BER and NER enzymes at the enhancer and polyA sites, favoring strand-specific repair of the GC dinucleotide at the 5' and 3' end of the chromatin loops.

Material and Method

Cells, treatment and transfections

Human breast cancer MCF-7 cells were grown at 37 °C in 5% CO_2 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). Cells were provided with fresh medium every 3 days. To evaluate the effect of estrogen challenge, cells were grown in phenol red-free DMEM containing 10% dextran–charcoal-stripped FBS for 1 to 3 days, before being challenged with 50 nM 17β-estradiol (E_2) for different times according to the experimental needs.

To obtain LSD1 dominant negative expression, the cells were transiently transfected, using a Neon® Transfection System, with LSD1 full-length cDNA and LSD1 dominat negative (ala) were inserted into the CMV 3xFLAG, Sigma-Aldrich, St. Louis, MO, USA, in medium without serum and incubation was continued for 48 h as previously published (Ambrosio et al., 2013).

Chromatin Immuno-Precipitation (ChIP)

The cells (~2.5 x 10^6 for each antibody) were fixed for 10 minutes at room temperature by adding 1 volume of 2% formaldehyde 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 10^6 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 (1 mg/ml). Precleared chromatin was divided in 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 FastStart Universal SYBR Green Master (Rox) (Roche Applied Science) with cycle conditions as follows: 95 °C 10 min; 40x (95 °C 10 sec, 55 °C 30 sec, 72 °C 30 sec); 72 °C 10 min.

Chromosome conformation capture (3C)

The 3C assay was performed as described previously (Dekker J. et al., 2002) with minor adaptations. Briefly: the AvaII and BamH1 restriction enzymes were used (Roche Applied Science). The cells (2.5×10^6) were crosslinked in 12 ml of PBS with 1% formaldehyde for 10 min at room temperature. The reaction was quenched by the addition of glycine to a final concentration of 125 mM. Fixed cells were harvested and the pellet resuspended in 1 ml of ice-cold lysis buffer (the same used for ChIP experiments). Nuclei were washed with 0.5 ml of restriction enzyme buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1mM Dithioerythritol, pH 7.5) at 37 °C, centrifuged and resuspended in 100 μl of restriction enzyme buffer. SDS was added to a final concentration of 0.1%, and nuclei were incubated at 37 °C for 15 min. Triton X-100 was added to the final concentration of 1% to sequester SDS. Digestion was performed with 100 U of the restriction enzyme at 37 °C for 16 h. The restriction enzyme was inactivated by the addition of SDS to 2% and incubation at 65 °C for 30 min. The reaction was diluted into 1 ml ligation reaction buffer (66 mM Tris-HCl, 5 mM MgCl2, 5 mM DTT, 1 mM ATP, pH 7.5) and incubated at 16 °C for 18 h with 50 U of T4 DNA Ligase (Roche Applied Science). EDTA (10 mM) was added to stop the reactions. Samples were treated with Proteinase K (200 μg/ml) and incubated for 5 h at 55 °C, and then overnight at 65 °C to reverse the formaldehyde crosslinks. The following day, the DNA was

purified by phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation. Samples were redissolved in 20 μl of TE buffer. To prepare a control template, we used a pool of plasmids containing an equimolar amount of the *PS2, BCL2 or CAV1* inserts spanning the genomic regions of interest. Five micrograms of plasmid DNA were digested with AvaII or BamH1 in 50 μl of 1x buffer for 8 h at 37 °C and then ligated in 20 μl with 5 U of T4 Ligase at 16 °C for 4 h. The efficiency of digestion at the end of 3C treatment was quantified by real time PCR, amplifying a fragment spanning two AvaII or BamH1 (uncut) in different 3C DNA preparations. Primer sequences are reported in Table 1. PCR were performed using FastStart Taq DNA Polymerase (Roche Applied Science) with cycle conditions as follows:

- BCL2 oligo 1-2/3-4 (Prenested) 95 °C 5 min; 5x (95 °C 45 sec, 57 °C 30 sec, 72 °C 40 sec) 25-44x (95 °C 45 sec, 55 °C 30 sec, 72 °C 30 sec); 72 °C 10 min;

- BCL2 oligo 1-2/3-4 (Nested) 95 °C 5 min; 5x (95 °C 45 sec, 57 °C 30 sec, 72 °C 30 sec) 30-40x (95 °C 45 sec, 54,5 °C 30 sec, 72 °C 30 sec); 72 °C 10 min;

- PS2 oligo 1-2 (Prenested) 95 °C 5 min; 5x (95 °C 45 sec, 58 °C 30 sec, 72 °C 40 sec) 35x (95 °C 45 sec, 56 °C 30 sec, 72 °C 30 sec); 72 °C 10 min;

- PS2 oligo 1-2 (Nested) 95 °C 5 min; 5x (95 °C 45 sec, 58 °C 30 sec, 72 °C 30 sec) 30-40x (95 °C 45 sec, 56 °C 30 sec, 72 °C 30 sec); 72 °C 10 min;

- CAV1 oligo 8-2 (Prenested): 95 °C 5 min; 5x (95 °C 45 sec, 60 °C 30 sec, 72 °C 30 sec); 25x (95 °C 45 sec, 58 °C 30 sec, 72 °C 30 sec); 72 °C 10 min;

- CAV1 oligo 8-2 (Nested): 95 °C 5 min; 5x (95 °C 45 sec, 62 °C 30 sec, 72 °C 35 sec); 35x (95 °C 45 sec, 60 °C 30 sec, 72 °C 35 sec); 72 °C 10 min;

PCR products were run on 1.2% agarose gels, stained with ethidium bromide and quantified with the imageJ program (Rasband WS, ImageJ, National Institutes of Health, Bethesda, Maryland, USA, http: //rsb.info.nih.gov/ij/). The amplified fragments at the end of the procedure were verified by DNA sequence analysis.

 Antibodies

 ERa ab32063 (Abcam); H3K4me2 ab32356 (Abcam); H3K4me3 ab1012 (Abcam); H3K9me2 ab1220 (Abcam); H3K9me3 ab8898 (Abcam); Total H3 ab1791 (Abcam); OGG1 sc-33181 (Santa Cruz Biotechnology); APE1 ab2717 (Abcam); LSD1 ab17721 (Abcam); JMJD2A sc-271210 (Santa Cruz Biotechnology); DNMT3a ab2850 (Abcam); TET1 GTX124207 (Genetex); TDG sc-292440 (Santa Cruz Biotechnology); XPG sc-13563 (Santa

Cruz Biotechnology); XPC sc-74411 (Santa Cruz Biotechnology); XPA ab85914 (Abcam); RPA ab2175 (Abcam); Normal rabbit IgG sc-2027 (Santa Cruz Biotechnology); Normal mouse IgG sc-2025 (Santa Cruz Biotechnology).

Ligation-dependent Probe Amplification (LPA)

Technique premise. The technique is a variant of Multiplex ligation-dependent probe amplification (MLPA) that has been developed to assess the presence of 8-oxoguanina base (8-oxoG). It involves the use two pairs of oligonucleotides each complementary to strand plus (+) and minus (-) of analyzed DNA regions. The junction point between the two oligonucleotides that recognize the same strand is located on a C (the 3'end of oligo L) and G (the 5'end of oligo R) of a CpG dinucleotide. The oxidation of both G located in the junction point produces a mismatch with used oligo pair, compromising the ligase reaction for both strands. To confirm that the mismatches is a consequence of 8oxoG we exploit the glycosylase and AP-lyase activity of FPG enzyme (8-oxoGuanine DNA Glycosylase from *E.coli*) that specifically recognizes the 8-oxoG, inducing a inefficiency ligation profile similar to the same samples not treated with FPG.

Experimental protocol. The LPA assay was performed as described previously (Jan P. Schouten et al., 2002) with minor adaptations. Briefly: 1μg of DNA samples were divided in 4 parts (each of them with 250 ng), two of them were digested with 1U of FpG enzyme (New England Biolabs) for 1 h at 37 °C and 5 min at 98 °C in a Buffer1 (10 mM Bis-Tris-Propane-Hcl, 10mM $MgCl_2$, 1mM DTT, pH 7.0) supplemented with BSA. The other two were used as not digest samples control to which were added the storage buffer of FpG (20 mM Tris HCl, 50 mM NaCl, 0.5 mM EDTA, 200 μg/ml and 50% glycerol).

Then, KCL 5x buffer (750 mM KCl, 50 mM Tris pH, 5 mM EDTA) and strand specific probes (12.5 fmol/sample) were added to digested and not digested samples and heated for 1 min at 95 °C and then incubated for 16 h at 60 °C for ERE, CnG, TATA and INTRON1 regions and 2 h at 60 °C for Poly_A region. For each analyzed region were outlined: a pair of primers complementary to the minus strand (-) (Right probe oligonucleotide (RPO(-))/(Left probe oligonucleotide (LPO(-)) and a pair of primers complementary to the plus strand (+) RPO (+)/ LPO (+)). The primers RPO (-) and RPO (+) were phosphorylated at t 5' end with 20U T4 kinase enzyme to 500 picomoles of oligo, in the presence of 100 nM ATP for 45 min at 37 ° C.

Ligation of annealed oligonucleotides was performed by diluting the samples to 60µl with dilution buffer (5 mM Tris HCl pH 8.5, 2.6 mM $MgCl_2$, 0.013% Triton x-100, 0.2 mM NAD) containing 1 U Ligase SALSA-65 enzyme (MRC Hollad), and incubation for 1h at 56 °C. The ligase enzyme was inactivated by heating at 98 °C for 15 min and ligation products were amplified by qPCR. For more experiments, 10 µl of the ligation reaction was added to 25 µl qPCR buffer. Real Time-qPCRs were performed using FastStart Universal SYBR Green Master (Rox) (Roche Applied Science) with cycle conditions as follows: 10 min at 95 °C; 35 cycles (30 s at 95 °C, 30 s at 60 °C and 40 s min at 72 °C); 10 min at 72 °C. DNA was amplified by PCR using primers listed in the Table 2.

Total genome analysis of 8-oxoG and 5hmC

MCF7 were pretreated for 2h with 2,5 µM of α-amanitin (AMA). After washing out the cells were stimulated with 50 nM E_2, or 10 nM 4-hydroxytamoxifen (4-OHT) or 10 µM parnate (Par) at different time (0, 15, 30, 45 and 60 min). DNA was extracted, clevaged with nuclease P1 and analyzed through spectrometry [15]. This experiment was done in collaboration with the laboratory of Prof. Ryszard Olinski (Department of Clinical Biochemistry, Nicolaus Copernicus University (Karlowicza)).

Revelation of hydroxymethylated citosine

The Oxidative Bisulfite assay was performed as described previously (Michael J Booth, et al. 2013) with a mirror adatpion. Briefly: 1µg of DNA samples were denatureted in NaOH 0.05 mM at 37 °C for 30 min and oxidated with 1µl of Potassium Perruthenate (KRuO4, Sigma) (15 µM) in ice for 1 h, vortexing every 15 min. As control, KRuO4 solvent was added to the samples. The oxidative and not oxidative samples were undergoes bisulfite conversion (EZ DNA Methylation-GoldTM Kit, Zymo Research) with cycle conditions as follows: 10 min at 98 °C, 4 h at 53 °C. DNA was amplified by PCR using primers listed in the Table 3 using HotStratTaq DNA polymerase (Qiagen) with cycle conditions as follows: 15 sec at 95 °C; 35 cycles (40 sec at 95 °C, 40 sec at 52 °C and 40 sec at 72 °C); 10 min at 72 °C. The samples were sequenced with Sanger method.

Statistical analysis

All data are presented as mean ± standard deviation in at least three experiments in triplicate (n≥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). All tests was performed using the JMP Statistical Discovery™ software by SAS, Statistical Analysis Software.

Results

Estrogen receptor clusters histone methylation-demethylation and repair enzymes at discrete genomic sites.

We first identified the genomic regions capable of recruiting, upon estrogen stimulation, estrogen receptor α (ERα), lysine specific de-methylase (LSD1) and 8-oxoguanine DNA glycosylase (OGG1). Immune-precipitated DNA fragments with specific antibodies to ERα, LSD1 and OGG1 from estrogen-stimulated cells (50 nM for 45 min) were used as probes to screen arrays of human promoters (Affymetrix, human promoter 1.0 R array). These arrays interrogate regions proximal 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 upstream through 2.45 kb downstream of 5' transcription start sites.

We focused our attention on the sites that synchronously recruit the estrogen receptor α and the enzymes indicated above upon E_2 stimulation. We mapped these regions in each chromosome but for simplicity we have shown in the figure 1A the percentage of ERα interaction with LSD1 and OGG1 in mono, bivalent and trivalent complexes on different genomic regions of chromosome 18. After E_2 stimulation, the co-recruitment of ERα, LSD1 and OGG1 increases on regulative regions. On estrogen responsive genes CAV1, BCL2 and PS2 (poised respectively on chromosome 7, 18 and 21) the stimuli makes the linkage of above mentioned factors happen on the same regions (Figure 1B).

Intertwined Cycles of histone H3 methylation and recruitment of DNA repair enzymes at ERE and polyA sites of E_2-induced genes.

Estrogens induce the transcription of several genes. Some genes are induced de novo, others are already transcriptionally active and the hormone super-induce them. We determined the histone code changes induced by the hormone in both types of genes. To do that we analyzed

the H3 histone methylation code of the lysine 9 and 4 on ERE and Poly_A chromatin regions of pS2 (induced by E_2), CAV1 and BCL2 (super-induced by E_2) genes. Figure 2 presents a summary of the most of chromatin changes and the recruitment of ERα, BER, NER and DNA methylating enzymes on PS2 gene. Receptor upon estrogen stimulation is highly and stably recruited to ERE and Poly_A at 1-2 hours (data not shown); in short kinetics (0-60 min E_2) this recruitment is not stable in fact ERα link rapidly ERE at 30-45 min and decrease at 60 min. The binding of the receptor oscillates at the 3' end Poly_A site during E_2 stimulation in the same way of ERE, suggesting that the 2 regions are synchronized or probably interact (Fig. 2A).

We observed the same kinetics on CAV1 (Fig. S1A) and BCL2 (Fig. S2A) regulative regions. In pS2 ERE and Poly_A regions, we found that estrogen induces cycle of methylation and demethylation. The stimuli leads to a peak of demethylation of the H3K9me3/me2 at 30 and 60 min and methylation at 45 min on analyzed regions. At the ERE region, H3K4me2 increases at 15 min and remains high up to 45 min, whereas H3K4m3 increases at 45 min. At the Poly_A region we observe an increase of H3K4me3 at 15 and 45 min, whereas the H3K4me2 increases only at 45 min (Fig. 1B).

After estrogen exposure on CAV1 ERE and Poly_A regions, we found that H3K9me2 was methylated at 30 min on ERE and at 45 min on Poly_A while H3K9me3 did not show remarkable oscillations. H3K4me3-me2 increase at 15 min and 45 min and decrease at 30 min (Fig. S1B).

In BCL2 ERE and Poly_A regions only H3K4m2 and H3K9m3 revealed oscillatory kinetics. In particular H3K4m2 had an unique peak at 30 min on both regions, whereas H3K9m3 increase at 15 min, decrease at 30 min and are re-methylated at 45 min (Fig. S2B).

Collectively, all the regions ERE and Poly_A sites of the estrogen-induced genes analyzed, show discrete oscillations of methylation and demethylation involving H3K9 and H3K4. In addition, we noted that the location of the ERE (intron or exon) and the basal transcriptional state of the gene may modify the intensity of the methylation- demethylation changes.

These dynamic methylation changes of K4 and K9 induced by E_2 exposure suggest that on these regions de-methylation enzymes were recruited; in particular on ERE region JMJD2A demethylates H3K9me3 (15 min E_2) and LSD1 demethylates H3K9me2 and H3K4me2 (15-45 min E_2), while the kinetic of these enzymes recruitment upon E_2 stimulation do not presents remarkable fluctuations on other one regulative region except of a peak of JMJD2A for CAV1 and BCL2 genes at 45 min. (Fig. 2C, Fig. S1C, Fig S2C).

Periodic recruitment of BER, NER enzymes at ERE and polyA sites. Transcription is associated with a burst of DNA oxidation dependent from histone demethylation [16] in particular the activation of a Fe^{++} dioxygenase (JMJD2A) and a FAD oxidase (LSD1) triggers local oxidation. Oxidized guanine (8-oxodG) is recognized by OGG1 [17]. In addition, it was recently reported that NER enzymes are essential for the formation of chromatin loops and DNA de-methylation induced by retinoic acid at the target genes [18]. These data suggest that both BER and NER enzymes participate to the formation of chromatin loops induced by nuclear hormones and are essential for the correct repair of DNA oxidation lesions. Therefore, we analyzed the recruitment of BER enzymes, XPC, XPG, XPA and RPA (NER) at the ERE and PolyA chromatin regions after E_2 induction. ChIP analysis shows that OGG1 is recruited very early (15' E_2) on ERE and Poly_A of studied genes, but on CAV1 and BCL2 show a second peak at 45 min (Fig. 2D, Fig. S1D, Fig. S2D). APE1, which process the abasic site (s) generated by OGG1, is recruited at 15 and 45 min on ERE of BCL2 and at 30 min on CAV1; the kinetics of recruitment is the same (45 min E_2) on Poly_A of all gene analyzed (Fig. 2D, Fig. S1D, Fig. S2D). Others analyzed enzymes do not show particular oscillations, except RPA that display a peak on Poly_A of CAV1 between 15 and 30 min, and TDG (Tymidine DNA Glycosilase which is required for base excision repair of deaminated methylcytosine, a frequent product of base oxidation) that bind the same region of CAV1 and BCL2 after 15 min of E_2 challenge (Fig. 2E, Fig. S1E, Fig. S2E).

In addition, we noticed important and selective differences in the recruitment of XPC and XPG: XPC is recruited earlier at the ERE than XPG (15 min versus 45 min), while at the Poly_A both XPC and XPG are recruited with the same timing (Fig. 2E, Fig. S1E, Fig. S2E).

These data suggest a hierarchy in the recruitment of BER and NER and between different types of NER enzymes. In fact, OGG1 is recruited early with XPC at the ERE this information is relevant, because XPC is part of the Global Genome NER that recognizes lesions caused by the distortion of the double helix, suggesting that it recognized the oxidative or apurinic site independently of the presence of stalled PolII.

Periodic recruitment of DNMT3a and TET1 enzymes at ERE and Poly_A sites. The gene transcriptional activation seems to be associated to DNA methylation/hydroxymethylation cycles, indispensable for a faithful transcriptional process. DNA oxidation involves not only G but also C. In fact C can be oxidized only when methylated in vivo. It has been reported

that estrogens induce cycles of the CpG methylation within the PS2 promoter through recruitment of DNMTs enzymes [14].

In this process a crucial role seems to be played by the DNA methyltransferase 3a (DNMT3a) which acts as a cofactor for OGG1, working closely with it in the repair of 8-oxo-guanine, through the methylation of cytosine which is opposed to the oxidized guanine; also TET1 plays a key role in repair because hydroxylating methylated cytosine by DNMT3a does not allow the tautomerization of C and allows the correct replacement of the oxidized G. We analyzed the recruitment of DNMT3a and TET1 at the ERE and Poly_A chromatin regions after E_2 induction. The results obtained reveal that on ERE region DNMT3a and TET1 are recruited after 15 min of stimulation, and they recruitment remain more or less constant up to 60 min E_2. On Poly_A the levels of both enzymes increased to 15 min E_2 and in the times of stimulation higher have slight fluctuations with a pick of DNMT3a at 30 min from the stimulation. We observed the same oscillations on all genes studied (Fig. 2F, Fig. S1F, Fig. S2F).

Estrogen induces strand specific oxidation and repair: coordinated repair of 8-oxodG and 5hmC.

Repair of 8-oxodG. Repair enzymes have a crucial role in transcriptional activation, in fact their silencing reduce significantly specific mRNA production and increase basal transcription (data not shown).

Among the events that lead to their recruitment there is the Gs oxidative burst on promoter region, this process is mediated by monoammino-oxidase activity of LSD1 that produces reactive oxygen species (ROS) that oxidizes G to 8-oxo-deoxyguanosine (8-oxoG). In first analysis we assessed the guanine oxidation induced by transcriptional activation with total genome analysis, evaluating the 8oxo-G accumulation in short kinetic after estrogens stimulation, after treatment with α-amanitin (transcription inhibitor), parnate (monoammino-oxidase inhibitor) and 4-hydroxytamoxifen (estrogen inhibitor) (Fig. 3). The results obtained underline that the oxidation occurs after 15 min of hormonal stimulation and that it is directly linked to transcription, chromatin remodeling depending to LSD1 activity and estrogen stimulation, in fact the treatment with specific inhibitors breaks down completely the phenomenon (Fig. 3). To directly document the DNA oxidative lesions and to monitor site-specific repair, we mapped precisely oxodGs using Ligation Proximity Amplification (LPA)

technique (see Material and Method). First of all we evaluated the technique sensitivity and specificity and sensitivity using it to amplify synthesized strands with 10% or 50% of modified bases such as 8-oxoG or formilC (deoxyformilcytosine). The method is sensible because detect molecules that contain also only the 10% of 8-oxoG, and specific because it is not influenced by other modified bases such as formilC. The treatment with FPG underlines that the efficiency reduction is due specifically to the presence of 8-oxoG (Fig. 4)

After we analyzed the kinetic of oxidation/repair of two strands of different regions of PS2 gene in estrogen kinetic stimulation. The data express the efficiency variation on basal conditions in which a lower efficiency corresponds to a higher oxidation degree and vice versa.

On ERE region of PS2 the damage repairing kinetics is asynchronous for stand plus (+) and strand minus (-); in fact both stands show oxidation at 5min but on strand (-) the first oxidation peak occurs to 15 min E_2 when the strand (+) is completely repaired. Also the amplification peak due to the complete repair of the strand (-) manifests at 30 min when the strand (+) is oxidized. The second oxidation cycle kinetic is similar on two strands probably the cells have lost the transcriptional synchronization (Fig. 5A).

To assess whether the oxidation was a peculiarity of CG dinucleotide we have analyzed a CnG region 24 bases upstream to ERE region (so we investigated if the penultimate base mismatch influences ligase SALSA 65 activity). The obtained data show an asynchronous oxidation/repair on two strands. The strand plus (+), after initial oxidation, which extends till to 15 min, shows the repairing peak (maximum efficiency of LPA) at 30 min, while the strand (-) is already repaired after 15 minutes from the stimulus. The kinetics of repair in this region is inverted instead of ERE region because repair of damage can not take place on two adjacent regions at the same time (Fig. 5B).

In TATA region analysis is important to consider that it presents the start of transcription, so the strand minus (3'→5') is the template for mRNA synthesis. In this region the strand (+) is more damaged in basal condition than strand minus but both strands undergo a gradual repair that extends till to 30 min for strand plus and till to 15 min for strand (-). The two strands keep on asynchronism of oxidation/repair but the strand (-) is the less damaged because it is transcribed (Fig. 5C).

On Poly_A region was interesting to observe how the strand (+) and (-) have a kinetic oxidation/repair analogous to ERE and TATA regions of same gene, suggesting that these regions could be physically close (Fig. 5D).

In addition to these regulatory regions it was analyzed an intronic region in which there is not observed a particular variation of the kinetics of oxidation and repair, in fact both strands have asynchronous low level of oxidation and repair. This underlines that the oxidative wave does not extend to that region (Fig. 5E).

Repair of 5hmC. Estrogen induces recruitment of DNMT3a and TET1 to regulate the wave of DNA methylation and oxidation. This coordinated recruitment of enzyme that methylate (DNMT3a) and hydroxymethylate (TET1) of cytosine may be crucial for repair. In addition, the presence of the TET1 oxidation products (5-hmC and 5-formylC) marks active transcription [19]. In first analysis we have assessed the cytosine oxidation induced by transcriptional activation with total genome analysis, evaluating the 5-OHdmC accumulation under the same conditions described above. The results do not show large fluctuations of oxidized Cs after estrogen induction, doing deduce that oxidized Cs are probably localized and unstable, so it is produced and repaired immediately (Fig. 6).

To determine if E_2-induced G and C oxidations are coordinated, we performed a modified bisulfite assay on ERE region of TFF1 to determine methylation and hydroxylation of specific Cs. The use of potassium perruthenate (KRuO4) selectively oxidizes 5hmC into 5- formyl-cytosine (5fC), which after bisulfite conversion is read as uracil in the sequence. We analyzed by oxBS-sequencing [20] five CpG on ERE region (Fig. 7). Figure 7 shows that the five CpGs upon E_2 challenge are differentially methylated in a strand specific fashion with a peak at 45' E_2, corresponding to the peak of histone H3K9 methylation (Fig. 2B), and they have a different levels of C hydroxymethylation. In particular the C -400 on strand (+) is methylated during all time of estrogen stimulation, while it has a peak of C hydroxyl-methylation at 15 and 30 min on strand (-) and then returns to baseline at 45-60 min (Fig. 7A). The C -388 has a low level of methylation/ hydroxyl-methylation on both strands it is oxidized with the same timing on the (+) and (-) strands, respectively (Fig. 7B). The C -370 on strand (+) is methylated for all the time of stimulation and formilated at 45 min E_2, while it is hydroxyl-methylated at 45 min and completely formilated at 60 min on strand (-) (Fig. 7C). The C -354 is formilated 30 min to 60 min on strand plus ad present as 5-OHdmC at 45 min E_2 on

the other strand (Fig. 7D), at last the C -301 is low formilated and hydroxyl-methylated respectively at 45 min and 15 min from stimulation (Fig. 7E).

We conclude that estrogens induce a sharp C hydroxyl-methylation peak at specific Cs at 15-30' E_2 in a strand-specific manner. These data indicate a precise order in timing and space of repair induced by E_2 and show that both oxidized C and G are precisely marked. We believe that methylation of C marks the abasic site and favors an efficient and precise repair.

DNA oxidation drives the formation of chromatin loops induced by estrogens.

Our data indicate that estrogens induce the same histone modifications in distal segments of target genes, suggesting that these regions are functionally synchronized and may be physically associated in an unique estrogen induced domain. To positively identify these domains, we performed the 3C analysis of three E_2-induced genes indicated, on chromatin isolated from E_2-induced cells. Briefly, fixed chromatin DNA was cleaved with a restriction enzyme and ligated under very diluted conditions to favor intra-molecular interactions. PCR and DNA sequence were then used to identify the ligated DNA segments.

The results obtained have shown that chromatin loops bridging the 5'-ERE and polyA sites are cyclically induced by E_2 at 15 min and 60 min. After stimulation BCL2 forms two loops: **1.** A loop connecting the ERE region and Poly_A (oligo 1-2), **2.** A loop connecting TSS and Poly_A (oligo 3-4). The first loop do not depend on estrogen stimulation in fact it is present in basal condition and it is formed with a cyclicality of 15 min then stabilized at 240 min when estrogen transcription ends. Instead the second loop is induced by the stimulus, it appears 15 min following E_2 exposure, disappear 30 min to 45 min and reform by 60 min (Fig. 8). Since ERE and promoter are contiguous in PS2 gene, we detected essentially one major loop connecting the 5' (promoter-ERE) with the 3' end of the gene (Poly_A region). This loop appear at 0 min, it becomes weak between 15 and 30 min and stabilizes between 45 and 60min from the stimulation (Fig. 9). Estrogen induced chromatin interactions in CAV1 gene at 15, 45 and 60 min, these connections involve ERE1 and Poly_A2 region (oligo 8-2) (Fig. 10). Data in the literature have shown that the transcriptional factory and looping gene is stabilized by RNA molecules that hybridize with the DNA, in fact the treatment with RNasiH destroys loop induced by estrogen in PS2 (Fig. 9B). We have also investigated if the loop formation depend to LSD1 histone demethylation activity. To demonstrate that LSD1 is

involved in loop formation, we transfected the cells with a dominant negative LSD1ala. This mutant is still enzymatically active, but is unable to target transcription factors; the LSD1ala mutant protein is difective in binding to the promoter or ERE element of studied genes following E_2 induction. The expression of this LSD1 variant inhibits E_2-induced chromatin loop altering the looping kinetics of analyzed genes (Fig. 8B, Fig. 9B, Fig. 10B).

Discussion

Transcription induced by estrogen is associated with chromatin modification dependent from histone methylation-demethylation cycles. Our data show that estrogens induce a change of the histone code on target genes. (Fig. 2B, S1B, S2B). Both lysines 9 and 4 methylation are stabilized after the first hour of induced-transcription, when the transcription become productive and this change can be depend to the location of the enhancer (intra or inter-gene) and the status of the basal transcription (see BCL2 and CAV1 genes). LSD1 and JMJD2A are the major histone demethylases responsible for histone modification induced by estrogen. We find a different timing of the recruitment of LSD1 and JMJD2A at the gene locations examined (Fig 2C, S1C, S2C). In transcriptional activation fundamental is the role of LSD1 histone demethylase that generates a DNA oxidative burst [16] and consequently a DNA repair enzymes recruitment: BER and NER enzymes (Fig. 2D/2E, S1D/E, S2D/E). The initial lesions are recognized by OGG1 and APE1, which display the same oscillations of demethylation-methylation cycles (15 and 45-60). Following OGG1, NER are recruited at 30 min. Interestingly, XPC is recruited earlier at the ERE than XPG (15 min versus 45 min), while at the Poly_A both XPC and XPG are recruited with the same timing (Fig. 2E, S1E, S2E). XPC is part of the Global Genomic NER, which recognizes helix distortions regardless of the presence of Pol II. On the other hand, XPG is part of Transcription-Coupled NER, which assists the RNA Pol II during recognition of the lesion [21]. These data suggest that the ERE and Poly_A lesions are differently recognized and repaired. OGG1 recognizes the 8-oxo Gs detecting a distortion of the helix. This can represent a signal for recruitment of XPC (GG-NER) at enhancer and can determine a stall of RNA Pol II at the Poly_A, which recruits XPG (TC-NER). The results obtained underline that the estrogen oxidative damage on regulative regions of estrogen-dependent genes is responsible of the interconnection between repair and transcription start systems. The recruitment of LSD1, in fact, acts as a pivot between transcription and repair and it simultaneously allows both the opening chromatin, then the

access to the promoter regions by the transcriptional machinery, that the production of ROS, responsible of promoter oxidative wave. Through a total genome analysis we have seen a transcriptional 8oxo-G accumulation (15 min) (Fig. 3); after we have analyzed the kinetic of oxidation/repair of specific Gs on two strands of different regions of PS2 gene to do that we used Ligation Proximity Amplification (LPA) technique. The data show that estrogens induce oxidation and repair differentially at the (-) and (+) strands of regulative regions (Fig. 5). In fact, in the transcription direction is repaired before the strand plus (TATA region) because it is not the template for mRNA synthesis and so it's free to protein factors; in not codogenic direction is repaired before the strand minus (CnG region). The asynchronism and specificity of strand oxidation can be explained by the need of the cell not to induce double strand break to each transcriptional cycle and to prevent a steric hindrance between the transcriptional machinery and repair enzymes. Our data show that DNMT3a is recruited at the enhancer and Poly_A sites at 15-30 min (Fig 2F, S1F, S2F), and it is essential for productive transcription induced by estrogens and for the correct repair of oxidize G, in fact DNMT3A protect the C point in front to 8-oxoG promoting the correct repair of damage base. The changes in methylation and hydroxymethylation of CpG are also associated with the recruitment of TET1 and TDG at enhancer and Poly_A sites. TET1 recognizes and modifies the methylcytosine, while TDG recognizes and removes a variety of oxydized and deaminated cytosines. TET1 is recruited 15–30 min to the enhancer and 60 min after estrogen stimulation to the Poly_A. TDG is recruited to the Poly_A and enhancer of BCL2 but not to CAV1 and pS2 enhancers (Fig 2D/2F, S1D/1F, S2D/2F).

We analyzed the methylation and hydroxylation status of DNA at single strand resolution. Bisulfite analysis shows that template and non template strands have a different levels and timing of methylation/hydroxymethylation, the changes of hydroxymethylation are associated with the timing and the location of oxidized Gs, suggesting that the 2 events (G and C methylation-oxidation) are tightly associated (Fig. 7). The local DNA oxidation following demethylation by LSD1 releases supercoiling and rigidity of the helix, so it allows the physical connections between distant intergenic regions, the chromatin loops with discrete 5' and 3' borders and transcriptional activation. These loops depend on LSD1 action in fact the expression of dominant negative (LSD1ala) alters the kinetics of gene looping and they are stabilized by mRNA that hybridized with DNA in effect the treatment with RNaseH destroys its interactions (Fig. 8, Fig. 9, Fig. 10).

References

1. Beato M., Wright R.H., Vicent G.P. *DNA damage and gene transcription: accident or necessity? (2015)*. Cell Research, 25:769-770.

2. Zuchegna C., Aceto F., Bertoni A., Romano A., Perillo B., Laccetti P., Gottesman M.E., Avvedimento E.V., Porcellini A. *Mechanism of retinoic acid- induced transcription: histone code, DNA oxidation and formation of chromatin loops.* (2014). Nucleic Acids Res, 42: 11040-11055.

3. Yi X., Jiang X., Li X., Ding-Sheng Jiang D.S. *Histone methyltransferases: novel targets for tumor and developmental defects.* (2015) Am J Transl Res 7(11):2159-2175.

4. Su Y., Ming G.L., Song H. *DNA damage and repair regulate neuronal gene expression.*(2015) Cell Res., 25:993-994.

5. Perillo B. , Di Santi A., Cernera G:, Ombra M.N., Castoria G., Migliaccio A. *Nuclear receptor-induced transcription is driven by spatially and timely restricted waves of ROS. The role of Akt, IKKα, and DNA damage repair enzymes* (2014). Nucleus, 5(5): 482–491.

6. Williamson L.M., Lees-Miller S.P. *Estrogen receptor α-mediated transcription induces cell cycle-dependent DNA double-strand breaks.*(2011) Carcinogenesis, 32(3):279-85.

7. Seo J., Kim K., Chang D.Y., Kang H.B., Shin E.C., Kwon J., Choi J.K. *Genome-wide reorganization of histone H2AX toward particular fragile sites on cell activation.* (2014) Nucleic Acids Res.; 42(2): 1016–1025.

8. Kim N., Abdulovic A.L., Gealy R., Lippert M.J., Jinks-Robertson S. *Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication.* (2007) DNA Repair (Amst), 6(9): 1285–1296.

9. Kim N. and Jinks-Robertson S. *Abasic Sites in the Transcribed Strand of Yeast DNA Are Removed by Transcription-Coupled Nucleotide Excision Repair.* (2010) Molecular and cellular biology, 30(13): 3206–3215

10. Beato M., Wright R.H., Vicent G.P. *DNA damage and gene transcription: accident or necessity?* (2015) Cell Res 25(7):769-70.

11. Dong C., Zhang H., Xu C., Arrowsmith C.H., Min J. *Structure and function of dioxygenases in histone demethylation and DNA/RNA demethylation.* (2014) IUCrJ 1, 540–549.

12. Perillo B, Ombra M.N., Bertoni A, Cuozzo C., Sacchetti S., Sasso A., Chiariotti L., Malorni A., Abbondanza C., Avvedimento E.V. *DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression.*(2008) Science 319(5860):202-6.

13. Müller U., Bauer C., Siegl M., Rottach A. , Leonhardt H. *TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation.* (2014) Nucleic Acids Res. 42(13):8592-604.

14. Métivier R., Gallais R., Tiffoche C., Le Péron C., Jurkowska R.Z., Carmouche R.P., 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, 452(7183):45-50.

15. Gackowski D., Starczak M., Zarakowska E., Modrzejewska M., Szpila A., Banaszkiewicz Z., Olinski R. *Accurate, Direct, and High-Throughput Analyses of a Broad Spectrum of Endogenously Generated DNA Base Modifications with Isotope-Dilution Two-Dimensional Ultraperformance Liquid Chromatography with Tandem Mass Spectrometry: Possible Clinical Implication*. (2016). Anal Chem, 20;88(24):12128-12136.

16. Perillo B., Ombra M.N., Bertoni A., Cuozzo C., Sacchetti S., Sasso A., Chiariotti L., Malorni A., Abbondanza C., Avvedimento E.V. *DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression.* (2008). Science 319(5860): 202-206.

17. Amente S., Bertoni A., Morano A., Lania L., Avvedimento E.V., Majello B. *LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc-induced transcription.* (2010). Oncogene 29 (25): 3691-3702.

18. Le May N., Fradin D., Iltis I., Bougneres P. and Egly J.M. *XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes.* (2012). Mol Cell. 47, 622-632.

19. Neri F., Incarnato D., Krepelova A., Rapelli S., Anselmi F., Parlato C., Medana C., Dal Bello F., Oliviero S. *Single-Base Resolution Analysis of 5- Formyl and 5- Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics.*(2015). Cell Rep. pii: S2211-1247(15)00009-1.

20. Booth M.J., Beraldi D., Bell N.M., Branco M.R., Reik W., Balasubramanian S. *25 Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine.* (2013) Nature Protoc,8(10):1841-51.

21. Gros L., Saparbaev M. K. & Laval J. *Enzymology of the repair of free radicals-induced DNA damage. (2002). Oncogene 21, 8905-8925.*

Figure 1
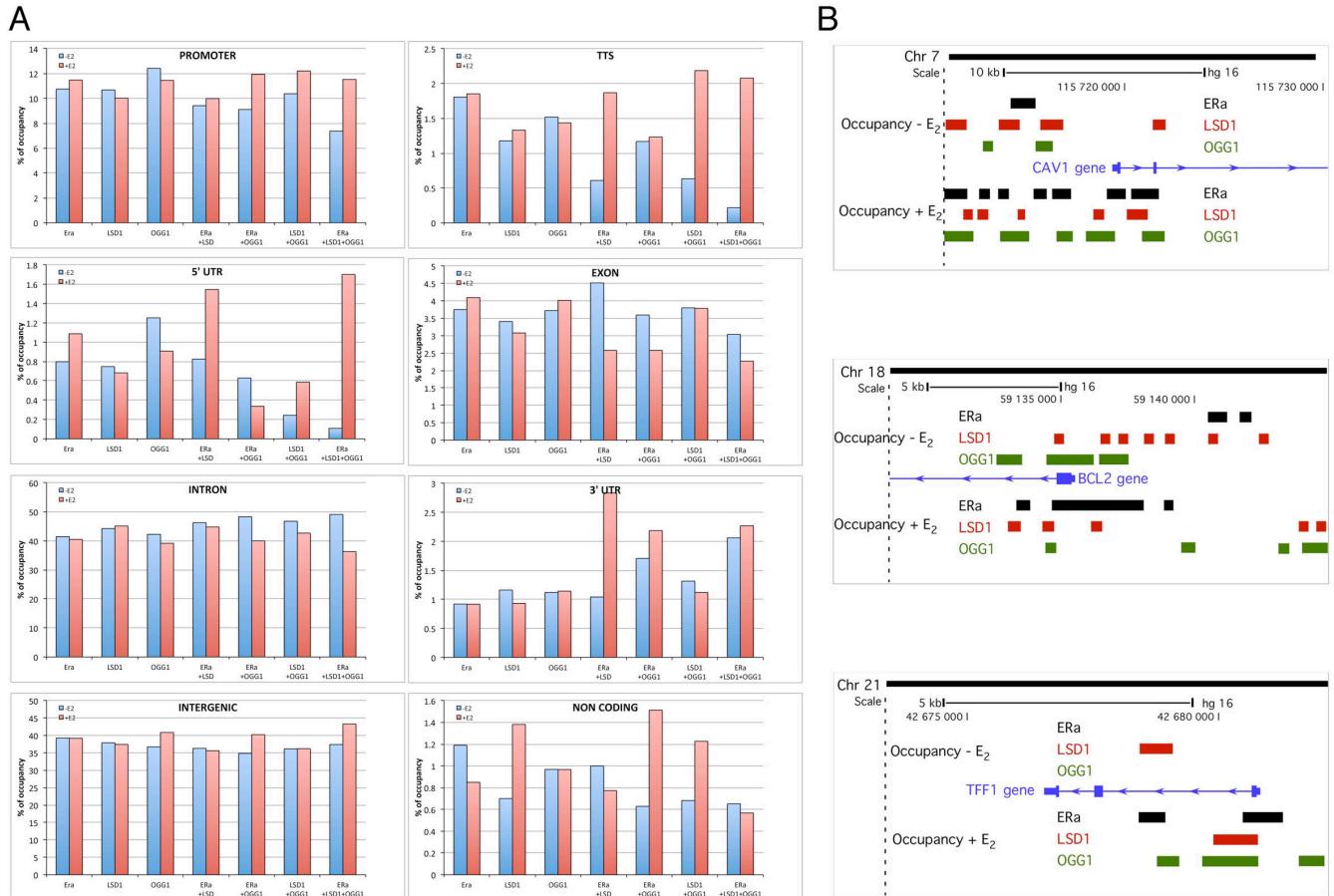
Estrogen receptor α synchronizes the recruitment of histone demethylase (LSD1) and DNA repair enzyme (OGG1) at multiple sites across the genome. Screening human promoter array for ERα, LSD1 and OGG1 recruitment sites induced by estrogen. MCF7 were grown in 10% of dextran-coated charcoal stripped serum for 3h and stimulated with 50nM of E_2 for 45 minutes. Formaldehyde-fixed chromatin was immunoprecipitated with antibody anti ERα, LSD1 or OGG1 and was resolved by GeneChIP Human promoter Array 1.0 (Affymetrix). Panel **A** shows the percentage of ERα interaction with LSD1 and OGG1 in mono, bivalent and trivalent complexes, on different genomic regions in presence (histogram red) or absence (histogram blue) of estrogen stimulation. Panel **B** shows occupation sites of ERα, LSD1 and OGG1 on CAV1, BCL2 and PS2 with or without E_2 stimulation. The black boxes represent ERα binding sites, red ones LSD1binding sites and green ones OGG1 binding sites 3'UTR-5'UTR orientation. Estrogen stimulation lead to an increase of ERα, LSD1 and OGG1 co-occupancy on gene regulative regions.

Figure 2
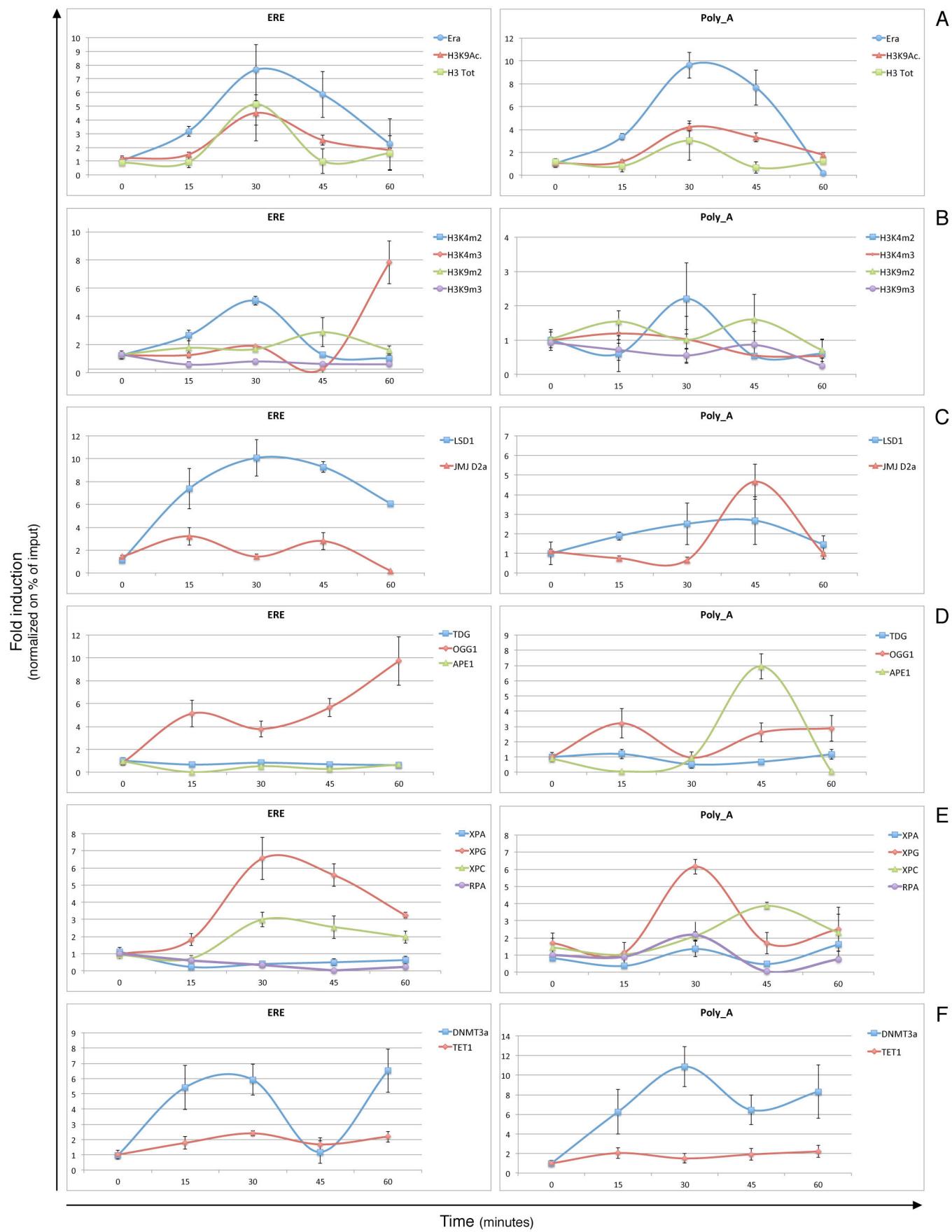
Intertwined Cycles of histone H3 methylation and recruitment of DNA repair enzymes at ERE and polyA sites of PS2 gene. MCF7 cells were serum starved and exposed to 50 nM E$_2$ at the indicated times (0, 15, 30, 45 and 60 minutes). **A-** qChIP was carried out using specific antibodies recognized estrogen receptor alpha (ERα), H3K9Ac and total H3 on the estrogen responsive elements (ERE) and polyA of *PS2* gene. Estrogen stimulation induces histone code changes; panel **B** shows H3K4m2/m3 and H3K9m2/m3 occupancy on PS2 ERE and Poly_A; The histone methylation and demethylation changes depend on the recruitment of histone demehylases on gene regulative regions. Panel **C** shows the recruitment of LSD1 and JMJD2A on above-mentioned regions. Histone methylation cycles overlap with the recruitment of BER, NER, DNMT3a and TET1. **D-** MCF7 cells were analyzed by qChIP using specific antibodies recognizing the OGG1, APE1 and TDG on ERE and Poly_A of PS2. Panel **E** shows the recruitment of NER enzymes (XPA, XPC, XPG and RPA) on studied regions; panel **F** shows the recruitment of DNMT3a and TET1 on the same regions of PS2. The results were normalized on estrogen unstimulated sample and on the percent of input. The statistical analysis derived from at least 3 experiments in triplicate (n ≥9; Mean±SD).

Figure 3
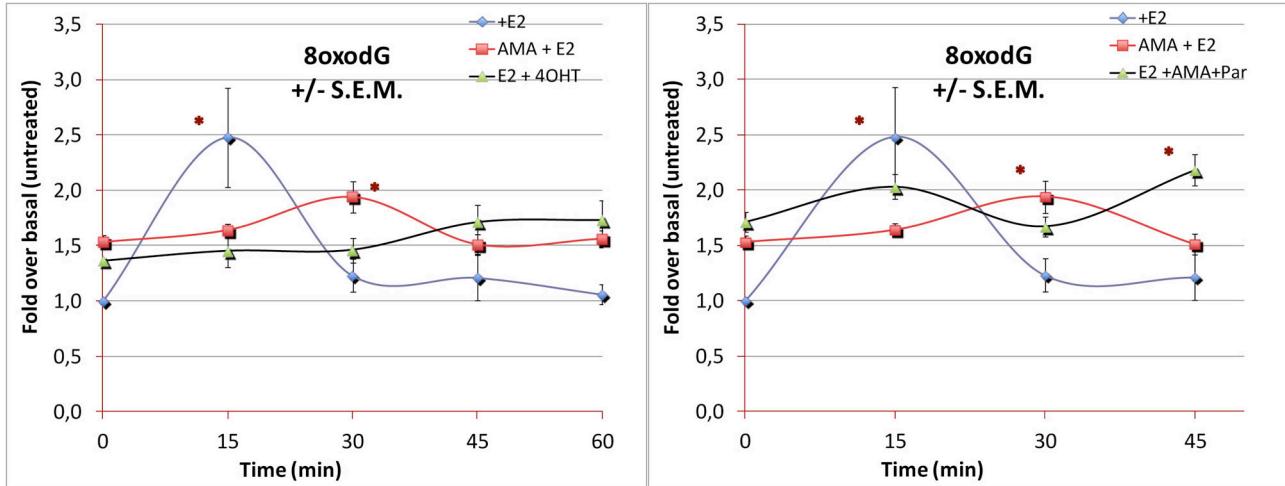
Total genome analysis of 8-oxo-deoxyguanosine (8-oxoG). DNA extracted from MCF7 stimulated with 50 nM of E_2 alone (blue lines) or in combination with 2,5 µM α-amanitin (AMA) (red line) or 10 nM 4-hydroxytamoxifen (4-OHT) (green line in panel on the left) or 10 µM parnate (Par) (green line in panel on the left) at different time (0, 15, 30, 45 and 60 min) was analyzed through spectrometry to evaluate the 8oxo-G accumulation in genome. The results shown are normalized on E_2 unstimulated sample. S.E.M was employed as statistical analysis.

Figure 4
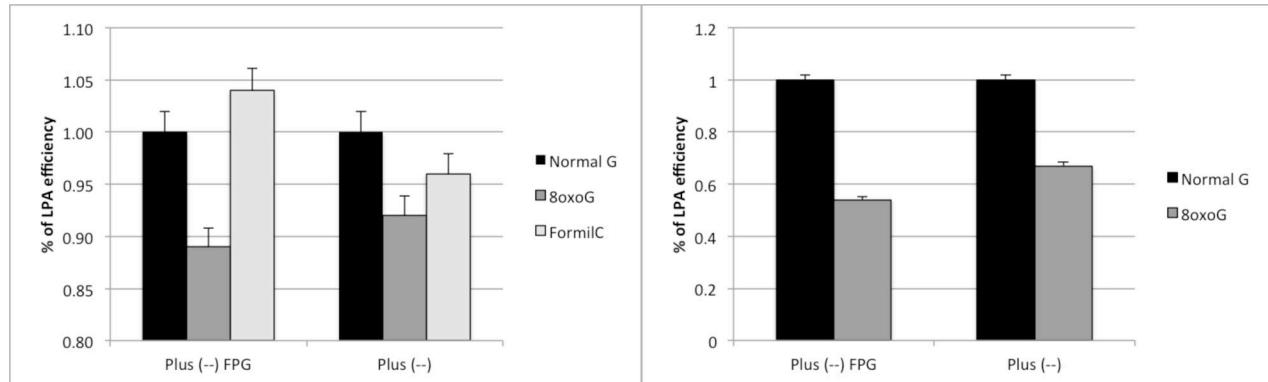
LPA assay control. DNA extrated by unstimulated MCF7 was used to syntetized DNA strands with normal G and with 10% and 50% of modified bases such as 8-oxoG and formilC (deoxyformilcytosine) in presence and absence of FPG (8-oxoGuanine DNA Glycosylase from E.coli) the treatment with FPG generates apurinic sites, where 8-oxoG was present. The generated nick do not support ligation and PCR of contiguous primers. The figure shows the results of LPA used to amplyfied strands with modyfied bases. Panel **A** shows amplification efficiency of plus strand syntetized with normal G and 10% of 8-oxoG and formilC with or without FPG treatment; panel **B** refers to amplification efficiency of plus strand syntetized with normal G and 50% of 8-oxoG with or without FPG treatment. The statistical analysis derived from at least 3 experiments in triplicate (n ≥9; Mean±SD).

Figure 5
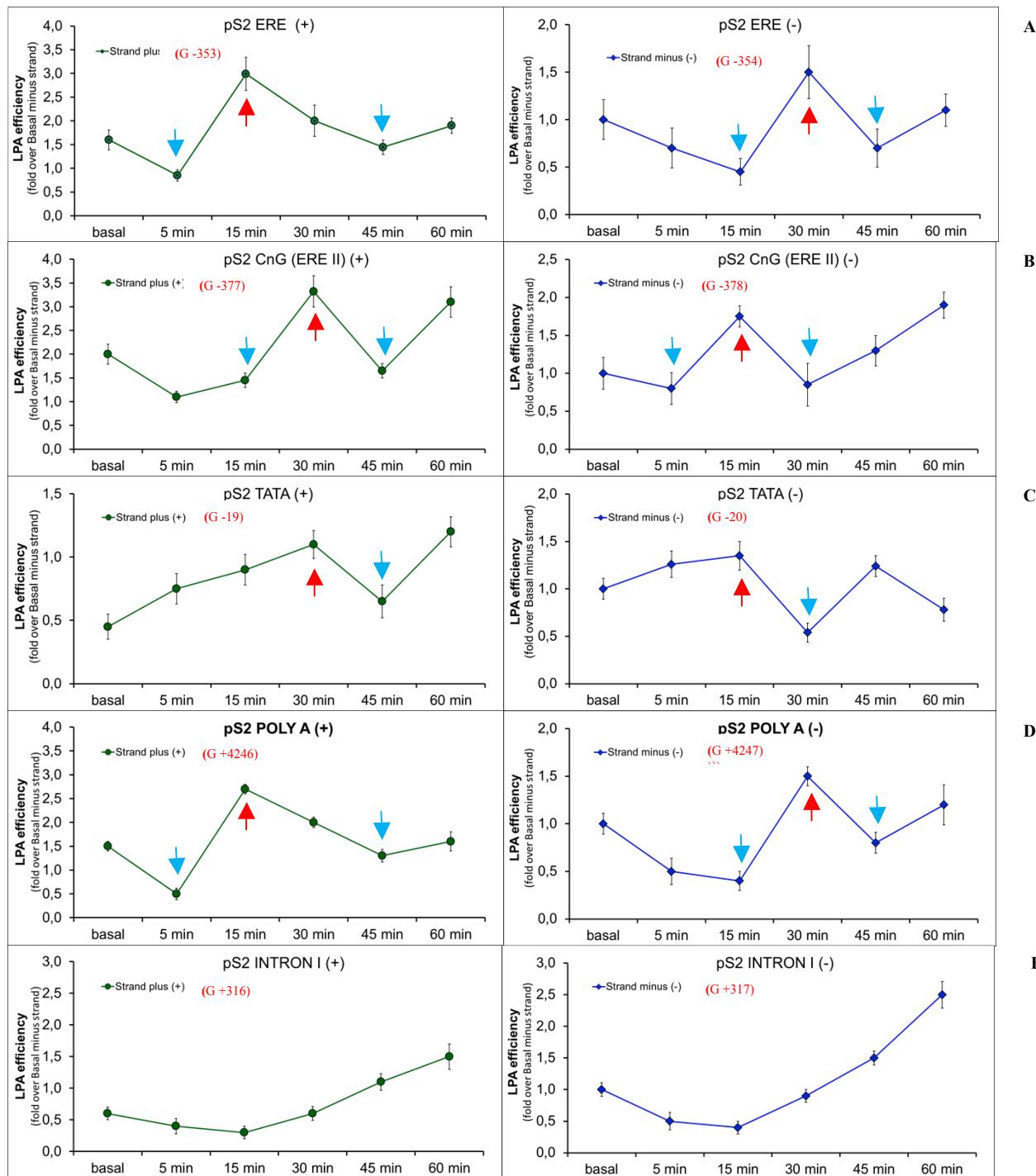
Estrogen induces strand specific oxidation and repair of 8-oxodG. MCF7 cells were serum starved and exposed to 50 nM E_2 at the indicated times (0, 5, 15, 30, 45 and 60 minutes). Resulting DNA was used to analized Gs (8-oxoG) oxidation level on different regulative regions of PS2 through Ligation-dependent Probe Amplification (LPA) method. Data is expressed as relative efficiency of minus strand; all data is normalized on PCR efficiency (not LPA standard) obtained on same DNA using primers couple for the same analyzed region. Panels A, B, C, D and E show the kinetics of oxidation/repair on plus strand (+) (on left) and minus strand (-) (on right) respectively on ERE, CnG, TATA, Poly_A and INTRON 1 regions at 0, 5, 15, 30, 45 and 60 min after estrogen induction.

Figure 6
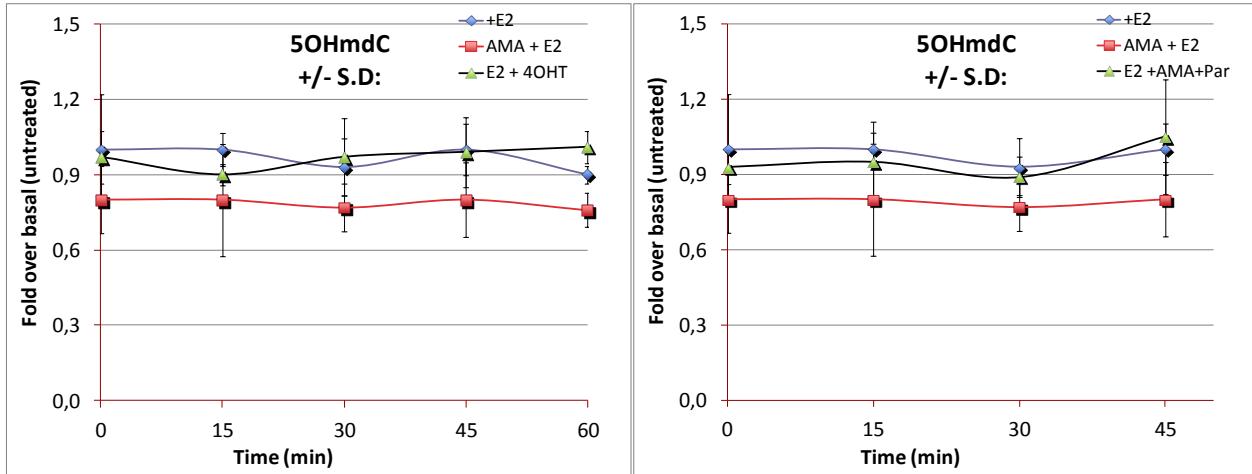
Total genome analysis of 8-oxo-deoxyguanosine (5OHmdC). DNA extracted from MCF7 stimulated with 50 nM of E_2 alone (blue lines) or in combination with 2,5 μM α-amanitin (AMA) (red line) or 10 nM 4-hydroxytamoxifen (4-OHT) (green line in panel on the left) or 10 μM parnate (Par) (green line in panel on the left) at different time (0, 15, 30, 45 and 60 min) was analyzed through spectrometry to evaluate the 5OHmdC accumulation in genome. The results shown are normalized on E_2 unstimulated sample. S.E.M was employed as statistical analysis.

Figure 7
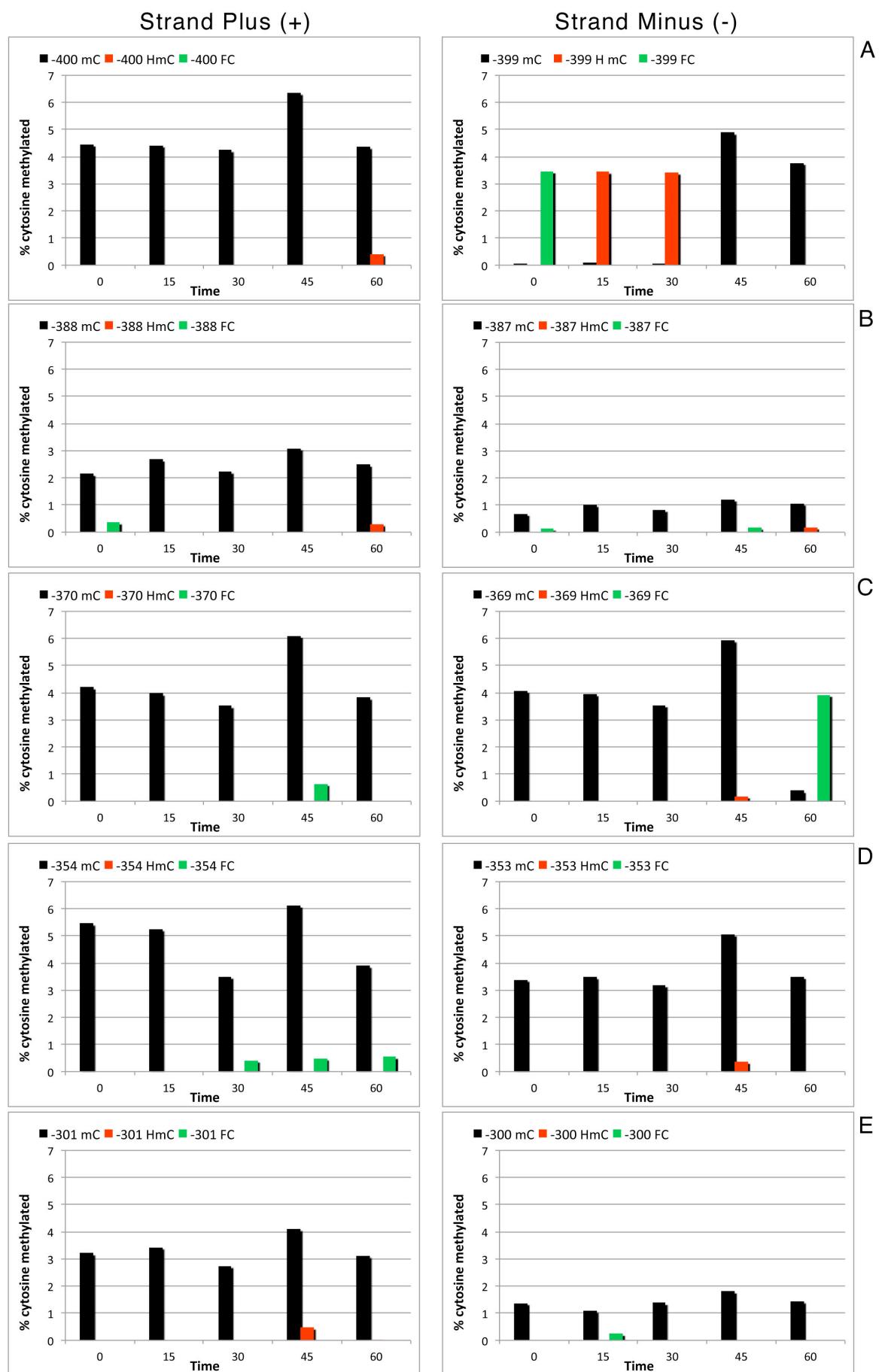
Estrogen induces strand specific oxidation and repair of 5hmC. MCF7 cells were treated with 50nM of E_2 at the indicated times (0, 15, 30, 45 and 60 min). DNA was extracted and used for OxBS-sequencing to determine the methylation and oxidation of C induced by estrogens of **5** CpG on ERE region. The use of potassium perruthenate (KRuO4) oxidizes 5-hydroxymethyl-cytosine (5hmC) into 5-formyl-cytosine (5fC), which after bisulfite conversion is read as uracil in the sequence. **A**, **B**, **C**, **D** and **E** panels show the percentage of modified Cs on plus strand (on the left) and minus strand (on the right) respectively of CG41, CG53, CG71, CG87 and CG140. The black task bar represent methylcytosine (mC), red ones hydroxymethyl-cytosine (HmC) and green ones formilcytosine (FC). The statistical analysis derived from at least 2 experiments in duplicate.

Figure 8
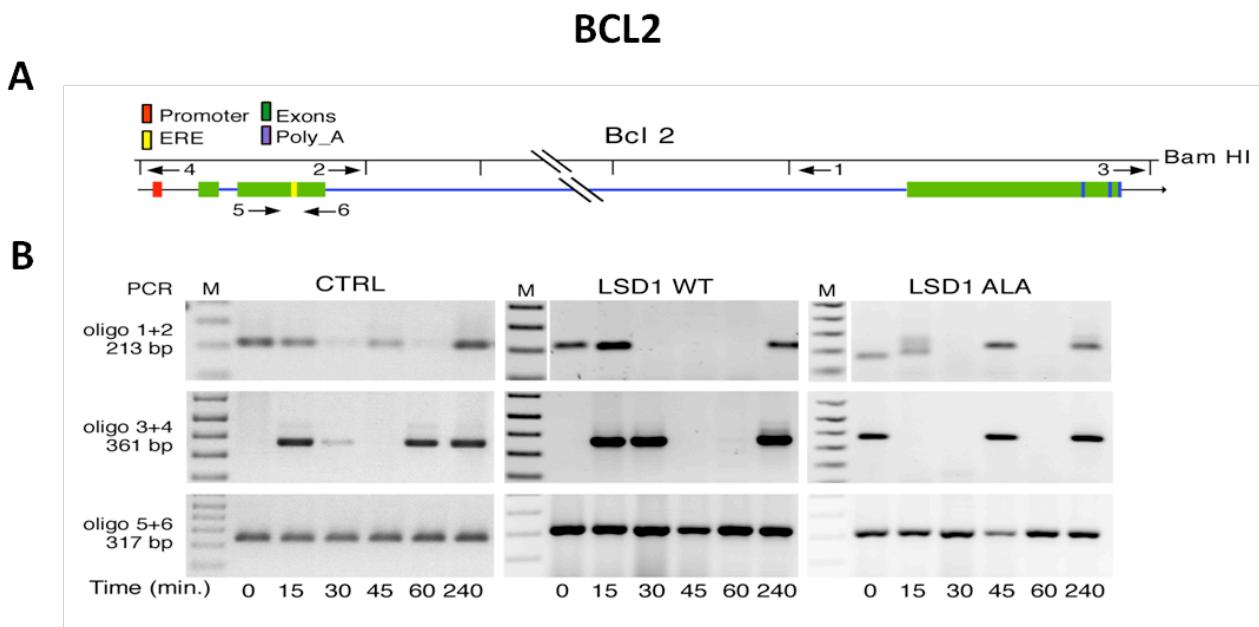
Formation of dynamic chromatin loops during early E$_2$-induced transcription. 3C analysis of *BCL2* chromatin in MCF7 cells exposed to 50 nM of E$_2$ for various periods of time. **A-** Schematic representation of gene and sites of Bam H1 enzyme. All the combinations of primers indicated, were performed on ligated chromatin. Each loop was detected with different primers pairs. **B-** Time course of chromatin looping during E$_2$ induction in control cells, cells transfected with LSD1 WT and with a dominant negative LSD1 ALA. 3C analysis was carried out as described in Methods and the loops shown in panels **B** were quantified by qPCR and verified by gel electrophoresis and DNA sequencing (data not shown). Oligos 1-2 in higher panel show the ERE-polyA interaction (213 bp stripe); the middle high panels show the TSS-polyA interaction (oligos 3-4, 361 bp stripe); the lower panels show control region (oligos 5-6, 317 bp stripe). Shown results derive from a least 3 experiments in triplicate (n≥9; Mean±SD).

Figure 9
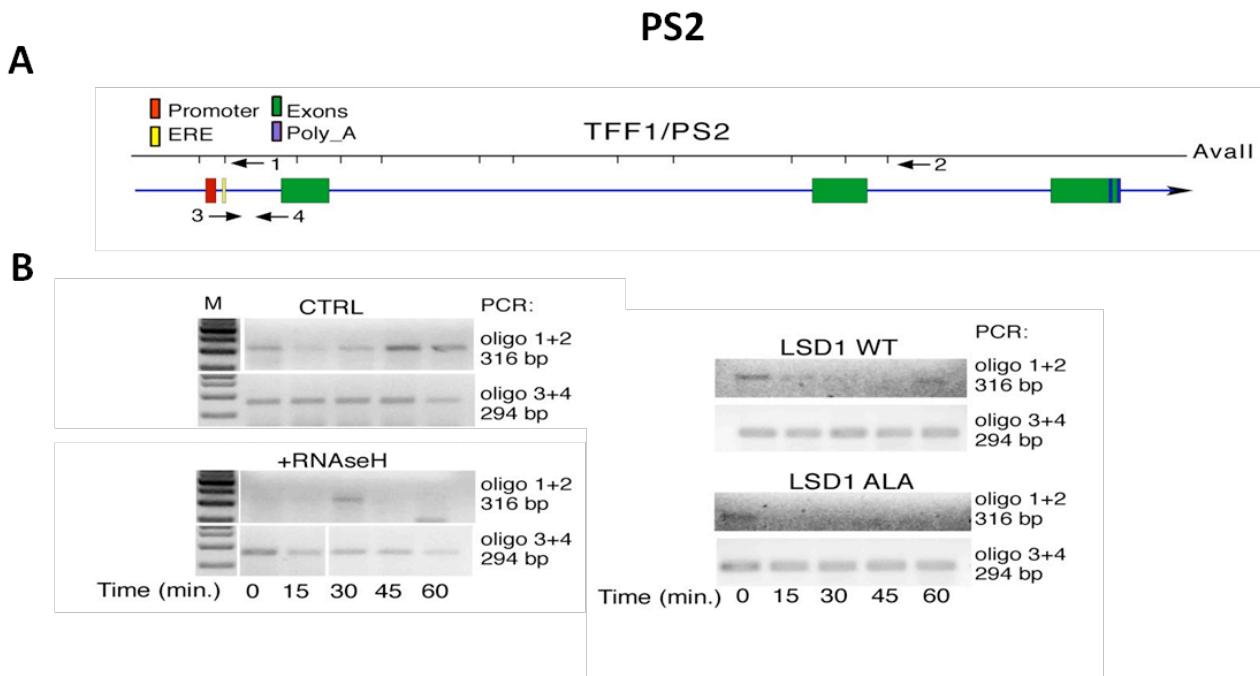
Formation of dynamic chromatin loops during early E$_2$-induced transcription. 3C analysis of *PS2* chromatin in MCF7 cells exposed to 50 nM of E$_2$ for various periods of time. **A-** Schematic representation of gene and sites of Ava II enzyme. All the combinations of primers indicated, were performed on ligated chromatin. Each loop was detected with different primers pairs. **B-** Time course of chromatin looping during E$_2$ induction in control cells, cells transfected with LSD1 WT and with a dominant negative LSD1 ALA. 3C analysis was carried out as described in Methods and the loops shown in panels **B** were quantified by qPCR and verified by gel electrophoresis and DNA sequencing (data not shown). Oligos 1-2 in panels show the ERE/TSS-polyA interaction (316 bp stripe); oligos 5-6 show the amplification of control region (294 bp stripe). Panel **+RNasi H** shows the kinetics of chromatin looping after the treatment with RNasiH. Shown results derive from a least 3 experiments in triplicate (n≥9; Mean±SD).

Figure 10

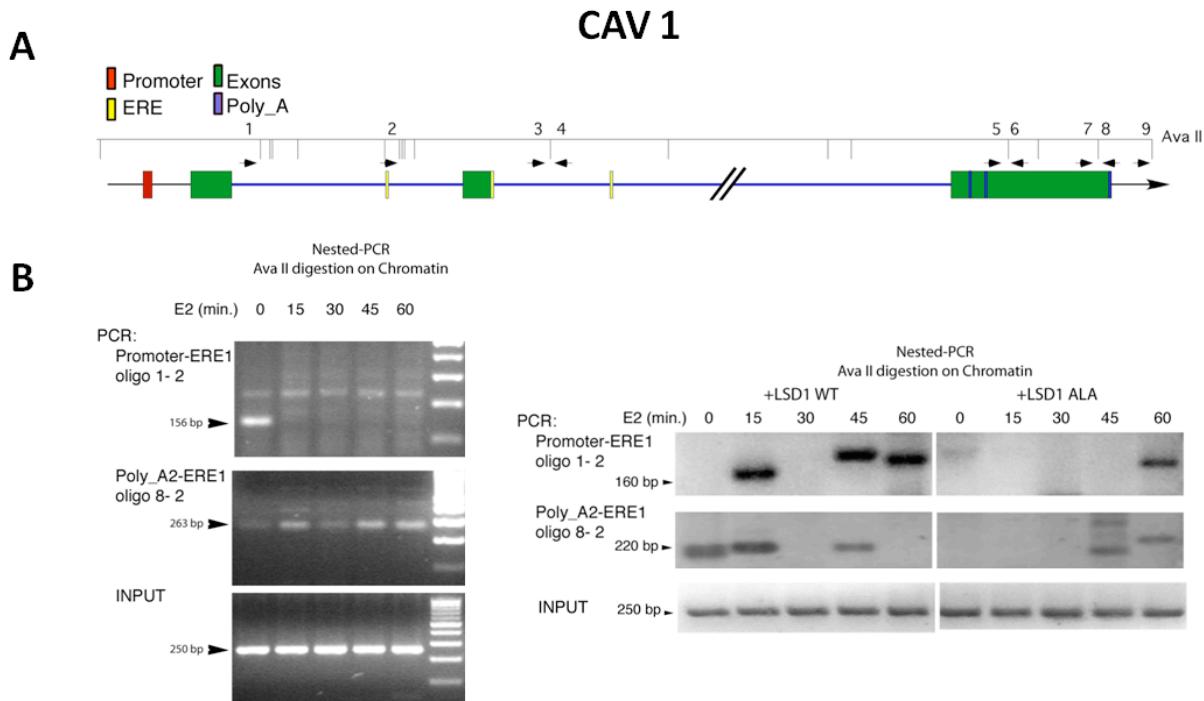**Formation of dynamic chromatin loops during early E$_2$-induced transcription.** 3C analysis of *CAV1* chromatin in MCF7 cells exposed to 50 nM of E$_2$ for various periods of time. **A-** Schematic representation of gene and sites of Ava II enzyme. All the combinations of primers indicated, were performed on ligated chromatin. Each loop was detected with different primers pairs. **B-** Time course of chromatin looping during E$_2$ induction in control cells, cells transfected with LSD1 WT and with a dominant negative LSD1 ALA. 3C analysis was carried out as described in Methods and the loops shown in panels **B** were quantified by qPCR and verified by gel electrophoresis and DNA sequencing (data not shown). Oligos 1-2 in panels show the Promoter/ERE1 interaction (156 bp stripe); oligos 8-2 show the polyA2-ERE1 interaction (263 bp stripe); last to panels refer to control region amplification (250 bp stripe). Shown results derive from a least 3 experiments in triplicate (n≥9; Mean±SD).
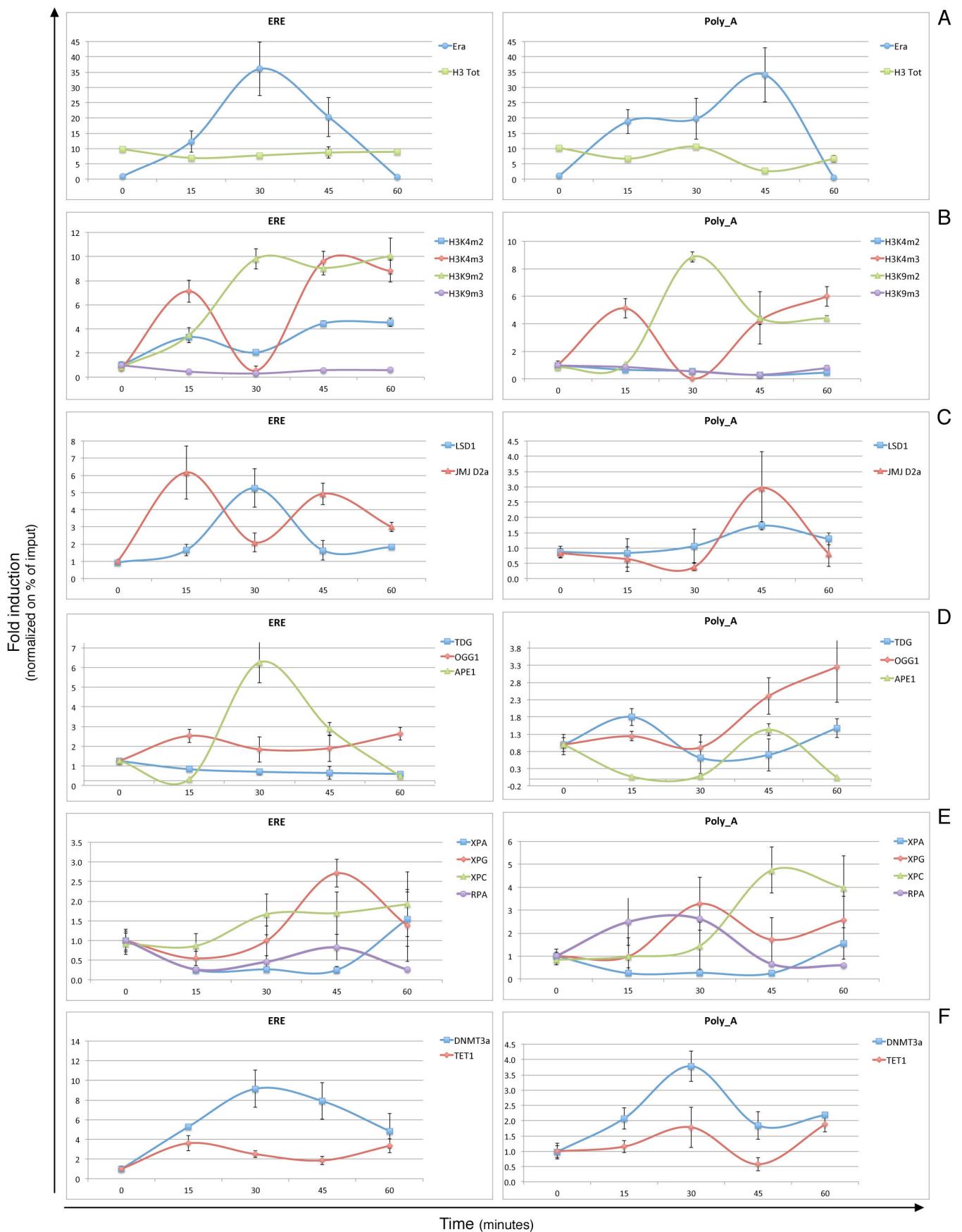
Figure S1

Intertwined Cycles of histone H3 methylation and recruitment of DNA repair enzymes at ERE and polyA sites of CAV1 gene. MCF7 cells were serum starved and exposed to 50 nM E_2 at the indicated times (0, 15, 30, 45 and 60 minutes). **A-** qChIP was carried out using specific antibodies recognized estrogen receptor alpha (ERα), H3K9Ac and total H3 on the estrogen responsive elements (ERE) and Poly_A of *CAV1* gene. Estrogen stimulation induces histone code changes; panel **B** shows H3K4m2/m3 and H3K9m2/m3 occupancy on CAV1 ERE and polyA; The histone methylation and demethylation changes depend on the recruitment of histone demehylases on gene regulative regions. Panel **C** shows the recruitment of LSD1 and JMJD2A on above-mentioned regions. Histone methylation cycles overlap with the recruitment of BER, NER, DNMT3a and TET1. **D-** MCF7 cells were analyzed by qChIP using specific antibodies recognizing the OGG1, APE1 and TDG on ERE and Poly_A of CAV1. Panel **E** shows the recruitment of NER enzymes (XPA, XPC, XPG and RPA) on studied regions; panel **F** shows the recruitment of DNMT3a and TET1 on the same regions of CAV1. The results were normalized on estrogen unstimulated sample and on the percent of input. The statistical analysis derived from at least 3 experiments in triplicate (n ≥9; Mean±SD).
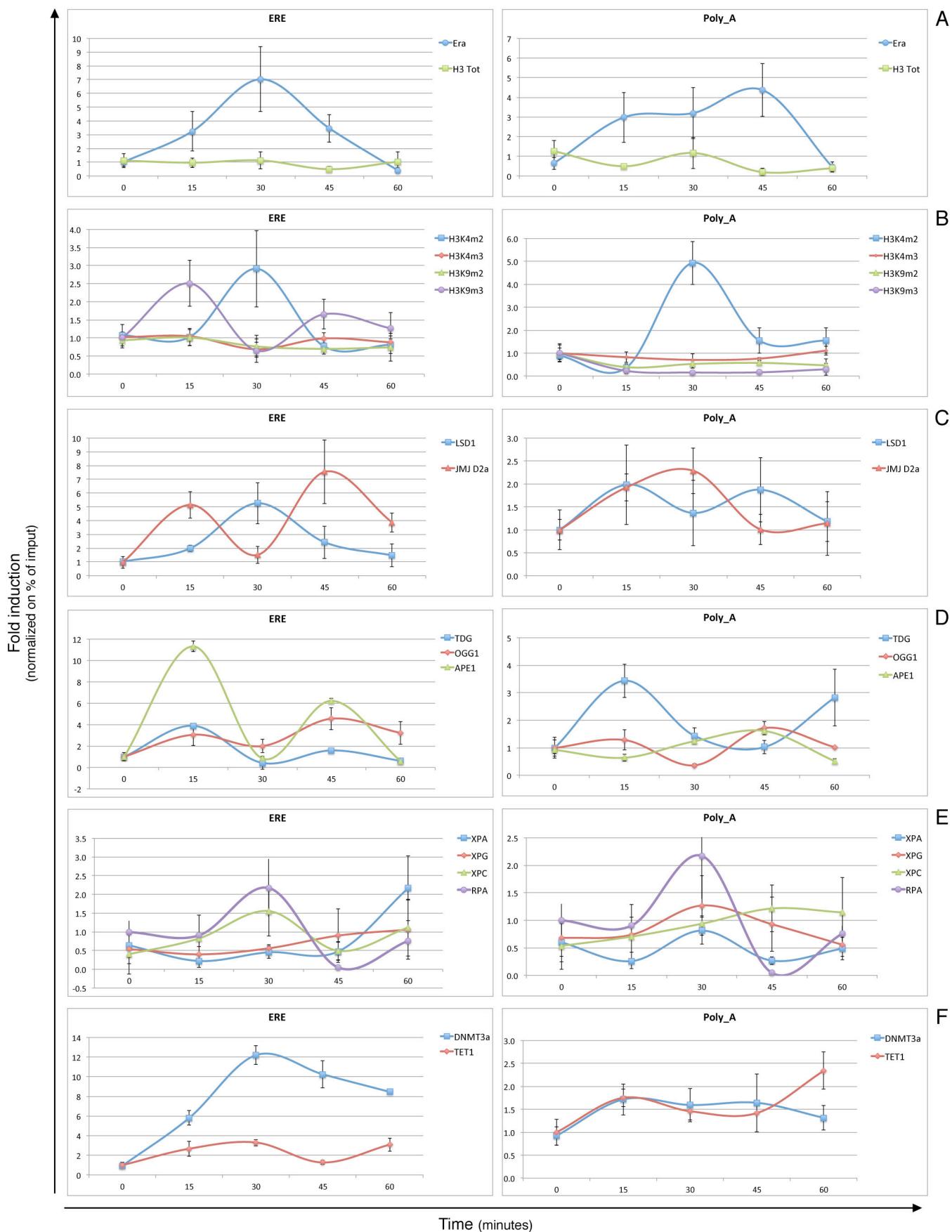
Intertwined Cycles of histone H3 methylation and recruitment of DNA repair enzymes at ERE and polyA sites of BCL2 gene. MCF7 cells were serum starved and exposed to 50 nM E_2 at the indicated times (0, 15, 30, 45 and 60 minutes). **A-** qChIP was carried out using specific antibodies recognized estrogen receptor alpha (ERα), H3K9Ac and total H3 on the estrogen responsive elements (ERE) and polyA of *BCL2* gene. Estrogen stimulation induces histone code changes; panel **B** shows H3K4m2/m3 and H3K9m2/m3 occupancy on BCL2 ERE and Poly_A; The histone methylation and demethylation changes depend on the recruitment of histone demehylases on gene regulative regions. Panel **C** shows the recruitment of LSD1 and JMJD2A on above-mentioned regions. Histone methylation cycles overlap with the recruitment of BER, NER, DNMT3a and TET1. **D-** MCF7 cells were analyzed by qChIP using specific antibodies recognizing the OGG1, APE1 and TDG on ERE and Poly_A of BCL2. Panel **E** shows the recruitment of NER enzymes (XPA, XPC, XPG and RPA) on studied regions; panel **F** shows the recruitment of DNMT3a and TET1 on the same regions of BCL2. The results were normalized on estrogen unstimulated sample and on the percent of input. The statistical analysis derived from at least 3 experiments in triplicate (n ≥9; Mean±SD).

Table 1

	PRIMERS FOR ChIP	LOCUS
ChIP ERE Fw	5'-CTAGACGGAATGGGCTTCAT-3'	pS2 / TFF1
ChIP ERE Rev	5'-TCTGAGAGGCCCTCCCGCCAG-3'	pS2 / TFF1
ChIP Poly_A Fw	5'-CTACTCACTGCGGATGCCCCAG-3'	pS2 / TFF1
ChIP Poly_A Rev	5'-GCTTCTGTATCCCTCCTCTGCTG-3'	pS2 / TFF1
ChIP ERE Fw	5'-CATTATAAGCTGTCGCAGAG-3'	BCL2
ChIP ERE Rev	5'-GAGGGTCAGGTGGACCACAG-3'	BCL2
ChIP Poly_A Fw	5'-AGTAAATGTGCCCAGCCTCT-3'	BCL2
ChIP Poly_A Rev	5'-TAGGGATGGTTCTCTGTTGC-3'	BCL2
ChIP ERE Fw	5'-GGATCTTAGATAAAGCTGGAAGG-3'	CAV1
ChIP ERE Rev	5'-GATCTCGCAGAGGACACCACAC-3'	CAV1
ChIP Poly_A Fw	5'-GATGTGATTGCAGAACCA-3'	CAV1
ChIP Poly_A Rev	5'-CAACAGCTTCAAAGAGTG-3'	CAV1
	PRIMERS FOR 3C	LOCUS
OLIGO 1 (Prenested)	5'-CACGGACACTGGATCTTGACTAT-3'	BCL2
OLIGO 1 (Nested)	5'-CCCCTTCCTCTTGCTGGTATC-3'	BCL2
OLIGO 2 (Prenested)	5'-CTTCTTGCCCTTCTGATGGTTG-3'	BCL2
OLIGO 2 (Nested)	5'-TATGCGCGTGGGAGGTGT-3'	BCL2
OLIGO 3 (Prenested)	5'-AATCCTCGGCTGTTCCCTGTTCAC-3'	BCL2
OLIGO 3 (Nested)	5'-TCTGCCTTCCCCATTCAACTATTA-3'	BCL2
OLIGO 4 (Prenested)	5'-CGACGACTTCTCCCGCCGCTACC-3'	BCL2
OLIGO 4 (Nested)	5'-GGACGGGGTGAACTGGGGGAGGAT-3'	BCL2
OLIGO 5	5'- AGTTGCTAACAGTGATGAGAGGCT -3'	BCL2
OLIGO 6	5'- ACCGAGACCCCTCTTGCTC -3'	BCL2
OLIGO 1 (Prenested)	5'-GATTAAGTCTCCTTCTGGGA-3'	PS2
OLIGO 1 (Nested)	5'-AGTGAACGTGGTTCACTC-3'	PS2
OLIGO 2 (Prenested)	5'-TCAGATTTCGTCATGCCTTC -3'	PS2
OLIGO 2 (Nested)	5'-CTCTATCCTAGGCAAGCTA-3'	PS2
OLIGO 3	5'-CCCATGGGAAAGAGGGACTTTC-3'	PS2
OLIGO 4	5'-GTGGCCATTGCCTCCTCTCTGC-3'	PS2
OLIGO 1 (Prenested)	5'-CTGGGCGGGGAGGTGAAGAG-3'	CAV1
OLIGO 1 (Nested)	5'-GAAGAGAAGCCAGGAATGTTTTAT-3'	CAV1
OLIGO 2 (Prenested)	5'-ACGGGAACGCGAAACAGGTGAAGC-3'	CAV1
OLIGO 2 (Nested)	5'-CCGGGAGAAGCCTGCGGCTGC-3'	CAV1
OLIGO 8 (Prenested)	5'-GAGGCAATACCAAGGTTCATCATC-3'	CAV1
OLIGO 8 (Nested)	5'-AAGTTGGGTGTGAGTGGATTTAAAG-3'	CAV1

Table 2

	PRIMERS FOR LPA	LOCUS
ERE LPO MINUS	5'-GGGTTCCCTAAGGGTTGGACCACTGCAGAAGTGATTCATAGTGAGAGATGGCC-3'	pS2 / TFF1
ERE RPO MINUS	5'-GGAAAAAGGCTTGGCCGTGACAACAGTGGCTCACGGGGTGTCTAGATTGGATCTTGCTGGCAC-3'	pS2 / TFF1
ERE LPO PLUS	5'-GGGTTCCCTAAGGGTTGGACACCCCGTGAGCCACTGTTGTCACGGCCAAGCCTTTTTCC-3'	pS2 / TFF1
ERE LPO PLUS	5'-GGGTTCCCTAAGGGTTGGACACCCCGTGAGCCACTGTTGTCACGGCCAAGCCTTTTTCC-3'	pS2 / TFF1
ERE RPO PLUS	5'-GGCCATCTCTCACTATGAATCACTTCTGCAGTGGTCTAGATTGGATCTTGCTGGCAC-3'	pS2 / TFF1
INTRON LPO PLUS	5'-GGGTTCCCTAAGGGTTGGACCAAAAGCAGGTGGAGAGTAACTCAGGGTGGCAGGGCCCCC-3'	pS2 / TFF1
INTRON RPO PLUS	5'-GGAGACCTTCGAGAAGTGCGACGAGGAGGGGGCTGTCTAGATTGGATCTTGCTGGCAC-3'	pS2 / TFF1
INTRON LPO MINUS	5'-GGGTTCCCTAAGGGTTGGACAGCCCCCTCCTCGTCGCACTTCTCGAAGGTCTCC-3'	pS2 / TFF1
INTRON RPO MINUS	5'-GGGGGCCCTGCCACCCTGAGTTACTCTCCACCTGCTTTTGGTCTAGATTGGATCTTGCTGGCAC-3'	pS2 / TFF1
TATA LPO PLUS	5'-GGGTTCCCTAAGGGTTGGACTAACGCTCTTTAAGCAAACAGAGCCTGCCCTATAAAATCC-3'	pS2 / TFF1
TATA RPO PLUS	5'-GGGGCTCGGGCGGCCTCTCATCCCTGACTCGGGGTCGTCTAGATTGGATCTTGCTGGCAC-3'	pS2 / TFF1
TATA LPO MINUS	5'-GGGTTCCCTAAGGGTTGGACGACCCCGAGTCAGGGATGAGAGGCCGCCCGAGCCCC-3'	pS2 / TFF1
TATA RPO MINUS	5'-GGATTTTATAGGGCAGGCTCTGTTTGCTTAAAGAGCGTTAGTCTAGATTGGATCTTGCTGGCAC-3'	pS2 / TFF1
CnG LPO PLUS	5'-GGGTTCCCTAAGGGTTGGACTTCCCCCTGCAAGGTCACGGTGGCCACCCCGTGAGCCACT-3'	pS2 / TFF1
CnG RPO PLUS	5'-GTTGTCACGGCCAAGCCTTTTTCCGGCCATCTCTCACTATGTCTAGATTGGATCTTGCTGGCAC-3'	pS2 / TFF1
CnG LPO MINUS	5'-GGGTTCCCTAAGGGTTGGACATAGTGAGAGATGGCCGGAAAAAGGCTTGGCCGTGACAACA-3'	pS2 / TFF1
CnG RPO MINUS	5'-GTGGCTCACGGGGTGGCCACCGTGACCTTGCAGGGGGAAGTCTAGATTGGATCTTGCTGGCAC-3'	pS2 / TFF1
Poly_A LPO PLUS	5'-GGGTTCCCTAAGGGTTGGACCTTTGACTACTCAAAATTGGCCTAAAAATTAAAAGAGATC-3'	pS2 / TFF1
Poly_A RPO PLUS	5'-GATATTAATCTGTGCTGTTCATTCTTTTAAAGAATATGAAGTCTAGATTGGATCTTGCTGGCAC-3'	pS2 / TFF1
Poly_A LPO MINUS	5'-GGGTTCCCTAAGGGTTGGACTTCATATTCTTTAAAAGAATGAACAGCACAGATTAATATC-3'	pS2 / TFF1
Poly_A RPO MINUS	5'-GATCTCTTTTAATTTTTAGGCCAATTTTGAGTAGTCAAAGGTCTAGATTGGATCTTGCTGGCAC-3'	pS2 / TFF1
Universal LPA Fw	5'-GGGTTCCCTAAGGGTTGGA-3'	pS2 / TFF1
Universal LPA Rw	5'-GTGCCAGCAAGATCCAATCTAGA-3'	pS2 / TFF1

89

Table 3

	PRIMERS FOR OXIDATIVE BISULFITE	LOCUS
ERE bis Minus Rw	5'-TAAACTTCATAAACTCCTTCCCTT-3'	pS2 / TFF1
ERE bis Minus Fw	5'-GGTTTTATTTATATTTGAGAGGTTTTTT-3'	pS2 / TFF1
ERE BIS plus Fw	5'-ATGGGTTTTATGAGTTTTTTTT-3'	pS2 / TFF1
ERE BIS plus Rw	5'-AAAATCTCCTCCAACCTAACCTTAA-3'	pS2 / TFF1

Chapter 4: DNA damage and direct site repair

DNA double-strand breaks (DSBs) is the most hazardous lesions for the cells. DSBs are generated when the two complementary strands of the DNA double helix are broken simultaneously at nearby sites. Physically dissociated ends can recombine inappropriately with other genomic sites, leading to chromosomal translocations. Eukaryotic cells have evolved complex, highly conserved systems to rapidly and efficiently detect these lesions and bring about their repair. These lesions are repaired either by homologous recombination (HR) or by nonhomologous end joining (NHEJ). HR can lead to gene conversion, a conservative process in which a donator DNA sequence with homology is copied into a broken locus, making the repaired locus an exact copy of donator sequence without altering the donator sequence.

Cuozzo et al. have shown that there is an association between DNA damage, repair with homologous recombination and DNA methastable methylation (methylation pattern generate as consequence of DNA damage repair). To explore the molecular mechanisms involved in these processes we have been used system pioneered by M. Jasin (Jasin M. 1996; Pierce A.J. et al., 1999) DR-GFP plasmid. In this construct can be induced double-strand break (DSB) in "reporter" gene through the cleavage with a meganuclease, the repair of the damage if occurs with the homologous recombination (HR), re-establish the continuity of codogenic region and it allow the expression of "reporter" protein. DR-GFP plasmid contains two copies of GFP gene oriented as direct repeats and separated by a drug selection marker, the puromycin N-acetyltransferase gene. An upstream cytomegalovirus (CMV) enhancer fused to the chicken β-actin promoter provided a strong and insulated transcriptional unit. Both copies of GFP do not produce a functional protein; the upstream GFP gene (cassette I) is transcribed but produces a truncated protein as it has been subjected to mutagenesis substituting 11 bp of the wild-type gene for introduce a recognition site for I-SceI, a rare-cutting endonuclease that does not cleave several eukaryotic genomes tested (Jasin M. et al., 1996) These substituted base pairs introduce two in frame stop codons that terminate translation. The downstream GFP (cassette II) is inactivated by upstream and downstream truncations. After DBS if the damage is repaired with HR and if occur a gene conversion between two GFP genes, in which the cassette II is the donor of wild-type sequence can be produced a functional copy of GFP. The low rate of GFP$^+$ cells after DSB generation isolated with flow cytometric analysis permitted to hypothesize that some wild-type GFP recombinants were silenced by methylation, in fact the inhibition of methylation with 5-aza-2'-deoxycytidine (5-AzadC)

increased the number of GFP$^+$ cells (clones with single integration) in a pool of cells carrying DR-GFP. In addition FACS analysis have shown that there is a bimodal fluorescence distribution caused by two different recombinant populations high (HR-H) and low-expressing (HR-L) cells, it happens because hypomethylated/hypermethylated molecules are generated after DSB/HR in a 1:1 ratio. *Morano et al.* proved that the transient stalling of RNA polymerase II induced by α-amanitin during repair increased GFP methylation in fact the numember of L clones increase while the fraction of H clones decrease. They proposed that stalled RNA polymerase during repair favors the recruitment of enzymes that methylate the repaired DNA, consolidating the methylation of L clones. This event occurs only during repair because stalling Pol II before or after DSB repair does not modify GFP methylation and expression, and they depend on the stall of Pol II because the treatment with actinomycin-D that depletes RNA polymerase II from chromatin prevent the accumulation of H clones. The original methylation profiles induced by HDR are remodeled in a transcription-dependent fashion during the first 15 days after repair, in fact the transcription convert a fraction of L to H cells by favouring loss of methylation. In particular the process consists in the introduction of a DSB in a copy of the gene (cassetta I) through the nuclease I-SceI, the lesion undergoes 5'→3' end resection, after the 3' free single strand end of strand plus invades the DNA of the GFP cassette II, while 3' free single strand end of strand minus DNMT3a (DNA methyl transferase de novo) is recruited in the first 24h after damage and adds methyl groups. At 48h on this strand the stalling RNA Pol II stalling on this strand leads to the recruitment of DNMT1(DNA methyl transferase de novo) and NP95 that amplify the methylation of region, while on strand plus is recruited GAD45 which promotes the loss of methylation from repaired strand. So after repair hemi-methylated DNA is generated. In cells which are in replication phase, the passage of the replication fork before that DNMT1 copy methylation pattern on the complementary strand, generates two copies of GFP gene converted with different degree of methylation inherited from different cells; H clones with high fluorescence and with hypomethylated GFP gene and L clones with low fluorescence and with hypermethylated GFP gene.

Cleared the link between DNA damage, repair and *de novo* methylation in this paper we have explored if DNA methylation patterns at repaired site are associated with change of chromatin structure (chromatin methylation and looping) after and before repair.

SCIENTIFIC REP RTS

4.1

DNA damage and Repair Modify DNA methylation and Chromatin Domain of the Targeted Locus: Mechanism of allele methylation polymorphism

Giusi Russo[1,*], Rosaria Landi[1,*], Antonio Pezone[1,*], Annalisa Morano[1], Candida Zuchegna[2], Antonella Romano[2], Mark T. Muller[3], Max E. Gottesman[4], Antonio Porcellini[2] & Enrico V. Avvedimento[1]

We characterize the changes in chromatin structure, DNA methylation and transcription during and after homologous DNA repair (HR). We find that HR modifies the DNA methylation pattern of the repaired segment. HR also alters local histone H3 methylation as well chromatin structure by inducing DNA-chromatin loops connecting the 5′ and 3′ ends of the repaired gene. During a two-week period after repair, transcription-associated demethylation promoted by Base Excision Repair enzymes further modifies methylation of the repaired DNA. Subsequently, the repaired genes display stable but diverse methylation profiles. These profiles govern the levels of expression in each clone. Our data argue that DNA methylation and chromatin remodelling induced by HR may be a source of permanent variation of gene expression in somatic cells.

DNA methylation in somatic cells is associated with aging, chromatin changes and efficiency of transcription[1,2]. There are two types of DNA methylation: 1. A stable and invariant form – imprinting - which is sex-specific and identical in individuals and cells[3]; and, 2. A metastable somatic type that changes with age and differs among individuals and cells[4,5].

We have used a system pioneered by M. Jasin, in which a double-strand break (DSB) in a GFP gene generated by the meganuclease I-SceI is repaired by gene conversion from a second copy of the gene[6,7]. We, and others, have shown that DNA damage and homology-directed repair (HR) induce *de novo* methylation of the repaired segment. This methylation pattern is stably transmitted to daughter cells[8–10]. In the absence of selection, such as a neutral gene like GFP, the distribution of differentially methylated clones in the population is essentially random. We find two populations of cell clones, those that express high levels of GFP and clones that express low levels of GFP, referred to as Rec H and Rec L clones, respectively. Relative to the parental gene, the repaired GFP is hypomethylated in Rec H clones and hypermethylated in Rec L clones. The altered methylation pattern is largely restricted to a segment immediately 3′ to the DSB along the direction of transcription. Hypermethylation of this tract significantly modifies the local chromatin structure and reduces transcription[8,11]. These data nicely account for the high polymorphism of methylation profiles in cells populations derived from individual somatic tissues[12]. However, genome-wide methylation analysis suggests that different mechanisms may explain the loss or gain of methylation: 1. Stochastic processes may produce the high rate of methylation polymorphism; 2. Deterministic events may contribute to the gain or loss of methylation at specific loci. Moreover, changes in DNA methylation are strictly associated with post-translational modifications (PTMs) of histones. It is unclear if histone PTM

[1]Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Istituto di Endocrinologia ed Oncologia Sperimentale del C.N.R., Università Federico II, 80131 Napoli, Italy. [2]Dipartimento di Biologia, Università Federico II, 80126 Napoli, Italy. [3]Epigenetics Division, TopoGEN, Inc., 27960 CR319, Buena Vista, Colorado, 81211, USA. [4]Institute of Cancer Research, Columbia University Medical Center, New York, New York, 10032, USA. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.P. (email: antonio.porcellini@unina.it) or E.V.A. (email: avvedim@unina.it)

94

variations drive or are induced by local DNA methylation. These events could generate cells with the same geno-type but with various levels of gene expression.

We address the following three questions in this paper. First, what is the relationship between chromatin changes and DNA methylation at the site of repair. Second, what is the origin of the polymorphism of somatic DNA methylation. Third, does, in fact, the extent and pattern of methylation following repair impart variation of gene expression in cell populations with an identical genotype? We chose to approach these questions in a system in which DNA damage and repair can be controlled temporally and spatially, focusing our attention on local transient as well as permanent changes induced by damage and repair.

Results

Spatial and temporal changes of the histone H3 methylation code after homologous repair of a DSB at the GFP locus.

The DRGFP system. The critical features of the system we use to study repair and methylation can be summarized as follows. A reporter construct (Direct-Repeat GFP: DRGFP) is randomly integrated in the genome of Hela cells at an average copy number of 1. I-SceI induces a double-strand break (DSB) in one GFP copy (I cassette) that can be repaired from the second copy (II cassette) by homologous recombination (HR), yielding GFP+ clones. 75–90% of the cells are repaired by NHEJ with or without small deletions at the *I-SceI* site[6,13]. Importantly, GFP+ cells can arise in this system only by HR[6]. I-SceI expression starts 2 h after transfection with an I-SceI plasmid, peaks at 24 h and slowly decays up to 48 h. At 48 h total HR, measured by qPCR, is approx. 5.0 to 10%. Cells exposed to I-SceI but which are not GFP+ are termed UnRec (unrecombinant). GFP+ cells fall into two expression classes: high and low expressers, Rec H and Rec L, respectively.

Transient histone H3 methylation changes induced by a DSB. Our previous work showed that levels of GFP expression reflected local DNA methylation patterns at the repaired *I-SceI* site[8]. To determine if these patterns were associated with changes in chromatin methylation, we probed the histone H3 PTMs, methylation of lysines 4 and 9, at the DSB and flanking sites 24 h, 48 h and 7 days after I-SceI expression. Chromatin was immune-precipitated with antibodies recognizing H3, H3K4m2/3 (an activation marker) or H3K9m2/3 (a repression marker). The DNA was amplified with primers corresponding to several sites upstream or downstream of the DSB along with a reference gene, exon 9 of TSHR (Fig. 1A). ChIP experiments were repeated three times in triplicates (see Supplemental Statistical Tables).

Major alterations in chromatin modification (loss of H3K4me2/3 and gain of H3K9me2/3) appeared just 3′ to the DSB 24 h after the onset of I-SceI expression: These changes disappeared 7 days later, when I-SceI levels decayed (Fig. 1B, panels 6 and 10; and ref. 11). Histone H3 decreased transiently 5′ to the DSB, but there were few other changes in the histone methylation pattern in this segment (panels 5 and 9). The H3K9 or H3K4 methylation patterns at the PolyA addition site (panels 7 and 11) or within the control TSHR gene (panels 8 and 12) were unaltered.

The major fraction of GFP+ cells did not appear until 48 h. Since loss of H3K4me2/3 was detectable as early as 24 h post transfection, is likely that loss is induced by the DSB itself or by NHEJ repair of the DSB. These local histone methylation changes did not involve nucleosome or histone eviction, since total H3 content did not change significantly at the *I-SceI* site (Fig. 1B, panel 2).

Dynamics of histone H3 methylation changes after DNA repair. At 7 days post-HR, we analysed H3K4me2/3 (Fig. 1C) and found no clear change in status in control cells (no DSB), cells subjected to I-SceI, sorted GFP− (UnRec) and HR cells, GFP+ (Rec). We did observe, however that H3K9me2/3 marks appeared to be selectively modified in HR processed DSBs (Fig. 1C). To explore these changes in H3K9 methylation, we sorted the high GFP+ (Rec H) and the low (Rec L) GFP+-expressing cells, which had been repaired by HR (Fig. 2A) and analysed the histone modification at various sites using primers in Fig. 1A. This analysis was performed at 14 days post I-SceI exposure (a time when there was no further HR). This analysis yielded the following results. First, we saw little change in H3K4me2/3 levels in the intron 5′ to the DSB (primers a/c) when we compared control cells to Rec H or Rec L cells (Fig. 2B, panel 1). Second, H3K9me2/3 levels in Rec H cells were significantly reduced relative to control and Rec L cells (Fig. 2B panel 4). Third, the region immediately 3′ to the *I-SceI* site (primers r/h) was enriched in H3K4me3 in Rec H cells and in H3K9me3 in Rec L cells (Fig. 2B, panel 2 and 5). Fourth, GFP+ cells (both Rec H and Rec L) were strikingly different from control cells at the polyA addition site (primers p/q). Specifically, Rec H cells were characterized by high H3K4me3 levels (Fig. 2B, panel 3) whereas Rec L cells showed elevated H3K9me2/3 levels (Fig. 2B, panel 6). These changes were detectable in the population of HR cells but not in the mass, unsorted, population (Fig. 1B, panels 7 and 11, and 1C). From these data, we conclude that chromatin corresponding to the repaired GFP gene in recombinant GFP+ cells showed significant localized changes of histone H3 K4-K9 methylation markers compared to control cells. H3K9me3 was elevated at the 3′ end of the DSB in Rec L cells, and depleted at the 5′ end of the DSB in Rec H cells. Conversely, H3K4me3 accumulated mainly at the polyA site in Rec H cells. Loss of the repressive H3K9me2/3 marks at the promoter region in Rec H cells and the dramatic increase in H3K9me3 just 3′ to the repaired DSB in Rec L cells are related to high and low GFP mRNA levels in this region in H and L clones, respectively[8].

The changes in the methylation represent a permanent, HR-associated, modification in Rec L cells. Recall that Rec L cells acquired new methylated CpGs at the DNA region immediately 3′ to the DSB after recombination, consistent with the idea that histone and DNA methylation are causally related.

DNA Methylation stabilizes the H3K9m2/3 marker. Why does the elevated H3K9m3 mark persist in Rec L cells? We hypothesized that HR-induced methylation of repaired DNA may maintain H3K9m3 at the repaired site. To test this idea, we de-methylated DNA in Rec L cells with 5-azadC and measured H3K9m2/3 content at three
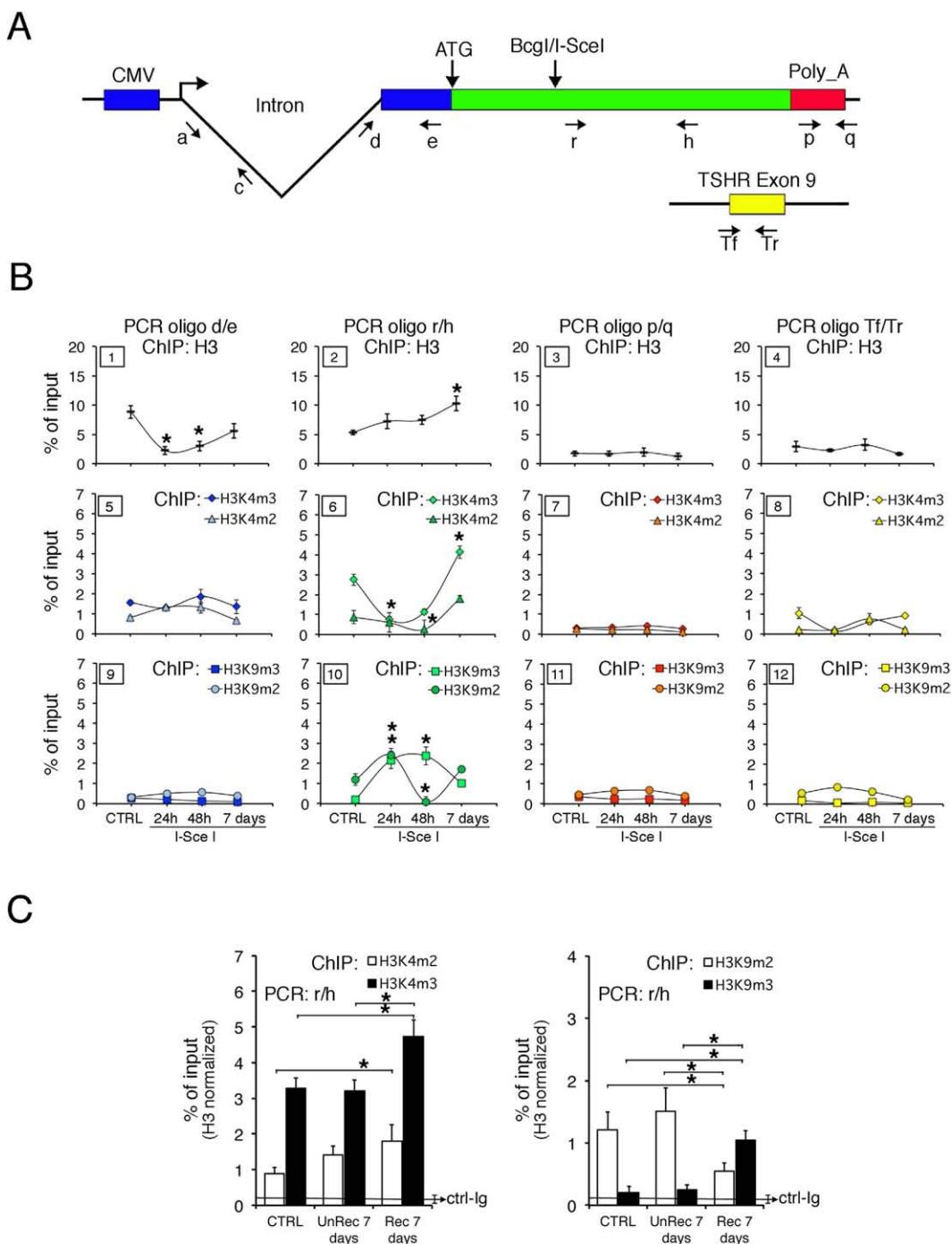
Figure 1. Spatial and temporal changes of histone H3 PTM after DSB at the GFP locus. (A) Schematic diagram of DRGFP plasmid and a reference locus (TSHR exon 9). Structure of the DRGFP plasmid integrated as single copy in different locations in Hela cells. Primers a and c cannot be used in cells transiently transfected with I-SceI plasmid because they are also present in intron 1 of the I-SceI expression vector. **(B)** H3K4m2/3 and H3K9m2/3 content of GFP and the reference gene. DRGFP cells were transfected with I-SceI and characterized 24 h, 48 h and 7 days later. Cells were fixed and the chromatin analyzed by ChIP with the indicated antibodies. qPCR on each immunoprecipitate was carried out with the primers indicated in A. The specific antibodies are indicated at the top of each column. Each panel is identified by a numbered box in the upper left side.*P < 0.01 (*t test*) as compared with untreated control or basal. **(C)** H3K4m2/3 and H3K9m2/3 content in cells sorted 7 days after transfection. CTRL are cells transfected with a control plasmid; UnRec were GFP⁻ cells sorted and separated from GFP⁺ Rec cells after I-SceI transfection. *P < 0.01 (*t test*) as compared with UnRec or CTRL. The detailed statistical analysis of the data shown in panels (B,C) is reported in Supplemental Statistical Tables 1 and 2.

96
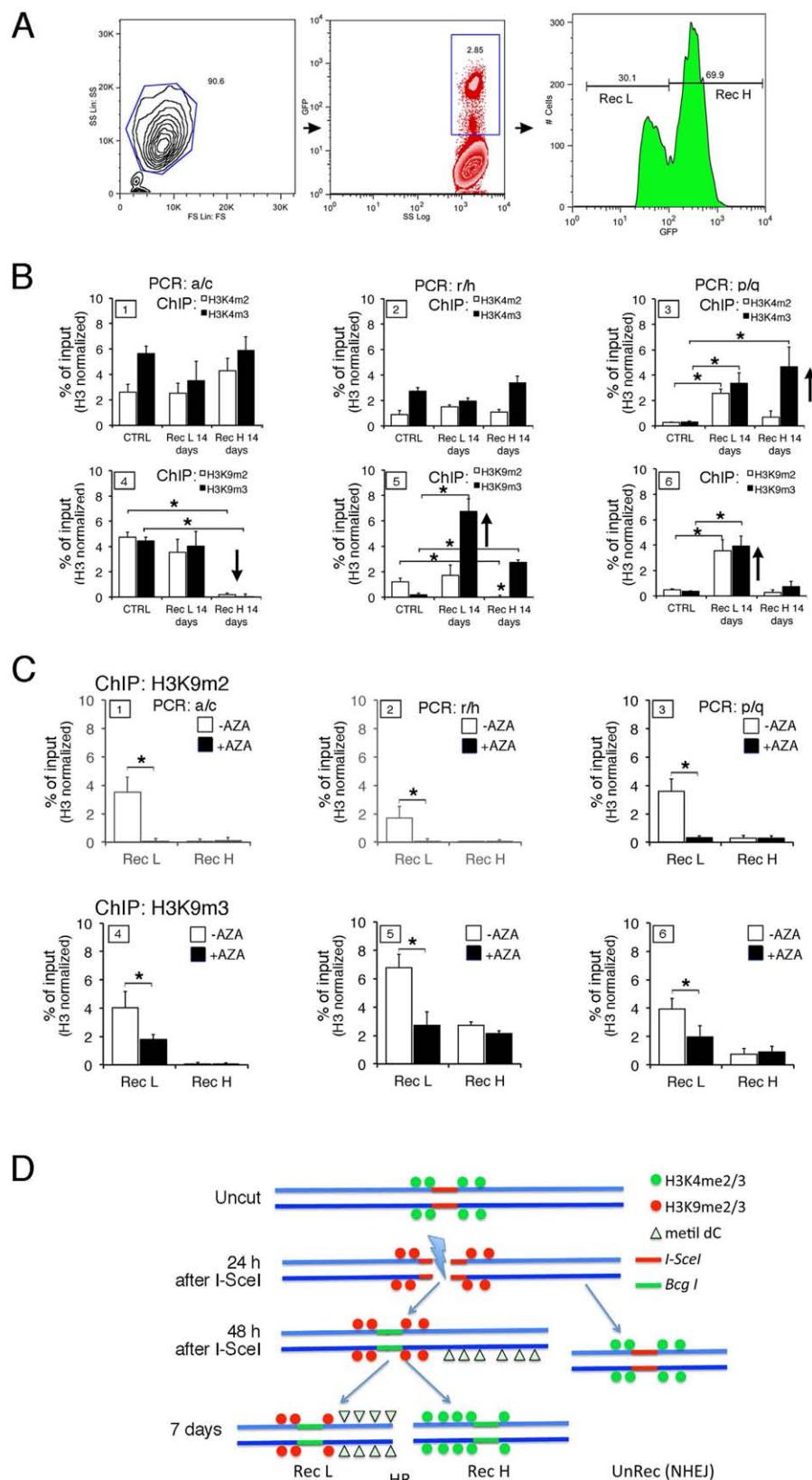
Figure 2. Spatial and temporal changes of histone H3 K4-K9 methylation after homologous repair at the GFP locus in HR cells. (**A**) DRGFP Hela cells were transfected with the I-SceI vector and sorted for GFP expression 14 days later as described in Methods. The panels from the left to the right show the gating strategy used to sort Rec H and Rec L cells. The percent of viable cells was 90.6. The fraction of the total GFP$^+$ cells (middle panel) or of Rec H and Rec L (left panel) is indicated. (**B**) H3K4m2/3 and H3K9m2/3 levels at GFP chromatin in purified Rec H and Rec L cells, 14 days after I-SceI transfection. The specific primers are indicated

at the top of each column as shown in Fig. 1A. The data are normalized to total H3 content. *P < 0.01 (*t test*) as compared with cells transfected with control plasmid (CTRL). (**C**) H3K9m2/3 content of GFP in chromatin derived from sorted Rec H or Rec L clones, untreated or treated with 5-azadC (10 μM for 3 days and analyzed 4 days later). ChIP analysis was performed with the indicated antibodies. qPCR on each immunoprecipitate was carried out with primers r/h, as shown in Fig. 1A. *p < 0.01 (*Wilcoxon rank-sum test*) as compared with untreated control or (primers a/c). (**D**) Histone H3 and DNA methylation changes following the DSB and HR, A schematic cartoon illustrating the major chromatin changes of H3K4me2/3 or H3K9me2/3 and DNA methylation (Δ) following the DSB and the HR or NHEJ in recombinant (HR) and non recombinant (NHEJ) cells. The detailed statistical analysis of the data shown in panels (B,C) is reported in Supplemental Statistical Tables 3 and 4.

· ·

locations in the GFP locus (Fig. 2C; primers defined in Fig. 1A). We found that 5-azadC significantly reduced H3K9m2 and H3K9m3 content in Rec L cells throughout the gene (Fig. 2C). Thus, DNA methylation at the repaired GFP locus is required to maintain chromatin H3K9 methylation within the region. Taken together, these data suggest that the extensive, rapid, localized and transient increase of H3K9m2/3 is induced by DSB formation in all treated cells (Fig. 1B, panel 10). Seven days later (after most of the breaks are repaired), this mark largely disappears and is replaced by the activation marker H3K4m2/3. In the low-expressing Rec L fraction, in which GFP DNA is hypermethylated following HR, the repressive H3K9m2/3 marker remains at the *I-SceI* site. A schematic model of the histone methylation changes induced by the DSB and HR is shown in Fig. 2D.

Chromatin looping induced by damage and repair. We suggest that *de novo* DNA methylation at the site of DSB repair stabilizes the H3K9m3 mark, which alters the chromatin structure of the entire GFP gene. To map specific domains modified by DNA damage and repair, we examined the structure of chromatin at the repaired locus in UnRec (repaired by NHEJ) and Rec cells by chromosome conformation capture (3C). Among the primers used (Fig. 3A,B), only a few were able to amplify specific DNA fragments. PCR and sequence analysis showed these segments to be contiguous in the chromatin but not in the DNA. Two looped segments are shown in Fig. 3A. Loops *A* and *C* are specific to recombinant cells and mark different regions of GFP cassette I at the 5′ end relative to the *I-SceI* site: loop *A* links a region that includes a GFP transcription start site driven by the chicken β-actin promoter (5′ end, identified by primer *d*, from −517 to −279; green loop). Loop *C* includes a more distal 3′ region of the GFP coding sequence located downstream to the *I-SceI* site (5′ end identified by primers *f*, *g*, from +70 to + 300; blue loop). Figure 3A, right panel, shows that 48 h after I-SceI expression, at which time repair was complete, both loops were detectable in the mass culture. Thus, these loops form soon after the repair process. Figure 3B shows the chromatin loops detected in Rec H and Rec L sorted cells after I-SceI transfection. In addition to loops A and C, two other chromatin loops juxtapose different elements of the DRGFP insert. One loop connects the PGK1-PolyA addition site of the puromycin acetyl transferase gene with its promoter (loop D in red, Fig. 3B). This loop, present both before and after DNA damage and repair, marks the border of the puromycin acetyl transferase gene transcription unit, and serves as an internal positive control. Loop B, present in both recombinant and UnRec cells, serves as an additional internal control (Fig. 3B, orange loop, 5′ end identified by primer *e*). Loops A and C mark selectively recombinant cells as shown in Fig. 3A. Specifically, the frequency of loop A is high in Rec H and low in Rec L cells, whereas loop *C* abundance was the converse, low in Rec H and high in Rec L cells. Note that the 5′ end of loop C corresponds to the segment of GFP that is *de novo* methylated after repair. It is possible that local DNA methylation influences the formation or stability of loop C in L cells. Consistent with this notion, Rec L cells exposed to 5-azadC (and consequently hypomethylated) for 72 h formed less loop C and more loop A. Thus, Rec L cells may convert to Rec H cells following loss of DNA methylation (Fig. 3C, panels 1 and 3). These data suggest that these chromatin loops are related to the transcription efficiency of the repaired gene. To clarify this point, we first determined RNA polymerase II (Pol II) occupancy by ChIP analysis at the GFP promoter, the translation start site, the DSB region and the 3′ polyA site (Supplemental Fig. S1A). Pol II concentrations were highest at the GFP promoter and polyA sites in non-recombinant or Rec H cells. Pol II occupancy was reduced at all sites in Rec L cells relative to UnRec or Rec H cells. Second, we inhibited transcription with low doses of actinomycin D and monitored formation of loops A and C. Actinomycin D reduced the abundance of loop A in both Rec H and Rec L cells and loop *C* in Rec L cells (Fig. 3C, panels 2 and 4), suggesting that both loops are either generated or maintained by transcription. The results of these experiments are summarized in Fig. 3D.

In summary, DNA repair and associated transcription permanently modify the structure of local chromatin, generating chromatin loops that juxtapose the 3′ end of the transcribed gene with various 5′ sites. These structures reflect the transcriptional status of the gene and are influenced by local DNA methylation.

BER enzymes remodel DNA methylation soon after HR. We previously reported that GFP DNA methylation in HR cells was modified by transcription after repair. Thus, inhibiting transcription with short pulses of actinomycin D shortly after HR permanently increased DNA methylation of the repaired GFP gene[11]. We hypothesized that transcription was associated with active demethylation[14,15]. To test this idea, we focused on BER enzymes, which promote DNA oxidation and cytosine demethylation during transcription[16–18]. We first asked if APE1, the BER apurinic site nuclease, was recruited to GFP chromatin before or after repair. Indeed, APE1 was enriched mainly at the promoter site of all cell types, but was specifically enriched at the 3′ end of the DSB in Rec H and Rec L cells (Supplemental Fig. S1B). We note that APE1 was present at high levels along the entire gene in Rec H cells, suggesting a role for APE1 in the transcription of the repaired gene (Supplemental Fig. S1B). We have previously shown that OGG1, the 8-oxoG glycosylase and APE1 are important in Myc[17], estrogen-[16] and retinoic acid-[18] induced transcription. We hypothesized that BER enzymes may control the rate
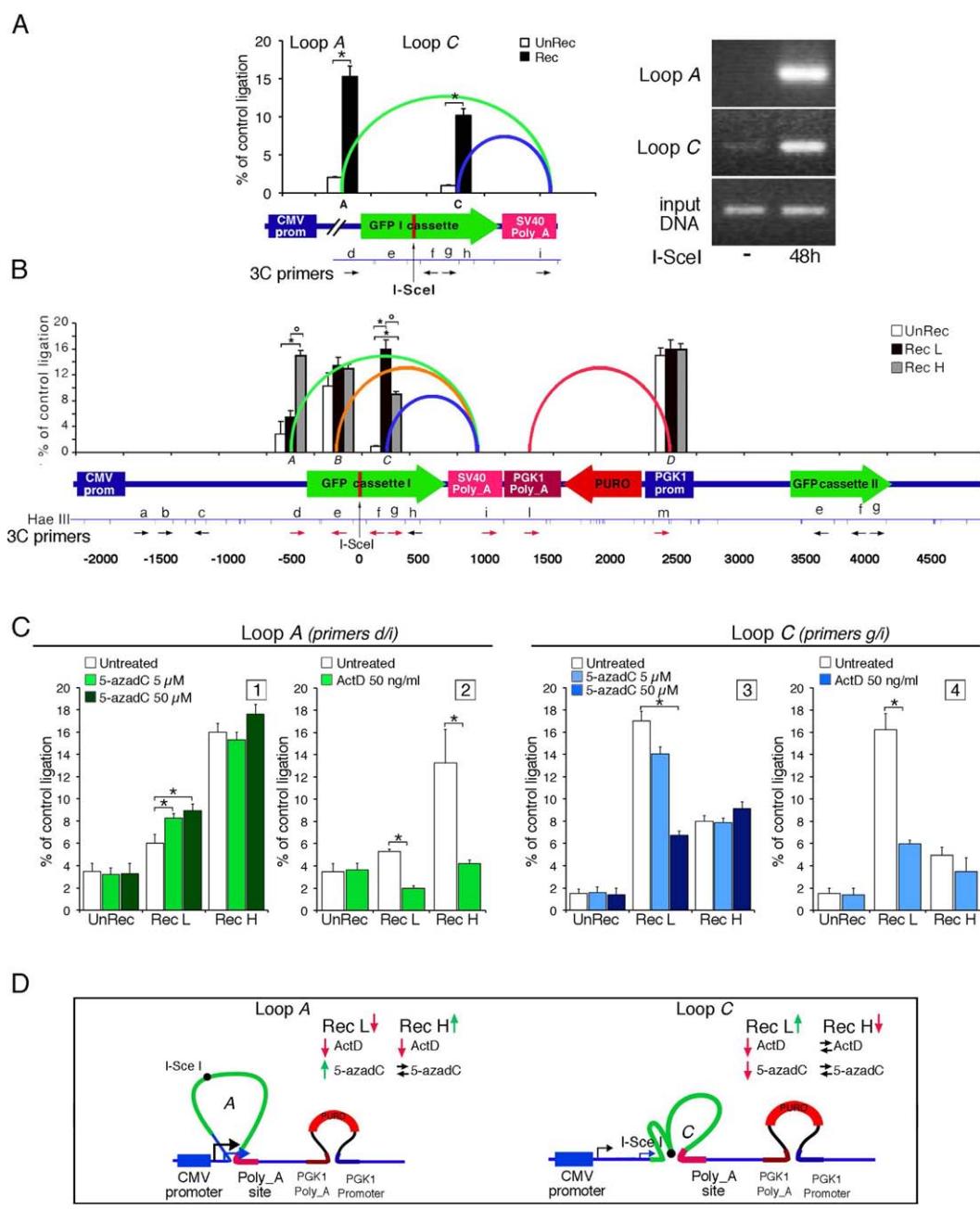
Figure 3. Chromatin-DNA domains induced by damage-repair. (**A**) **Left.** GFP chromatin loops in recombinant (Rec, black) and unrecombinant (UnRec, white) cells 7 days after DSB. qPCR in 3C, performed with the primers indicated (HaeIII map), show the frequency of ligation (mean ± SD) of the DRGFP HaeIII fragments amplified with the specific primers. The ends of the loops are shown by the green (loop A) and blue (loop C) lines, corresponding to the regions indicated by arrows and boxes. Ligation efficiency is relative to the DRGFP plasmid digested with HaeIII, ligated and amplified by qPCR. Distance, in bp, is relative to *I-SceI* site (vertical arrow), *p < 0.01 (*t test*) Rec *vs* UnRec. **Right.** A representative gel of the ligated fragments (loops A and C) in cells exposed or not for 48 h to I-SceI. (**B**) GFP loops in Rec H and Rec L cells. The 5′ end of the loop A (green) includes an alternative GFP transcription start site, identified by primer *d*. Differences between non-recombinant, H or L cells *p < 0.01 (*t test*); °p < 0.01 (*t test*) H vs L cells. (**C**) Inhibition of methylation (panels 1, 3) or transcription (panels 2, 4) alters the chromatin loops induced by HR. **Panels 1 and 3.** Sorted GFP⁻ (UnRec) or Rec H or Rec L cells were exposed to 5 or 50 μM 5-azadC for 24 h + 24 h in standard medium. Loops A and C were monitored as described above. The results derive from at least 3 experiments in triplicate. *p < 0.01 as compared to untreated samples (*Wilcoxon rank-sum test*). **Panels 2 and 4.** Sorted GFP⁻ (UnRec) or Rec H or Rec L cells were exposed to 5 μM actinomycin D for 24 h, washed and analysed 48 h later. *p < 0.01 (*Wilcoxon rank-sum test*) compared to untreated cells. (**D**) Model summarizing the features of Loops A and C. The red and green arrows indicate decrease or increase of loop formation, respectively, while the black arrows indicate no change. The two GFP transcription start sites are indicated by black and blue arrows.
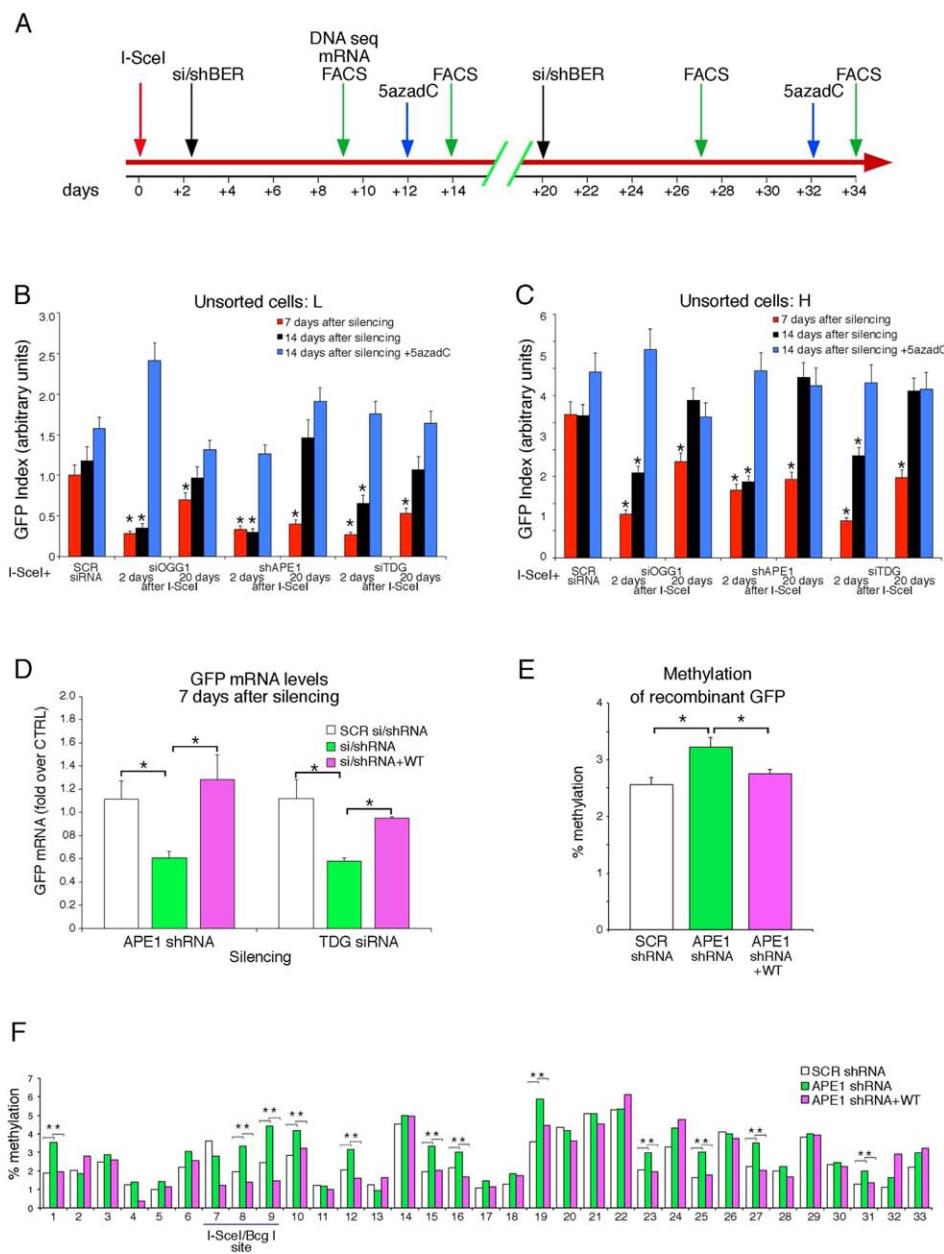
Figure 4. Inhibition of BER early after repair reduces transcription and increases methylation of the repaired DNA. (**A**) Chronology of BER-silencing experiments. Time 0 indicates I-SceI transfection and arrows the time of analysis or treatments. GFP+ cells were 10% ± 2% in all treatments. At days 12 and 32, 5-azadC (10 μg/ml, blue arrows) was added for 24 h, removed and 24 h later cells were analyzed. (**B,C**) Analysis was performed at day 7 (red) or 14 (black) after treatments. GFP index is the product of GFP intensity and reciprocal cell fraction in the fluorescence gate to normalize frequency of GFP+ cells in H and L gates to intensity of signal, to compare different experiments. The data shown derive from 20 independent experiments. (**B,C**) Show GFP index in Rec H and Rec L cells, respectively. *p < 0.01 (*Matched t test*) compared to scrambled control. (Supplemental Fig. S3 and Table S2). (**D**) GFP mRNA levels 7 days after APE1 and TDG-silencing by qPCR for recombinant GFP. TDG protein levels are in Supplemental Fig. S3C. *P < 0.01 (*Wilcoxon rank-sum test*) compared to scrambled control. (**E,F**) Methylation analysis of GFP in mass cultures of I-SceI-transfected cells, in which APE1 levels were modified after HR (48 h after I-SceI and analyzed 7 days later, panel A). DNA was subjected to bisulfite analysis and sequenced with Myseq Illumina (Supplemental Table S3). Panel E shows the average methylation of the recombinant GFP. All the cells were exposed to I-SceI and these vectors: SCR shRNA (white shaded) scrambled shRNA; APE1 shRNA (green) shRNA APE1; APE1 shRNA + WT (purple) shRNA APE1 and APE1 expression vector. The percent of methylation of recombinant GFP in all samples is normalized to the recombinant GFP cassette as shown in Supplemental Table S3; data were expressed as the mean ± SEM. *p < 0.01 (*t test*) compared APE1 shRNA *vs* SCR shRNA or *vs* APE1 shRNA + WT. Panel F shows the percent of CpG methylation in recombinant GFP in cells in which the levels of APE1 were modified early after repair. The position of the *I-SceI/BcgI* site corresponds to CpG 7 and 8. *p < 0.01 (*Pearson's chi-squared test*) compared APE1 shRNA *vs* SCR shRNA (+SceI) or *vs* APE1 shRNA + WT.

100

of cytosine demethylation during transcription of the repaired gene in a precise time frame after repair. To confirm and define this critical period necessary to establish permanent DNA methylation changes, we depleted BER enzymes at 2 or 20 days after DSB formation (Fig. 4A). We selectively inhibited APE1 and two other BER enzymes, OGG1, and TDG (Supplemental Figs S2 and S3). TDG has also been directly implicated in active DNA demethylation[19–21].

The effects on GFP expression were evaluated beginning 7 and 14 days after silencing, when the mRNA and protein levels expressed by the targeted genes had returned to normal (Supplemental Fig. S2C). The timing of the knock-downs relative to the formation of the DSBs is shown in Fig. 4A. Figure 4B,C display the effects of silencing OGG1, APE1 and TDG on GFP expression as assayed by cytofluorimetry. GFP expression is represented by a GFP index, which takes into account the distribution of Rec H and Rec L cell populations (% of Rec H and Rec L peaks) and the fluorescence intensity (see legend of Fig. 4). Silencing was initiated at 2 or 20 days after DSB formation and Rec H and Rec L peak percentages and fluorescence intensity were measured. The color code of the histograms shown in Fig. 4B,C indicates the time of analysis after initiation of silencing (7 and 14 days, red bars and black bars, respectively), and the effects of treating the samples with 5-azadC on GFP expression (blue bars). Since HR is complete by 2 days following exposure to I-SceI, BER depletion at or after this time had no effect on the frequency of GFP$^+$ cells (Supplemental Figs S2 and S3), although it significantly altered the levels of GFP expression (Fig. 4B–E).

Figure 4B,C show that silencing of BER enzymes early after HR permanently inhibited GFP expression in both Rec H and Rec L cell populations, even when the concentration of the depleted proteins returned to pre-treatment levels (Supplemental Fig. S2C). Inhibition of GFP expression was due to specific reduction of APE1 and TDG mRNA, since it was reversed with plasmids expressing APE1 and TDG (Fig. 4D). APE1 silencing resulted in an elevation of DNA methylation (Fig. 4E) and consistent with this observation, 5-azadC restored GFP expression to normal levels in BER-depleted cells (Fig. 4B,C, blue bars). Depletion of the BER enzymes 20 days after DSB formation did not modify GFP methylation or expression (Fig. 4B,C). To gain insight into the methylation status of repaired GFP in cells in which APE1 levels were manipulated early after repair, we performed deep sequencing of bisulfite-treated DNA derived from of mass cultures (data set are available at Figshare; DOI:10.6084/m9.figshare.3470099). We analysed at least 8,000 GFP molecules/samples (Supplemental Table S3) and measured total methylation of GFP in cells in which APE1 levels were down-regulated after repair. Total methylation did not differ significantly between recombinant and unrecombinant or control cells, because HR cells include both hypo- (H) and hyper-methylated (L) clones. However, GFP methylation increased in cells in which APE1 had been depleted early after repair. Reconstitution of APE1 eliminated this increase and restored methylation to control (scrambled shRNA) levels (Fig. 4E).

We next analysed the methylation status of individual CpGs in the HR region. We found that APE1 levels significantly affected the methylation pattern.

The percentage of methylation of each CpG, (numbered from 1 to 33), is tabulated within the *I-SceI* region. APE1 depletion stimulated methylation at a subset of these sites 1, 8, 9, 10, 12, 15, 16, 19, 23, 25, 27, 31. The methylation gain at the majority of these sites was eliminated when APE1 levels were restored by expressing the wild-type protein from an expression vector (*).

In conclusion, the data shown in Fig. 4F provide an extensive window on the methylation changes that follow repair. Several sites, which we call seeds, are preferentially methylated and demethylated after HR, suggesting that the DNA methylation status early after repair is subject to extensive remodelling by transcription and BER-associated demethylation.

Discrete DNA methylation patterns mark clones with distinct GFP gene expression levels.
DNA methylation status is highly polymorphic and can be reshaped during and after DNA damage-repair events. Over time, the DNA methylation profiles of Rec H and Rec L cells stabilize and generate cells with different but heritable GFP expression levels. These clones are characterized by specific GFP chromatin domain patterns (Figs 1,2 and 3).

To relate the extent and placement of DNA methylation with gene expression, we compared the location and the number of methylated CpGs (mCpGs) with GFP expression levels in the most frequently modified GFP molecules isolated from sorted Rec H and Rec L cells. We ordered the methylated GFP molecules into families that share mCpGs at identical locations to define epigenetic haplotypes. Specifically, we asked if molecules with the same number of mCpGs, but located at different sites in the gene (i.e., different haplotypes), expressed similar levels of GFP. We also included in this analysis the most frequent UnRec molecules to reveal potential relationships between GFP molecules present before and after HR. Figure 5A shows the similarity of the repaired GFP molecules on the basis of the position of mCpGs and GFP expression levels. There are two main branches in the tree: the first, indicated as I, includes essentially L clones with 10 or more mCpGs; the second, indicated as II and III, contains GFP molecules with intermediate frequencies of mCpGs (from 3 to 7) and includes both H and L clones. The arrows shown in the cartoon below the tree indicate that the L and H clones in these groups are very similar in terms of mCpG content. The clones with intermediate levels of mCpGs contain the same number of mCpGs but carry them in different locations. This reveals that the location of the mCpGs, rather than their absolute frequency is critical for GFP expression. We have mapped the mCpGs that characterize Rec H (green) and Rec L (red) clones. We find that some positions (for example mCpGs 9–10) are specific to Rec H clones, while others characterize Rec L clones (mCpGs 17–20; Fig. 5B). These mCpGs that characterize H and L clones were detected in a large unsorted pool of GFP$^+$ molecules isolated from a mass culture of cells exposed to I-SceI (Fig. 4F). These mCpGs are stable over time and can be recovered with the same frequency after three years of continuous culture[8,11].
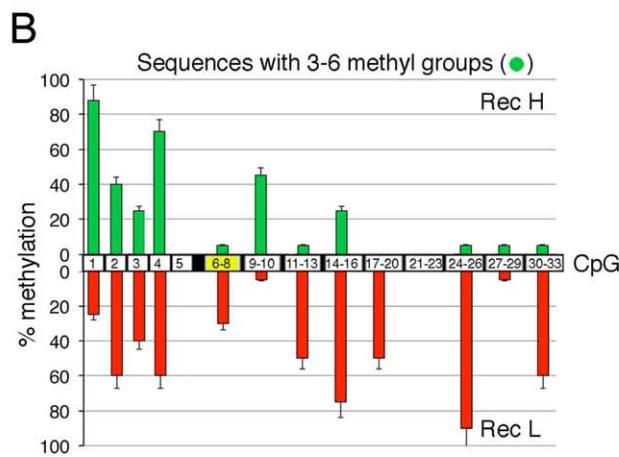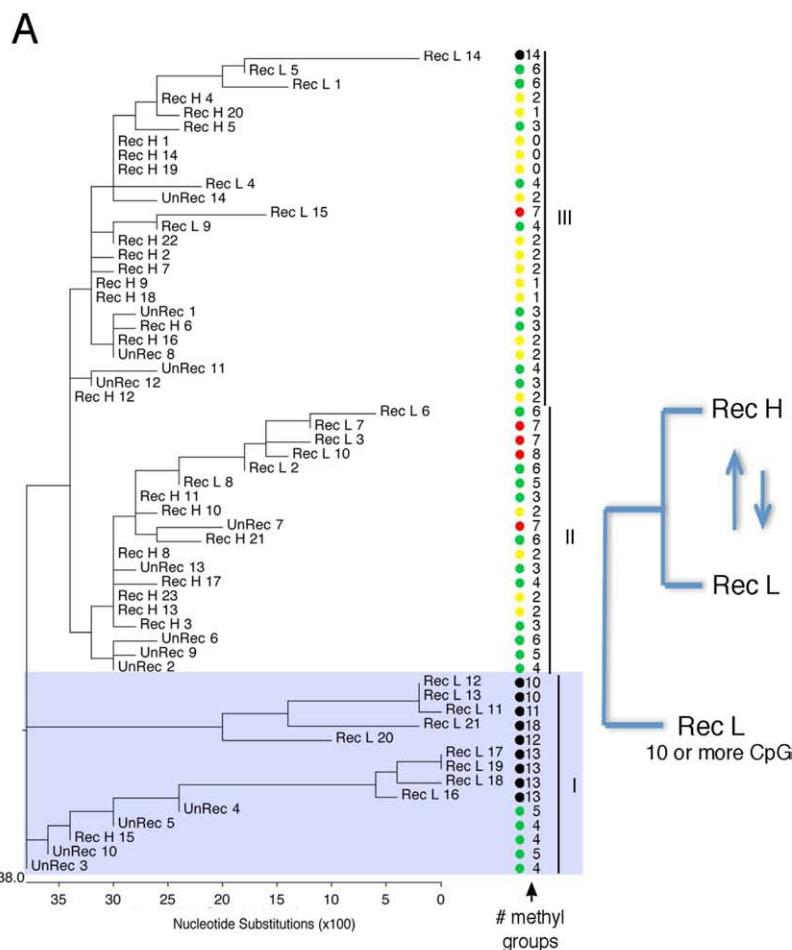
Figure 5. Methylated GFP molecules are polymorphic. (A) Qualitative analysis of GFP methylation profiles in UnRec, Rec H and Rec L clones. We compared the location of methylated CpGs in the most abundant GFP molecules (above 5–10%) derived from recombinant (Rec H and Rec L) and non-recombinant (UnRec,) cells. The sequence at the *I-SceI* site in UnRec molecules was edited to *BcgI* to eliminate the differences in the sequence between Rec and UnRec molecules and to permit the comparison of Rec and UnRec GFP molecules only on the basis of methylation. Cluster analysis (ClustalW) shows three main families of methylated molecules: I, represented essentially by Rec L clones; II and III represented by Rec H and Rec L clones. Colored circles indicate the number of methylated CpGs/molecule. **(B)** Specific CpGs are methylated in Rec H and Rec L clones following HR. Molecules containing 3 to 6 mCpGs were sorted from Rec H or Rec L cells pools and compared. The location of methylated CpGs at the 5′ and 3′ ends of the DSB is shown relative to the DSB (black-yellow box centered at the 6–8 CpG). Methylation of CpG from 1 to 5 is not modified by DSB or HR (ref. 11). The histograms show the percentage of methylation of the specific CpGs in Rec H (green) and Rec L (red) clones.
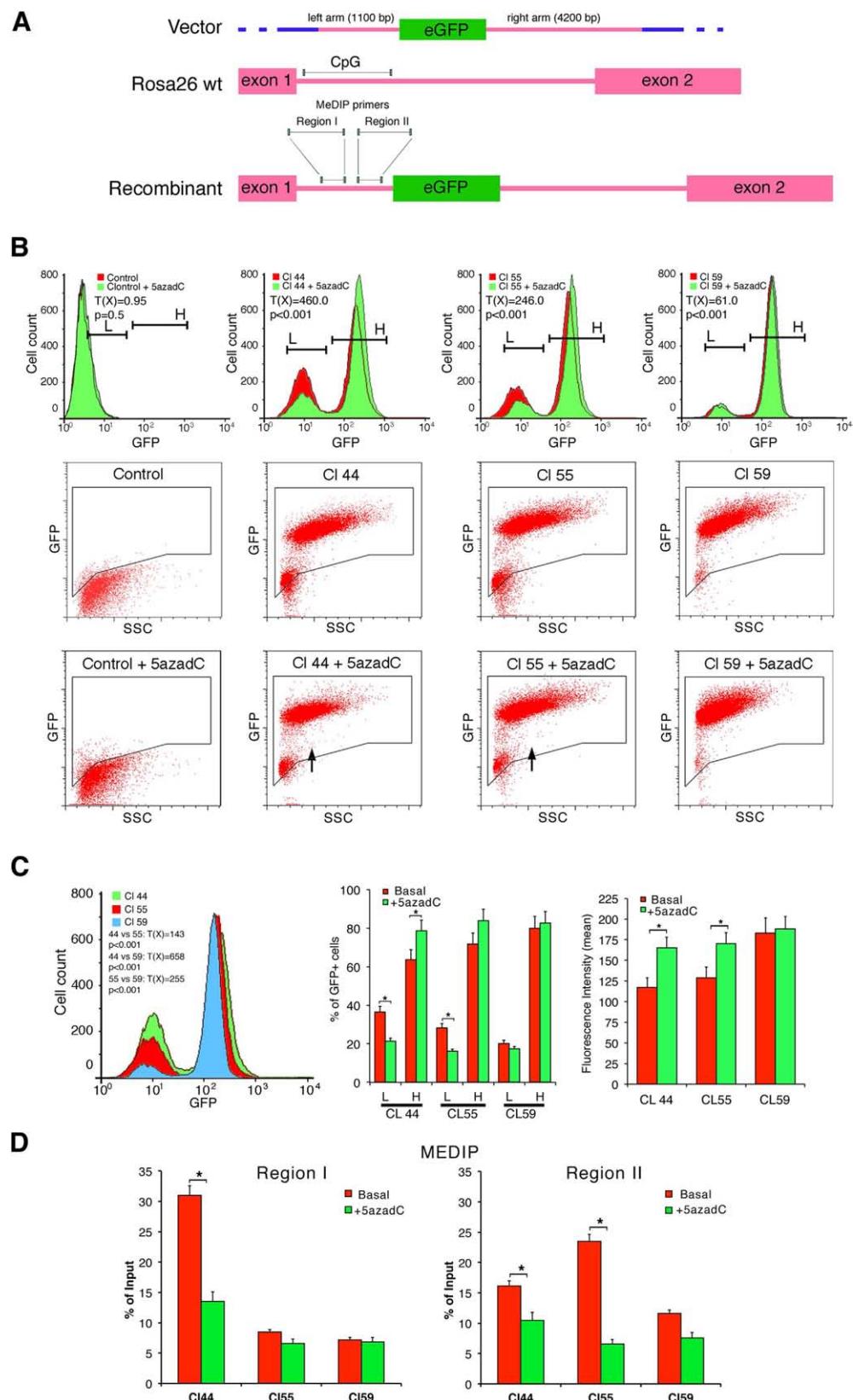
Figure 6. Homologous targeting of GFP to the mouse Rosa26 locus generates ES clones with variable DNA methylation and GFP expression. Three independent ES clones in which the EGFP gene was targeted to the mouse Rosa26 locus, (kindly provided by A. Simeone and D. Acampora, Intern. Inst. Gen. Biophys., IGB, Naples, Italy). These clones, containing a single copy integrated GFP, were purified 7 days after transfection with the targeting vector, amplified and analyzed by cytofluorimetry as described in Methods. The same clones were exposed to 5-azadC ($0.5\,\mu$M) for 4 days and analyzed 48 h later. Panel (A) shows: 1. the structure of the targeting vector; the two homologous regions are shown in red; 2. the structure of the mouse Rosa26 locus and; 3. the

<div align="center">103</div>

structure of the targeted locus. A segment at the 5′ end, containing a CpG island, is shown (lines). The primers for MEDIP analysis are located in regions I and II. Panel (B) shows the cytofluorimetric analysis of the 3 clones exposed or not to $0.5\,\mu M$ 5-azadC for 4 days and analyzed 48 h later. Dot Plot scans are shown to illustrate the composition of GFP+ or GFP− cells. The arrows indicate the shift of the L population after 5-azadC treatment. Differences in GFP expression between control and 5-azadC treated cells were tested for statistical significance using the Chi Square test, T(X), (Population Comparison module of the FlowJo software from Tree Star). Cl 44, untreated vs 5-azadC T(X) = 460, p > 0.001; Cl 55, untreated vs 5-azadC T(X) = 246, p > 0.001; Cl 59, untreated vs 5-azadC T(X) = 61, p < 0.001. (C) The panel on the left shows the overlapping profiles of the three clones without treatment to compare the relative GFP expression levels. Cl 44 vs Cl55 T(X) = 143, p < 0.001; Cl 44 vs Cl59 T(X) = 658, p > 0.001; Cl 55 vs Cl59 T(X) = 255, p > 0.001. The central and the left panels show the quantitative analysis of GFP expression in Rec H and Rec L cells before or after 5-azadC treatment as % of GFP+ cells and mean of fluorescence intensity. In the left panel, Rec H and Rec L clones were analyzed together. Differences between treatments were tested for statistical significance using matched pairs t test: *p < 0.001. Panel (D) shows MEDIP analysis of region I and II, respectively in the 3 clones. *p < 0.001.

Homologous targeting of GFP in ES cells also generates clones with various levels of GFP expression and DNA methylation.

The data presented above indicate that HR repair of a DSB changes the methylation pattern of the repaired segment. In these experiments the DSB was artificially generated by the I-SceI meganuclease. To study DSBs, created by a different mechanism, we targeted CMV-GFP to a mouse DNA locus (Rosa26) by homologous recombination, which requires formation and repair of a DSB (Fig. 6A). Based on the *I-SceI* data, we predicted that the expression level of the inserted gene would differ in genetically identical clones. We isolated 3 mouse ES clones carrying a single copy of CMV-GFP targeted to the Rosa26 locus (A. Simeone and D. Acampora, unpublished observations and Fig. 6A). These clones, 44, 55 and 59, were characterized for GFP expression. Clones 44 and 55 contain two populations that differ in GFP expression levels, whereas clone 59 contains primarily cells that express high GFP levels (Fig. 6B). We propose that the distinct progeny of clones 44 and 55 are equivalent to the hypomethylated Rec H and hypermethylated Rec L clones found in DRGFP HeLa cells, and that clone 59 only generated Rec H progeny. To test this hypothesis, we treated the cells with 5-azadC and measured GFP expression. Exposure to 5-azadC shifted Rec H cells to the right (higher expression) and reduced the number of Rec L cells in clones 44 and 55. Demethylation slightly affected GFP expression in derivatives of clone 59 (Fig. 6B,C).

We also determined the DNA methylation status of the CpG island at the 5′ end of the homologous targeting sequence with primers specific to sub-regions I and II (Fig. 6A). MEDIP analysis showed that region I in clone 44 was hypermethylated compared to clones 55 and 59. Region II in clone 55, and to a lesser extent in clone 44, was hypermethylated compared to clone 59, which is not methylated (Fig. 6D). The differences in mCpG content were largely reversed by treatment with 5-azadC (Fig. 6D). On the basis of the *I-SceI* data. we hypothesize that the methylated subregion is located at the 3′ end of the DSB, which initiates HR along the direction of transcription. We propose that the DSB occurring during homologous pairing upstream of region I generated clone 44 and that a DSB between regions I and II generated clone 55. Clone 59 is equivalent to high expressor clones found in GFP+ cells.

These data extend the notion of HR-induced methylation and suggest a general mechanism that modifies expression of targeted genes by homologous recombination.

Discussion

The data reported here shed light on somatic DNA methylation and consequent histone modification induced by damage and homologous repair. They suggest that cell-to-cell variations in gene expression are dependent on the different DNA methylation profiles and chromatin structures of the expressed gene acquired during or soon after HR.

Transient and stable cis and trans chromatin changes induced by DNA damage and repair. Concurrent with DSB formation by I-SceI and repair by HR, chromatin near the lesion becomes enriched with the repressive chromatin mark, H3K9m2/3. This modification has been reported to be essential to recruit other histone-modifying enzymes and ATM to the site of damage[22]. The I-SceI-treated cells also transiently lose H3K4 methylation, but restore the H3K4m2/3 mark 2 to 7 days after exposure. The rapid appearance of H3K9me2/3 and loss of H3K4me3 at the DSB region after I-SceI exposure (Fig. 1B, panels 6, 10) suggest that these changes are induced by the formation of DSBs. Subsequent purification reveals that H3K9m2/3 is selectively retained after HR only in Rec L clones (Fig. 2A,B). Maintenance of H3K4m2/3 is secondary to DNA methylation. Thus, treatment of Rec L cells with the DNA demethylating agent, 5-azadC, significantly reduced the levels of H3K9m2/3 on GFP chromatin (Fig. 2C).

De novo methylation of the repaired segment was also responsible for stabilization of the chromatin loop specific to Rec L cells (loop *C* in Fig. 3). Repressive methylation H3K9m2/3 marks in Rec L cells were also present at sites physically distant to the repaired DSB. However, these sites were, in fact, juxtaposed and linked by loop *C* (Fig. 3). At other physically distant sites (e.g., the puromycin-resistance gene), histone marks were not modified by damage and repair (data not shown). Note that a similar series of events occurs during transposon integration and silencing. Chromatin repressive marks (H3K9me2/3) are induced early during integration of the transposable element (TE) followed by methylation of the integrated segment[23].

We propose that the initial functionally relevant event in HR-directed gene modification is the formation of H3K9me3 at the DSB. This modification is induced by the DSB or by NHEJ in the majority of cells. Histone methylation is carried out by the histone methyltransferase SUV39, which is recruited to the DSB[11,24,25]. The increase of H3K9me3 contributes to the repression of local transcription induced by DNA damage[26,27]. The H3K9me2/3 at

<div align="center">104</div>

the DSB is progressively lost after repair except in Rec L cells, and to a lesser extent in Rec H clones (see Fig. 2B, panel 5). Rec L and Rec H clones are characterized by discrete chromatin loops (Fig. 3) and specific H3K4me2/3 and H3K9me2/3 profiles along the GFP gene (Fig. 2B). These marks appear very early after DSB formation in unsorted mass culture and precede stabilization (Fig. 3A). The most striking feature of the HR cells is the stability and the inheritance of the chromatin and DNA changes at the repaired locus. Our data also indicate that the initial *de novo* methylation of the repaired segment can be revised. Early after exposure to I-SceI, the ratio of Rec H to Rec L clones was approximately 1:1. After 7–14 days, the ratio changed to ≃4:1, depending on the clone[8]. The ratio was stabilized permanently by day 21. The repressive chromatin marks (H3K9me2/3) were maintained only in Rec L cells, suggesting that the DNA methylation profiles after repair stabilize repressive chromatin markers (Fig. 2C).

BER enzymes reshape stable methylation. Overall, the final methylation status of the repaired gene is polymorphic (Figs 4F and 5A). We find that depletion of BER enzymes two days after DSB formation modified permanently the expression of repaired GFP (Fig. 4 and Supplemental Figs S2 and S3). We, and others, have reported that recruitment of BER and NER enzymes to promoter sites of several nuclear hormones or HIF-induced genes is essential for transcription[16–18,28,29]. Additionally, the BER enzyme, TDG, which suppressed GFP DNA methylation after DSB repair (Supplemental Fig. S2B,C) is involved in active CpG de-methylation during transcription[19–21]. These data and those shown in Fig. 4 and Supplemental Figs S2 and S3 indicate that other BER enzymes, (e.g., OGG1 and APE1), which recognize and process oxidized G, are also involved in transcription-associated demethylation. Indeed, although the absolute methylation levels of the repaired segments do not change dramatically, depletion of BER enzymes significantly changes the qualitative methylation profiles of the repaired segment, 3′ to the *I-SceI* site (Fig. 4). We propose, therefore, that BER enzymes perform transcription-associated demethylation at GFP at the repaired locus. This process is partly stochastic with respect to a particular CpG, since some CpGs are preferentially methylated (Figs 4F and 5B,C). These may seed further methylation along the repaired gene. Methylation revision generates polymorphic methylation profiles, which influence GFP expression depending on the location of the methylated CpG (Fig. 5). We wish to stress that CpG preferentially methylated in sorted H or L clones were also identified in mass culture sequencing (Figs 4F and 5B).

Polymorphism of somatic methylation profiles. Comparing the levels of expression of GFP with the methylation profiles of single DNA molecules, we find a relationship between the profile and the expression of the repaired gene. We describe clones carrying the same number of mCpGs (Fig. 5), but whose GFP enzyme levels vary dramatically. The differences of GFP expression between Rec H and Rec L clones are reduced or erased by 5-azadC treatment, indicating that the variations in expression reflect the methylation status of specific CpGs.

In conclusion, we propose that the variability of GFP expression is affected by editing of local methylation by transcription and active demethylation (Fig. 4; ref. 19), which together with DNA damage and repair represent a major source of polymorphism of methylation in somatic cells.

Is DNA methylation in somatic cells deterministic or stochastic? It was recently reported that methylation of INK4-ARF suppressor gene is induced by a specific Ki-Ras oncogene transcriptional program[30]. This observation supports a general deterministic model for DNA methylation in which locus-specific targeting of DNMT enzymes induces and maintains DNA methylation. The choice of target is not random, but determined by specific affinities of transcription factors and chromatin modellers. Eventually, the preference of DNMT1 for hemi-methylated DNA stabilizes the methylation profiles. This deterministic model may rationalize clustering of methylated sites in the same DNA region, but fails short of explaining the extreme polymorphism of methylated alleles found with a deeper sequence coverage of the genome (Fig 5 and ref. 12).

Our data suggest that both deterministic and stochastic factors govern stable DNA methylation profiles. In the system described here, we can quantify the deterministic and the stochastic factors that contribute to the final methylation status of GFP epialleles. HR and specific factors, recruited to the DSB, establish the location and the strand that will be methylated (Figs 2D and 5 and refs 10,11). In contrast to this deterministic modification of the epigenome, stochastic editing of methylation by transcription and BER enzymes generates polymorphism of the methylated GFP alleles.

We propose, therefore that both mechanisms contribute to the final methylation status of DNA in each cell. Importantly, our work impacts in the area of genome editing, which is largely driven by HR. The final penetrance of a repaired gene will be directed by the events related to DNA methylation revision described here.

Methods

Cell culture, transfections and plasmids. HeLa cells lines were cultured at 37 °C in 5% CO_2 in RPMI medium supplemented with 10% fetal bovine serum (Invitrogen), 1% penicillin-streptomycin, and 2 mM glutamine. HeLa-DR-GFP cells were obtained by stable transfection of HeLa cells with the pDR-GFP plasmid as described in ref. 11. We used the same conditions of growth (~40% confluent cells starting from freshly frozen aliquots). The structure of the pDR-GFP and other plasmids is described in Supplemental Information. The expression vectors for OGG1WT and for the K338R/K341R OGG1 mutant were the FLG-Tagged vectors previously described in ref. 31.

DNA extraction and qRT-PCR and qPCR. Genomic DNA extraction was performed as described in Supplemental Information. Reverse Transcription Polymerase Chain Reactions (qRT-PCR) and Quantitative Polymerase Chain Reactions (qPCR) were performed on a 7500 Real Time PCR System (Applied Biosystems) using the SYBR Green-detection system (FS Universal SYBR Green MasterRox/Roche Applied Science). The complete list of oligonucleotides is reported in the Supplemental Table S1.

FACS analysis. For the FACS analysis, HeLa-DR-GFP cells were harvested and re-suspended in 500 μl of PBS at density of 10^6 cells/ml. Cell viability was assessed by Propidium Iodide (PI) staining: before FACS analysis cells

were incubated with $3\,\mu$M PI for 10 min. Cytofluorimetric analysis was performed on a 9600 Cyan System (Dako Cytometrix), PI positive cells were excluded from the analysis by gating the PI-negative cells on a FSC-Linear vs FL2H-Log plot; GFP + cells were identified with a gate (R1) on a FL1H-Log vs SS-Log plot. Rec L and Rec H cells were identified on a FL1H Histogram of the R1-gated cells with 2 range-gate (see Fig. 4A). The same gate was used for all flow cytometry experiments.

Population comparison was performed using the FlowJo software (Chi-Squared Test). Differences in fluorescence intensity (mean) were determined using the matched pairs Student's t test.

Bisulfite DNA preparation, PCR, and sequence analysis.   (The detailed protocol is described in the Supplemental Methods). A total of $2\,\mu$g of genomic DNA was bisulfite_converted according to the EZ DNA Methylation Kit (Zymo Research). Methylation status was assessed through a strategy based on the locus- specific amplification of bisulfite-treated genomic DNA, amplifying each amplicon separately, followed by Illumina MiSeq sequencing. The sequence of the bisulfite-specific primers used for this analysis is reported Supplemental Table S1. The methods involved two PCR steps, following Illumina recommended procedure. The pool of amplicons was subjected to sequencing using MiSeq system (V3 reagents kits). Sequencing was performed by 281 cycles (paired-end sequencing). Sequences in FASTQ format by Illumina sequencing machine were initially processed with Paired-End reAd mergeR (PEAR) data for an initial quality filtering and assembling (R1 plus R2). Only those sequences with a threshold quality score of ≥ 30 and an overlapping region within paired-end reads of 40 nt were processed with PReprocessing and INformation of SEQuence (Prinseq) to obtain FASTA for further analysis. Reads were aligned to the bisulfite converted reference sequence. Reads with ambiguous calls at the CpG dinucleotide were removed. After filtering, an average of 25,607.875 (range: 2,389–81,923) amplicon reads were obtained from each sample. Methylation states were estimated by counting the number of base calls (T/C) at CpG sites in the mapped reads compared to the referent on both strands (% methylation).

Chromatin Immune-Precipitation (ChIP).   Cells were transfected and/or treated as indicated in the legends of the figures. The cells ($\sim 2.5 \times 10^6$ for each antibody) were fixed for 10 min at room temperature by adding 1 volume of 2% formaldehyde to a final concentration of 1%; the reaction was quenched by addition of glycine to a final concentration of 125 mM. Fixed cells were harvested and the pellet was re-suspended in 1 ml of lysis Buffer containing $1\times$ protease inhibitor cocktail (Roche Applied Science). The lysates were sonicated to have DNA fragments 300 to 600 bp. Sonicated samples were immunoprecipitated as described in Supplemental Information and the DNA was recovered and subjected to qPCR using the primers indicated in the legend of the specific figures, primers sequences are described in Supplemental Table S1.

MeDIP.   Cells were transfected and/or treated as indicated in figure legends. A total of $\sim 5 \times 10^6$ cells were harvested and Genomic DNA extracted as described above. Ten micrograms of total genomic DNA was digested in $200\,\mu$l for 16 h with Restriction Endonuclease mix containing 30 U each of Eco RI, Bam HI, Hind III, XbaI, Sal I (Roche Applied Science), phenol/chloroform extracted, ethanol precipitated and resuspended in $50\,\mu$l of TE buffer. An aliquot (1/10) of digested DNA was used as input control to determine DNA concentration and digestion efficiency. MeDIP was performed essentially as described[32] except that $2\,\mu$g of antibody specific for 5 mC (Abcam) were used to precipitate methylated DNA from $5\,\mu$g of total genomic DNA.

Chromosome conformation capture (3C).   The 3C assay was performed as described previously[18,33] with minor modifications. Briefly, the Hae III restriction enzyme, which cleaves pDR-GFP and generates 55 fragments was used. Digestion was performed on formaldehyde-fixed nuclei with 150 U of restriction enzyme at 37 °C for 16 h. The restriction enzyme was inactivated by addition of SDS to 2% and incubation at 65 °C for 30 min. The reaction was diluted into 4 ml ligation buffer containing 50 U of T4 DNA Ligase and incubated at 16 °C for 18 h. Samples were de-crosslinked by incubation at 65 °C in the presence of proteinase K for 15 min, purified by phenol/chloroform extraction and ethanol precipitated. Samples were re-dissolved in $20\,\mu$l of TE buffer. Primer sequences are shown in Supplemental Table S1.

Statistical analysis.   All data (with exception of Fig. 4E) are presented as mean ± standard deviation in at least three experiments in triplicate ($n \geq 9$). Data sets were analyzed statistically using *JMP Statistical Discovery*™ software by SAS and tested for normality using the Shapiro-Wilks test ("normal distribution fit" tool -JMP software). Two-tailed significance tests were performed with $p < 0.05$ considered significant. Non-parametric analyses were done with the Mann-Whitney-U-Test (Wilcoxon rank-sum test), parametric with the t-test. Detailed statistical analysis of the data shown in Figs 1 and 2 is reported in Supplemental Statistical Tables 1, 2, 3 and 4.

References

1. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biology* **14,** R115 (2013).
2. Huh, I., Zeng, J., Park, T. & Yi, S. V. DNA methylation and transcriptional noise. *Epigenetics Chromatin* **6,** 9 (2013).
3. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366,** 362–365 (1993).
4. Fu, A. Q., Genereux, D. P., Stöger, R., Laird, C. D. & Stephens, M. Statistical inference of transmission fidelity of DNA methylation patterns over somatic cell divisions in mammals. Ann *Appl Stat* **4,** 871–892 (2010).
5. Baylin, S. B. & Ohm, J. E. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? *Nat Rev Cancer* **6,** 107–116 (2006).
6. Jasin, M. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet* **12,** 224–228 (1996).
7. Pierce, A. J., Johnson, R. D., Thompson, L. H. & Jasin, M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev* **13,** 2633–2638 (1999).
8. Cuozzo, C. *et al.* DNA damage, homology-directed repair, and DNA methylation. *PLoS Genet* **3,** e110 (2007).
9. O'Hagan, H. M., Mohammad, H. P. & Baylin, S. B. Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. *PLoS Genet* **4,** e1000155 (2008).

106

10. Lee, G. E., Kim, J. H., Taylor, M. & Muller, M. T. DNA methyltransferase 1-associated protein (DMAP1) is a co-repressor that stimulates DNA methylation globally and locally at sites of double strand break repair. *J Biol Chem* **285,** 37630–37640 (2010).
11. Morano, A. *et al.* Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene. *Nucleic Acids Res*. **42,** 804–821 (2014).
12. Landan, G. *et al.* Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat Genet*. **44,** 1207–1214 (2012).
13. Golding, S. E. *et al.* Double strand break repair by homologous recombination is regulated by cell cycle-independent signaling via ATM in human glioma cells. *J Biol Chem*. **279,** 15402–15410 (2004)
14. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473,** 343–348 (2011).
15. Müller, U., Bauer, C., Siegl, M., Rottach, A. & Leonhardt, H. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res*. **42,** 8592–8604 (2014).
16. Perillo, B. *et al.* DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. *Science* **319,** 202–206.16 (2008).
17. Amente, S., Bertoni, A., Morano, A., Lania, L., Avvedimento, E. V. & Majello, B. LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc-induced transcription. *Oncogene* **29,** 3691–3702 (2010).
18. Zuchegna, C. *et al.* Mechanism of retinoic acid-induced transcription: histone code, DNA oxidation and formation of chromatin loops. *Nucleic Acids Res*. **42,** 11040–11055 (2014).
19. Wu, S. C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* **11,** 607–620 (2010).
20. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502,** 472–479 (2013).
21. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair *Cell* **146,** 67–79 (2011)
22. Rebollo, R. *et al.* Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet. Sep* **7(9),** e1002301 (2011).
23. Vaissière, T. & Herceg, Z. Histone code in the cross-talk during DNA damage signaling. *Cell Res*. **20,** 113–115 (2010).
24. Shanbhag, N. M., Rafalska-Metcalf, I. U., Balane-Bolivar, C., Janicki, S. M. & Greenberg, R. A. ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell*. **141,** 970–981 (2010).
25. Bhakat, K. K., Mokkapati, S. K., Boldogh, I., Hazra, T. K. & Mitra, S. Acetylation of human 8-oxoguanine-DNA glycosylase by p300 and its role in 8-oxoguanine repair *in vivo*. *Mol Cell Biol*. **26,** 1654–1665 (2006).
26. Sun, Y., Jiang, X. & Price, B. D. Tip60: connecting chromatin to DNA damage signaling. *Cell Cycle* **9,** 930–936 (2010).
27. Sun, Y. *et al.* Histone H3 methylation links DNA damage detection to activation of the tumour suppressor Tip60. *Nat Cell Biol*. **11,** 1376–1382 (2009).
28. Stewart, M. D., Li, J. & Wong, J. Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1recruitment. *Mol Cell Biol*. **25,** 2525–2538 (2005).
29. Le May, N., Fradin, D., Iltis, I., Bougnères, P. & Egly, J. M. XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol Cell*. **47,** 622–632 (2012).
30. Pastukh, V. *et al.* An oxidative DNA "damage" and repair mechanism localized in the VEGF promoter is important for hypoxia-induced VEGF mRNA expression. *Am J Physiol Lung Cell Mol Physiol*. **309,** L1367–L1375 (2015).
31. Serra, R. W., Fang, M., Park, S. M., Hutchinson, L. & Green, M. R. A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype. *Elife*. **3,** e02313 (2014).
32. Lee, B., Morano, A., Porcellini, A. & Muller, M. T. GADD45α inhibition of DNMT1 dependent DNA methylation during homology directed DNA repair. *Nucleic Acids Res*. **40,** 2481–2493 (2012).
33. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–1311 (2002).

Acknowledgements

Author Contributions

E.V.A., A. Porcellini, M.E.G. and M.T.M. wrote the main manuscript text, A. Pezone a prepared Figure 4 and A. Porcellini prepared figures 1–3, 5, 6, S1-S3; G.R., R.L., A. Pezone, A. Porcellini, A.M., C.Z. and A.R. performed experiments; E.V.A. and A. Porcellini designed experiments; All authors reviewed the manuscript.

Additional Information

107

Chapter 5: *Relationship between chromatin remodelling and transcriptional activation in the terminal differentiation cellular model Treg.*

Crucial role of ENO1/MBP1 in immunosuppressive activity of a-iTreg.

Treg cells

Regulatory T (Treg) cells or T suppressor are a T-cell subpopulation that play a key role in immune suppression, in the maintenance of immunological self-tolerance and homeostasis (Sakaguchi S. et al., 2008) (Rudensky A.Y. et al., 2011). They represent about 5%-10% of CD4$^+$ T lymphocytes and are characterized by the expression of specific surface markers as interleukin-2 receptor α-chain (CD25) and the transcription factor forkhead box P3 (FOXP3).

Tregs can be divided into two sub-groups: natural regulatory T cells (nTreg) and induced regulatory T cells (iTreg) (Sakaguchi S. et al., 2008). The gene expression profile of iTregs and nTregs is significantly different (Haribhai D. et al., 2009). nTregs arise in the thymus and are strictly controlled by the thymus microenvironment. They are strongly stimulated by Tcell receptor (TCR) and CD28 co-stimulating signals from dendritic cells (Jordan M.S. et al., 2001) (Salamon B. et al., 2000). iTregs are derived from naïve CD4$^+$ cells (nTreg) in the peripheral lymphoid tissues or in vitro when activated in the presence of TGF-β and IL-2. iTreg induction does not require a strong TCR stimulation but requires appropriate antigenic stimulation. Over-intense activation of signaling downstream of the TCR can interfere with iTreg development.

CD4$^+$CD25$^+$ Tregs express a variety of proteins, such as cytotoxic T lymphocyte-associated antigen-4 (CTLA-4), tumor necrosis factor receptor (TNFR) superfamily member, Toll-like receptor 4, C-X-C chemokine receptor type 4 (CXCR4), programmed death-1 (PD-1), and lymphocyte antigen complex 6. Numerous molecular studies have identified other membrane antigens on Treg, such as CD103, galectin-1, LY6, OX-40, 4-IBB, TNFR2, TGF-βR1, and neuropilin 1. However, these proteins lack specificity because they are also commonly expressed by activated effector Tcells (Shu Y. et al., 2016).

Forkhead box p3 (FOXP3) is a member of the fork head transcription factor family and is exclusively expressed in CD4$^+$CD25$^+$ Treg cells (Walker M.R. et al., 2003). FOXP3 plays an important role in the growth, differentiation, and function of Tregs and is recognized as a marker of CD4$^+$CD25$^+$ Treg cells. The survival of Treg in the periphery depends critically on the interaction of the TCR with self antigens, and by the action of IL-2 and TGF-β. Despite Treg do not produce IL-2, their development depends on this cytokine in the thymus and their periphery in survival. Furthermore, studies conducted on knock out TGF-β mice have shown that the IL-2 is necessary for the maintenance of the expression of FOXP3 and for the regulatory function, but not for their development in the thymus (Brunkow M. E. et al., 2001).

In addition to natural Treg, it's possible to get in vitro another group of regulatory T cells (iTreg), through exposure of lymphocytes T CD4$^+$ CD25$^-$ (Tconv) with particular stimulation (for istance anti-CD28 antibodies and anti-TCR, TGF- β1, vitamin D3). Inducible CD4$^+$ CD25$^+$ cells (iTreg) express FOXP3 and show the same regulatory properties of nTreg. The ways in which the Treg exert their action are numerous and not yet fully defined. It's been postulated that CD4$^+$ CD25$^+$ cells could directly inhibit antigen presenting cells (APC). However, it has been shown that Treg can also operate through direct interaction cell-cell independent from APC.

Different diseases with presumed or proven immune or autoimmune pathogenesis have been associated with an alteration of Treg activity, such as multiple sclerosis, autoimmune polyglandular syndrome type 2, myasthenia gravis, psoriasis, diabetes mellitus type I and lupus erythematosus. In these pathologies it was proved that there is a reduction of the expression of FOXP3 and cytokines produced by Treg, in the absence of their depletion, indicating a lack of regulatory function as qualitative as quantitative.

FOXP3 a key gene in T regulatory cells

FOXP3 gene is a master regulator for the development of Tregs, and its constitutive expression is required for Treg-suppressive function. Mutations of FOXP3 impair the development and function of nTreg cells, leading to a severe immunological disorder, called immune dysregulation polyendocrinopathy enteropathy X-linked (IPEX) syndrome, in humans multiple endocrine disease, X chromosome linkage syndrome. In addition, ectopic expression of FOXP3 is able to confer suppressive activity on conventional T (Tconv) cells (Fontenot J. D. et al., 2003) (Hori S. et al., 2003). The human FOXP3 gene is located at the small arm of the X-chromosome (p11.23-13.3), with 11 different coding exons, 3 noncoding exons, and104 introns (Brunkow M. E. et all 2001). The 2 extreme 5′-non coding exons (−2a and −2b) are spliced to a second non coding exon (−1) and separated by 640bp. The −2b and −1 exons are separated by approximately 5000 bp and have conserved noncoding sequences CNS2 (+2177~+2189bp), intron CNS3 (+4393~+4506bp). Except a conserved promoter region, the FOXP3 gene has three adjacent conserved non coding intronic sequences: CNS1, CNS2 and CNS3. These introns are highly conserved, and FOXP3 expression is heavily influenced by the methylation state of certain motifs within them (Zheng Y. et al., 2010). The methylation status of the these conserved CpG island determines FOXP3 expression level and the stability of Treg cells (Toker A. et al., 2011), more specifically the CNS1 and CNS3 are

targets of histone modifications, while the CNS2, also known as TSDR (Treg-cell specific demetilated region) is a region rich in CpG whose selective demethylation is critical for the stability of the expression of FOXP3 in Treg cells. A detailed study of FOXP3 mRNA identifies two different transcription start sites, one proximal and one distal; as a result of alternative splicing seven different isoforms originate: "full length", two free exon 2 (FOXP3Δ2), two depleted of exon 7 (FOXP3Δ7) and two of both exons, 2 and 7 (FOXP3Δ2Δ7).

The suppressive activity of i-Treg is based on the expression of FOXP3E2, the variant splicing form containing exon 2, which is altered in autoimmune diseases, including multiple sclerosis (a severe chronic inflammatory demyelinating disease that affects the central nervous system) and diabetes mellitus type I.

Preliminary data

The human FOXP3 gene encodes for at least 7 different transcript variants, 4 major forms have been extensively characterized. These variants differ both for the untranslated 5' end and for the alternative splicing of exon 2.

We focused our attention on the differential usage of transcription start sites (TSS) of the FOXP3 gene and on the differential expression of FOXP3 transcripts, during the very early phases of activation of human nTreg and Tconv cells. In particular we evaluated the different transcripts originating either from the 5' distal start (DS), one containing the exon 2 (named DS-FOXP3E2) and the other one lacking it (named DS-FOXP3Δ2), or from the 5' proximal start (PS), one containing the exon 2 (named PS-FOXP3E2) and the other one lacking it (named PS-FOXP3Δ2) (Fig. 14).

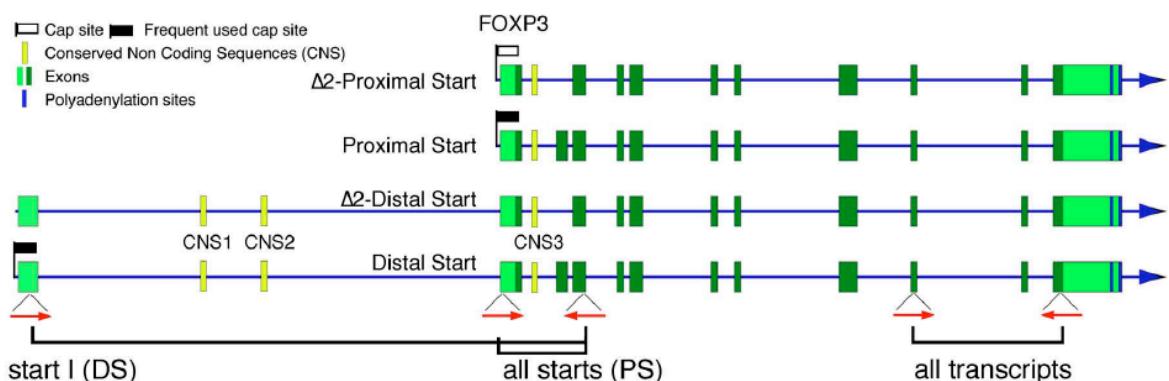**Figure 14** Schematic representation of alternative transcripts of the *FOXP3* gene with the position of oligonucleotides used to measure the different mRNAs. Alternative mRNAs are shown aligned from 5' to 3' on a virtual genome where introns have been shrunk to a minimal length. Exon size is proportional to length, dark green areas represent putative protein coding regions, light green boxes represent the 5'UTR (on the left) and 3'

UTR (on the right). Flags identify validated cap site on the 5' side. Filled flags correspond to frequent events, while empty flags correspond to lesser frequent events. The oligonucleotides are represented as red arrows and their location is shown under the mRNAs. The black brackets indicate the pairs of oligonucleotides used for each amplification.

We evaluated the expression of all transcripts in basal conditions and upon TCR stimulation (anti-CD3/CD28) in the early phases of T cell activation. Our data show that unstimulated nTreg cells have very high levels of the DS-transcripts (high basal level of FOXP3) which were downmodulated upon 12 h of TCR stimulation; while Tconv cells, which do not express FOXP3 in basal conditions, increase FOXP3 mRNA levels upon T cell activation (Fig. 15 A). In particular, unstimulated nTreg expresses both DS-FOXP3E2 that DS-FOXP3Δ2 transcripts (generated by the activity of the transcription start 1). After 30'-60' of TCR stimulation are reduced both transcripts containing the DS-start, and between 6-12h the only splicing variant detectable is the Δ2. PS-transcripts are not found in unstimulated cells and are induced between 30' to 60' of TCR stimulation and remain stable up to 12h. In conclusion in TCR-stimulated nTreg cells the FOXP3E2 variant derive exclusively from the PS-start site (Fig. 15B lower panel). Tconv cells, upon TCR engagement induce the FOXP3 gene by both distal start 1 and proximal start 2 (Fig. 15A). After 15' of stimulation, we observed a small, but significant, induction of both DS- and PS-FOXP3Δ2 transcripts (Figure 15 B higher panel); the DS- rapidly subsided at 30' to reappear at 60' and further increase at 12h. Conversely, the PS- forms increased overtime during TCR stimulation and reached a peak at 12h. Of note, full length DS- and PS-transcripts were absent in unstimulated Tconv cells and were induced after 15' of TCR stimulation from the PS-start, and only upon 12hrs from the DS-start.

To evaluate if the differential FOXP3 gene regulation was correlated with changes histone chromatin modifications, we analyse the FOXP3 promoter region, the conserved non-coding region 3 (CNS3) and Poly_A for the presence of di- and tri-methylations of the lysin 4 and 9 of the H3 histone (H3K4m2, H3K4m3, H3K9m2 and H3K9m3, respectively). ChIP analysis revel that in nTreg on CNS3 the level of di- and tri-methylations of the lysine 4 increase (this region is near to the putative proximal start) while it decrease on distal promoter (in fact we detect an enrichment of H3K9m2/m3) (Fig. 16B). TCR stimulation in Tconv cells induced a rapid H3K4 tri-methylation in both the promoter (within 30') and CNS3 (within 15') (Fig. 16A). Finally, we analyzed the histone code at 3' end of the FOXP3 gene (Poly_A site) in both nTreg and Tconv cells. In nTreg cells we observed a rapid increase (15'-12h) in H3K4m3 and a consequent decrease of H3K4m2 (Figure 16B higher right panel), whereas H3K9 was demethyled at 15 min and methylated at 30 min (Figure 16B lower right panel). In Tconv

cells we found a rapid increase in H3K4m3 (similar to nTreg) (Figure 16A higher right panel)accompanied by an early (15'-30') H3K9m2/m3 methylation (Figure 16A lower right panel). These preliminary data have allowed me to study the altered immunosuppressive activity in individuals with autoimmune disorders such as multiple sclerosis.

During the activation of regulatory T cells of these patients, there is a compromise of the intermediate glycolytic metabolism. This phenomenon is associated to an alteration of suppressive activity of iTreg cells connected with levels of FOXP3-E2 variants. It is known that glycolysis is indispensable for the suppressive function of a-iTreg (activated inducible Treg) cells and it is implicated in the induction of specific splicing variants of FOXP3E2. Our studies have demonstrated a crucial role for the glycolytic enzyme enolase-1 and its isoform (i.e., MBP-1) in modulating the expression of specific splicing forms of FOXP3 that are indispensable for the suppressive function of iTreg cells. The treatment with a inhibitor of glycolysis 2DG diminishes the abundance of the FOXP3-E2 splice forms, necessary for the regulatory function of iTreg cells. The splice form that accounted for the suppressive activity of iTreg cells was the 47-kDa form, which contains the region encoded by exon 2. Under conditions of inhibition of glycolysis (treatment with 2DG), there is an increase in enolase-1 on both the FOXP3 promoter and FOXP3 CNS2, and this is associated with reduced expression of total FOXP3 and the FOXP3-E2 splice variant. The results obtained allowed to conclude that during the activation of Tconv cells enolase-1 enzyme (ENO1) is localized in the cytoplasm and it is engaged in the glycolytic process, when glycolysis is inhibited (in iTreg-2DG cells) ENO1/MBP-1 translocate to the nucleus, where it is localized on the regulatory elements of FOXP3 to repress and modulate the expression to total FOXP3 and the FOXP3-E2 isoform. This leads to the generation of cell-to iTreg with a low expression of FOXP3E2 and a reduced suppressive activity. Silencing of ENO1 in the absence of treatment with 2DG (in iTreg-CTR cells) lead to a reduction in FOXP3 expression that could be ascribed to impairment of the glycolytic activity of enolase-1. Our data have identified a previously unknown molecular mechanism that links a specific glycolytic enzyme to the regulation of a "master gene" (FOXP3) whose product is necessary for the induction and function of Treg cells.
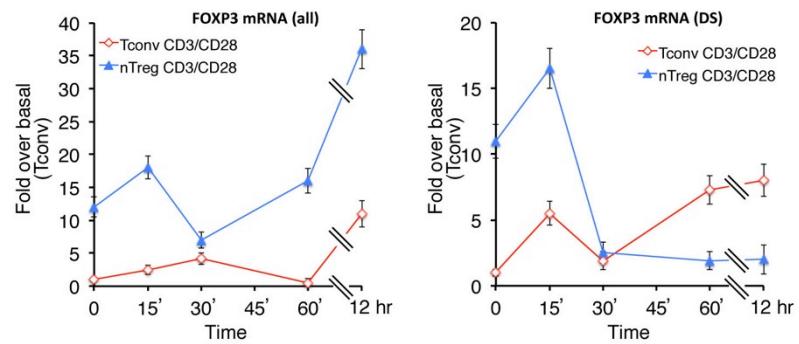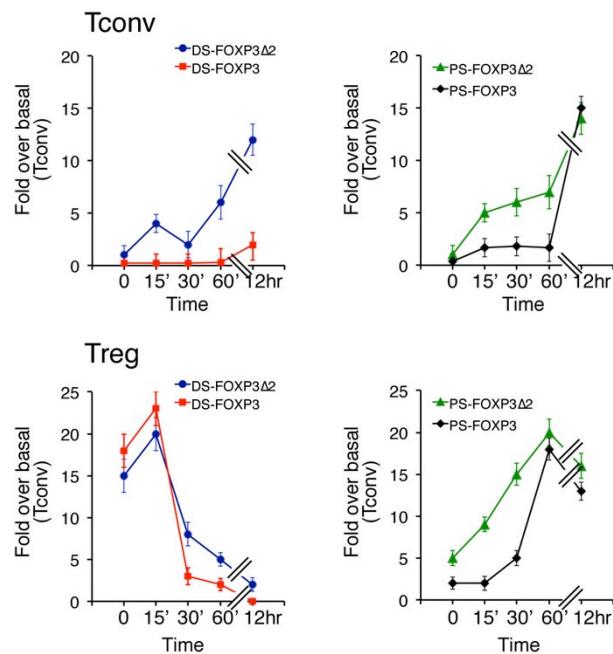
Figura 15 *FOXP3* mRNAs transcripts levels in CD4+ T cells. **A.** Total RNA was prepared from Tcov and nTreg 1 h after culturing them in medium alone or with anti-CD3/anti-CD28 for 15, 30, 60 minutes and 12 hours and analyzed by RT-qPCR (Reverse Transcription, Real Time PCR) with specific primers (all transcripts) (panel A on the left) or specific oligonucleotides for the 5' distal start (DS-mRNAs, see Fig. 15 A on the right) normalized on 18s RNA levels. **B.** Analysis of Exon 2 alternative splicing derived from total (PS) and distal (DS) transcription start sites (TSS) in nTreg and Tconv cells during the early phases and after 12 h of anti anti-CD3/anti-CD28 stimulation. Total RNA was extracted and levels of spliced mRNA forms were monitored by semi-quantitative RT-PCR using specific primers. All the cDNAs were tested for template amount by RT-qPCR for 18s RNA. The statistical analysis derived from at least 3 experiments in triplicate (n ≥9; Mean±SD).
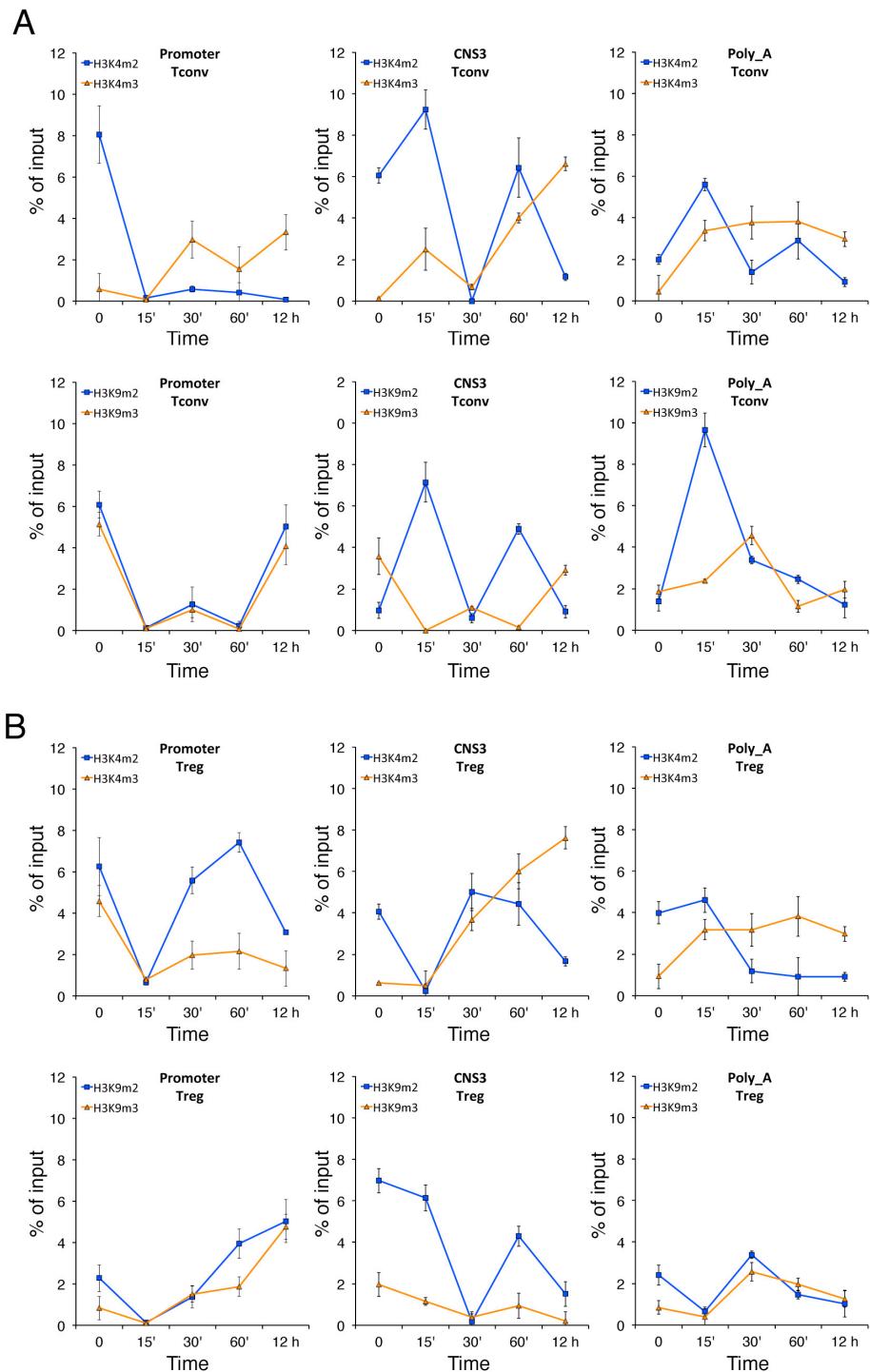
Figura 16 Early methylation-demethylation assessed in nTreg and Tconv cells. Methylation-demethylation cycles of histone H3K4/K9 on *FOXP3* promoter, CNS3 and Poly_A chromatin induced by anti-CD3/anti-CD28 in Tconv and nTreg cells. Tconv and nTreg cells were isolated and cultured for 1 h in medium alone and then treated or not with anti-CD3/anti-CD28 for 15, 30, 60 minutes and 12 hours. qChIP was carried out using specific antibodies recognizing H3K4me2/me3 and H3K9me2/me3. Panel **A**, H3K4me2/me3 occupancy on *FOXP3* promoter, CNS3 and Poly_A in Tconv. Middle panel **A**, H3K9me2/me3 occupancy on *FOXP3* promoter, CNS3 and Poly_A in same cells. Panel **B**, H3K4me2/me3 occupancy on *FOXP3* promoter, CNS3 and Poly_A in Treg. Lower panel **B**, H3K9me2/me3 occupancy on *FOXP3* promoter, CNS3 and Poly_A in same cells. The results were normalized on the percent of input. The statistical analysis derived from at least 3 experiments in triplicate (n ≥9; Mean±SD).

Methods of preliminary data

RNA extraction and qRT-PCR and qPCR.

Total RNA was extracted using Triazol (Gibco/Invitrogen). cDNA was synthesized in a 20 μl reaction volume containing 1 μg of total RNA, 100 units of Superscript III Reverse Transcriptase (Invitrogen), and 1 μl of random primers (200 ng/μl). mRNA was reverse-transcribed for 1 h at 50 °C, and the reaction was heat inactivated for 15 min at 70 °C. The products were stored at -20 °C until use. Quantitative (q)RT-PCR and qPCR were performed three times in six replicates on a 7500 RT-PCR System (Applied Biosystems) using the SYBR Green-detection system (FS Universal SYBR Green MasterRox/Roche Applied Science). The complete list of oligonucleotides used is reported in Table 1.

Chromatin Immuno-Precipitation (ChIP).

Cells were treated as indicated in the legends of the figures. The cells (2.5 x 10^6 for each antibody) were fixed by adding formaldehyde to a final concentration of 1% for 10 min at room temperature and 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, extracted with phenol/chloroform and precipitated to determine DNA concentration and shearing efficiency (input DNA). The sheared chromatin was precleared for 2 h with 1 μg of non-immune IgG (Normal rabbit IgG sc-2027, Normal mouse IgG sc-2025, Santa Cruz Biotechnology Inc., Dallas, TX) and 20 μl of Protein A/G PLUS-Agarose (sc-2003, Santa Cruz Biotechnology) saturated with salmon sperm (1 mg/ml). Precleared chromatin was divided in aliquots and incubated at 4 °C for 16 h with 1 μg of the specific antibody (anti-H3K4m2 ab32356, anti-H3K4m3 ab1012, anti- H3K9m2 ab1220, anti-H3K9m3 ab8898 from Abcam, Cambridge, MA, anti-c-Rel 4727 from Cell Signaling Technology Inc., Danvers, MA) and non-immune IgG (Normal rabbit IgG sc-2027, Normal mouse IgG sc-2025, Santa Cruz Biotechnology Inc., Dallas, TX) respectively. The immuno-complexes were recovered by incubation for 3h at 4 °C with 20 μl of protein-A/G agarose, beads were washed with wash

116

buffers according to the manufacturer's instructions and immunoprecipitated DNA was recovered and subjected to qPCR using the primers indicated in the legend of the specific figures and in Table1.

Table 1

Primers mRNA	
FOXP3 mRNA Fw (All transcripts)	5'-AACATGCGACCCCCTTTCACCTAC-3'
FOXP3 mRNA Rev (All transcripts)	5'- GCCCCCTTCTCGCTCTCCAC-3'
FOXP3 StartI 5Fw (DS)	5'-AGCCAGGCTGATCCTTTTCTGT-3'
FOXP3 StartI 5Rev (DS)	5'-ATTTGGGAAGGTGCAGAGCAGT-3'
FOXP3 All Starts Fw (PS)	5'-GGACAAGGACCCGATGCCCAACC-3'
FOXP3 All Starts Rev (PS)	5'-ATTTGGGAAGGTGCAGAGCAGT-3'
Primers ChIP	
FOXP3 Promotor Fw	5'-GGTTGGCCCTGTGATTTATTTTAG-3'
FOXP3 Promotor Rev	5'-GTGTGGAAGCCGCAGACCTC-3'
FOXP3 CNS3 Fw	5'-GGACAAGGACCCGATGCCCAACC-3'
FOXP3 CNS3 Rev	5'- CTCCCGCCCAGTGCCACAGTAAAG-3'
FOXP3 Poly_A Fw	5'-CTGCCACTCACCTCACTCCCATTCC-3'
FOXP3 Poly_A Rev	5'-GCCCACTGACCGCAACCCTTAT-3'

nature immunology

5.1 Glycolysis controls the induction of human regulatory T cells by modulating the expression of *FOXP3* exon *2* splicing variants

Veronica De Rosa[1,2,12], Mario Galgani[1,12], Antonio Porcellini[3,12], Alessandra Colamatteo[2,4], Marianna Santopaolo[5], Candida Zuchegna[3], Antonella Romano[3], Salvatore De Simone[1], Claudio Procaccini[1], Claudia La Rocca[1], Pietro Biagio Carrieri[6], Giorgia Teresa Maniscalco[7], Marco Salvetti[8], Maria Chiara Buscarinu[8], Adriana Franzese[9], Enza Mozzillo[9], Antonio La Cava[10] & Giuseppe Matarese[4,11]

Human regulatory T cells (T_reg cells) that develop from conventional T cells (T_conv cells) following suboptimal stimulation via the T cell antigen receptor (TCR) (induced T_reg cells (iT_reg cells)) express the transcription factor Foxp3, are suppressive, and display an active proliferative and metabolic state. Here we found that the induction and suppressive function of iT_reg cells tightly depended on glycolysis, which controlled *FOXP3* splicing variants containing exon 2 (Foxp3-E2) through the glycolytic enzyme enolase-1. The Foxp3-E2–related suppressive activity of iT_reg cells was altered in human autoimmune diseases, including multiple sclerosis and type 1 diabetes, and was associated with impaired glycolysis and signaling via interleukin 2. This link between glycolysis and Foxp3-E2 variants via enolase-1 shows a previously unknown mechanism for controlling the induction and function of T_reg cells in health and in autoimmunity.

Distinct subsets of CD4⁺CD25⁺ human regulatory T cells (T_reg cells) are involved in the maintenance of immunological self-tolerance and the control of autoimmunity[1]. T_reg cells are classified into two main subgroups according to their developmental origin, and both express the transcription factor Foxp3 (refs. 2,3). One subgroup arises from the thymus as a distinct lineage, and the other subgroup derives from the peripheral conversion of CD4⁺CD25⁻ conventional T cells (T_conv cells)[4,5]. Experimental evidence indicates that T_reg cell differentiation relies on multiple signaling pathways, such as those derived from the cytokine milieu, engagement of the T cell antigen receptor (TCR), the costimulatory molecule CD28, and signaling via interleukin 2 (IL-2) and its receptor (IL-2R). For example, the cytokine TGF-β can induce Foxp3 expression in T_conv cells stimulated via the TCR, which leads to their conversion into inducible T_reg cells (iT_reg cells) with strong *in vitro* suppressive capacity[6,7]. Additionally, chronic activation of CD4⁺ T cells in the presence of TGF-β can induce the differentiation of a T_reg cell subset that suppresses antigen-specific T cell responses in both mice and humans[6,7]. However, cytokines can be dispensable in the generation of human iT_reg cells, as these cells can also be generated by *in vitro* stimulation of T_conv cells in a cytokine-independent manner[8,9]. In this context, homeostatic proliferation of T_conv cells *in vivo* can produce a population of CD25⁺ T cells with low proliferative capacity and the ability to suppress antigen-specific T cell responses[10]. *In vitro* and *in vivo* studies have shown that the extent of signaling via the TCR and associated costimulatory molecules can affect the outcome of T cell differentiation[11,12]. In this context, culture of CD4⁺ T cells in the presence of dendritic cells presenting low concentrations of antigen results in T_reg cell proliferation together with the conversion of T_conv cells into iT_reg cells[13]. Therefore, the density and affinity of TCR ligation seem to control the induction of Foxp3, since maximal TCR stimulation seems to be detrimental to the differentiation of T_reg cells, whereas optimal induction of Foxp3 is associated with suboptimal TCR engagement[14,15]. Accordingly, antigen-specific T_reg cells can be induced efficiently in mice when an agonist peptide is administrated in sub-immunogenic doses, as supra-physiological stimulation leads to the proliferation of CD4⁺CD25⁺ T cells without Foxp3 expression[16].

Distinct metabolic pathways control the function and differentiation of T cells[17–19]. The activation of CD4⁺ T cells requires metabolic reprogramming characterized by diminished lipid oxidation and increased glycolysis[17–19]. Metabolic enzymes can influence T cell fate

by modulating both lineage-specific differentiation and cytokine production[20,21]. Here we found that highly suppressive human iT$_{reg}$ cells were generated in the absence of exogenous regulatory-type cytokines (i.e., TGF-β or IL-10) following suboptimal stimulation of T$_{conv}$ cells via the TCR. They represented the highly glycolytic and metabolically active fraction of proliferating T$_{conv}$ cells and depended for their induction on the expression of *FOXP3* splicing variants containing exon 2 (*FOXP3*-E2). Among all splicing variants of human Foxp3, the variant produced from *FOXP3*-E2 (called 'Foxp3-E2' here) has been shown to serve a major role in conferring suppressive ability onto T$_{reg}$ cells[22–24]. Therefore, we focused on the metabolic determinants that led to the induction of Foxp3-E2 in human iT$_{reg}$ cells and found that glycolysis controlled the generation of iT$_{reg}$ cells through localization of the glycolytic enzyme enolase-1 to the nucleus. Enolase-1 directly affected the expression of Foxp3-E2 after binding to *FOXP3* regulatory regions, such as the promoter and conserved noncoding sequence 2 (CNS2). We confirmed our findings in studies of subjects with the autoimmune diseases relapsing-remitting multiple sclerosis (RRMS) or type 1 diabetes (T1D), in whom we observed impaired glycolysis and Foxp3-E2 expression in iT$_{reg}$ cells.

RESULTS

Generation of iT$_{reg}$ cells after suboptimal TCR stimulation

To determine whether the induction of human iT$_{reg}$ cells from T$_{conv}$ cells could be achieved through weak stimulation of the TCR in the absence of exogenous cytokines, we obtained peripheral blood mononuclear cells (PBMCs) from healthy human subjects, negatively selected T$_{conv}$ cells (purity, >98%) from those cells and activated them (via the TCR) for 36 h *in vitro* with beads coated with monoclonal antibody (mAb) to the invariant signaling protein CD3 plus mAb to CD28 (at a density of 0.1 bead per cell) (**Supplementary Fig. 1**). At 24 h after activation, we assessed cellular metabolism (glycolysis, mitochondrial respiration and fatty acid oxidation (FAO)) by measuring the extracellular acidification rate (ECAR) and oxygen-consumption rate (OCR). T$_{conv}$ cells underwent an increase in their mitochondrial respiration rate (OCR) and used both glucose and fatty acids, as indicated by an increase in glycolysis and FAO (**Fig. 1a,b**). At 36 h after activation, we sorted T$_{conv}$ cells by flow cytometry into three subsets on the basis of their cell-surface expression of the T cell–activation marker CD25. We subsequently assessed the proliferation marker Ki67, phosphorylation of S6 (a downstream target of the metabolic checkpoint kinase mTOR)

Figure 1 Generation of human iT$_{reg}$ cells from T$_{conv}$ cells during weak stimulation of the TCR. (**a**) ECAR, as an indicator of glycolysis (left), and OCR, quantifying mitochondrial respiration (right), of T$_{conv}$ cells after 24 h of *in vitro* stimulation with medium alone or with mAb to CD3 plus mAb to CD28 (CD3+CD28) (key), in the presence of (above plots) glucose, the ATP-synthase inhibitor oligomycin and 2DG (left), or oligomycin, the uncoupling reagent FCCP and the respiratory-chain inhibitors antimycin A plus rotenone (right), administered sequentially (dotted downward arrows). (**b**) OCR, quantifying FAO (left), of T$_{conv}$ cells stimulated for 24 h *in vitro* as in **a**, right, in the presence of the specific FAO substrate palmitate alone (+ palm) or palmitate plus the FAO inhibitor Etx (+ palm + inh), and their FAO index (right), calculated as ratio of the FCCP-stimulated OCR in the presence of palmitate to the FCCP-stimulated OCR in the presence of palmitate plus inhibitor. (**c**) Flow cytometry of CD4+CD25+ T cells sorted by flow cytometry on the basis of surface CD25 expression (far left), and expression of Ki67, phosphorylation of S6 and expression of Foxp3 (right) in CD25^{hi}, CD25^{int} and CD25^{lo} cells gated as at left. Numbers in plots indicate percent CD25^{hi} cells (top), CD25^{int} cells (middle) or CD25^{lo} cells (bottom) or mean fluorescence intensity (MFI) (in parentheses) of CD25 in those cells (far left), or percent Ki67+ cells (left), cells with phosphorylated S6 (middle) or Foxp3+ cells (right), as well as the MFI (in parentheses) of Ki67 (left), phosphorylated (p-) S6 (middle) or Foxp3 (right). (**d**) Proliferation (left) of CFSE-labeled CD4+ T cells stimulated for 96 h *in vitro* with mAb to CD3 plus mAb to CD28 and cultured alone (CD4+ alone) or in the presence of flow cytometry–sorted CD25^{hi}, CD25^{int} or CD25^{lo} T cells (above plots), and percent proliferation of CD4+ T cells in those conditions (far right). Numbers in plots (left) indicate percent CFSE dilution in CD4+ T cells cultured alone (top left); numbers above bracketed lines indicate percent CFSE dilution in CD4+ T cells cultured with iT$_{reg}$ cells. *$P < 0.05$ and **$P < 0.0001$ (paired two-tailed Student's *t*-test (**b**) or Wilcoxon matched-pairs test (**d**)). Data are from one experiment representative of two experiments with technical duplicates (**a**; mean ± s.e.m.), three independent experiments with technical duplicates (**b**; mean ± s.e.m. (left) or mean ± s.e.m. of $n = 18$ (right)), one experiment representative of eight experiments (**c**), one experiment representative of three experiments (**d**, left) or three independent experiments with technical triplicates (**d**, right; mean ± s.e.m. of $n = 9$ replicates).

119

and Foxp3 in cells with high CD25 expression (CD25^{hi}), intermediate CD25 expression (CD25^{int}) or low CD25 expression (CD25^{lo}) and found that CD25^{hi} cells had highest levels of all these (**Fig. 1c**). In parallel, we evaluated the ability to suppress *in vitro* the proliferation of CD4$^+$ T cells labeled with the division-tracking dye CFSE and stimulated with mAb to CD3 plus mAb to CD28. We found that the subset with the highest expression of CD25, Ki67 and Foxp3 and greatest phosphorylation of S6 also had the strongest suppressive activity (**Fig. 1d**) and that the suppression was contact dependent (**Supplementary Fig. 2a**). We also evaluated the methylation status of *FOXP3* CNS2 to assess the stability of *FOXP3* expression (a more highly methylated CNS2 is related to diminished stability). The frequency of CNS2 methylation in iT$_{reg}$ cells was low and similar to that observed in freshly isolated T$_{reg}$ cells (**Supplementary Fig. 2b**). The low CNS2-methylation status, high suppressive activity and expression of T$_{reg}$ cell–characteristic surface markers (CTLA-4, PD-1, GITR and CD71) were retained for up to 10 d of culture in the presence of IL-2, at levels similar to those of the initially generated iT$_{reg}$ cells (at 36 h) (**Supplementary Fig. 2b–d**). The generation of iT$_{reg}$ cells did not depend on TGF-β and IL-10, since neutralization of these cytokines *in vitro* did not alter the induction of iT$_{reg}$ cells (data not shown). These results indicated that human iT$_{reg}$ cells could be induced from T$_{conv}$ cells stimulated by weak engagement

of TCR in the absence of exogenous cytokines and that they retained their suppressive capacity overtime.

Glycolysis and FAO in the generation and function of iT$_{reg}$ cells
Since the generation and function of T cells depend on underlying metabolic programs[17–19], we evaluated the metabolic pathways involved in the generation of iT$_{reg}$ cells. We activated T$_{conv}$ cells in the presence or absence of a specific inhibitor of glycolysis (2-deoxy-D-glucose (2DG)) and a specific inhibitor of FAO (etomoxir (Etx)) (**Supplementary Fig. 3**). 2DG-treated T$_{conv}$ cells had lower glycolysis than that of untreated T$_{conv}$ cells and had mitochondrial respiration (**Fig. 2a**) and FAO (**Fig. 2b**) similar to that of untreated T$_{conv}$ cells. In contrast, treatment with Etx affected the FAO rate (**Fig. 2b**) and decreased the mitochondrial respiration without any substantial effect on glycolysis (**Fig. 2a**). Although both compounds inhibited the upregulation of CD25 expression, Foxp3 expression was much lower in CD25^{hi} T cells generated in the presence of 2DG and sorted by flow cytometry (iT$_{reg}$-2DG cells) and was higher in sorted CD25^{hi} T cells generated in the presence of Etx (iT$_{reg}$-Etx cells) than in untreated (control (CTR)) CD25^{hi} T cells (iT$_{reg}$-CTR cells) (**Fig. 2c**). To determine whether the metabolic perturbations were associated with an altered regulatory function, we assessed the ability of iT$_{reg}$-2DG cells and iT$_{reg}$-Etx cells to suppress
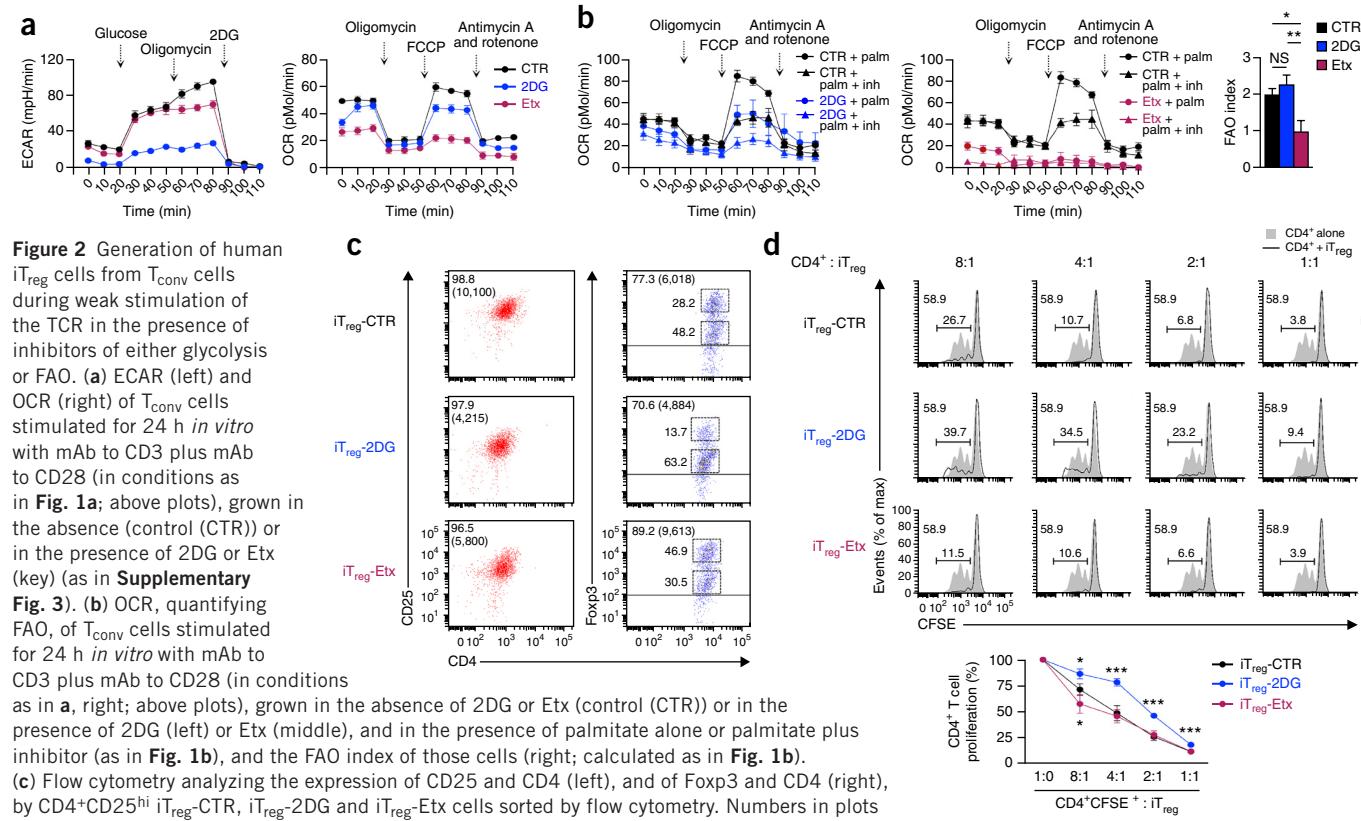

Figure 2 Generation of human iT$_{reg}$ cells from T$_{conv}$ cells during weak stimulation of the TCR in the presence of inhibitors of either glycolysis or FAO. (**a**) ECAR (left) and OCR (right) of T$_{conv}$ cells stimulated for 24 h *in vitro* with mAb to CD3 plus mAb to CD28 (in conditions as in **Fig. 1a**; above plots), grown in the absence (control (CTR)) or in the presence of 2DG or Etx (key) (as in **Supplementary Fig. 3**). (**b**) OCR, quantifying FAO, of T$_{conv}$ cells stimulated for 24 h *in vitro* with mAb to CD3 plus mAb to CD28 (in conditions as in **a**, right; above plots), grown in the absence of 2DG or Etx (control (CTR)) or in the presence of 2DG (left) or Etx (middle), and in the presence of palmitate alone or palmitate plus inhibitor (as in **Fig. 1b**), and the FAO index of those cells (right; calculated as in **Fig. 1b**). (**c**) Flow cytometry analyzing the expression of CD25 and CD4 (left), and of Foxp3 and CD4 (right), by CD4$^+$CD25^{hi} iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells sorted by flow cytometry. Numbers in plots (top left corner) indicate percent CD25$^+$ cells or MFI (in parentheses) of CD25 (left) or percent Foxp3$^+$ cells or MFI (in parentheses) of Foxp3 (right); numbers adjacent to outlined areas (right) indicate percent Foxp3^{hi} cells (top) or Foxp3^{int} cells (bottom). (**d**) Proliferation of CFSE-labeled CD4$^+$ T cells stimulated for 96 h *in vitro* with mAb to CD3 plus mAb to CD28 and cultured alone (filled gray curves) or in the presence of iT$_{reg}$-CTR, iT$_{reg}$-2DG or iT$_{reg}$-Etx cells (black lines) at various ratios (CFSE$^+$CD4$^+$ T cells to iT$_{reg}$ cells, 8:1 to 1:1; above plots). Numbers in plots indicate percent CFSE dilution in CD4$^+$ T cells cultured alone (top left); numbers above bracketed lines indicate percent CFSE dilution in CD4$^+$ T cells cultured with iT$_{reg}$ cells. Below, frequency of proliferation of CD4$^+$ T cells in the conditions above ('–' indicates CD4$^+$ T cells alone). NS, not significant ($P > 0.05$); *$P < 0.05$, **$P < 0.001$ and ***$P < 0.0001$, iT$_{reg}$-2DG versus iT$_{reg}$-CTR or iT$_{reg}$-Etx versus iT$_{reg}$-CTR in **d** (paired two-tailed Student's *t*-test (**b**) or Wilcoxon matched-pairs test (**d**)). Data are from one experiment representative of two experiments with technical duplicates (**a**; mean ± s.e.m.), three independent experiments with technical duplicates (**b**; mean ± s.e.m. (left) or mean ± s.e.m. of $n = 18$ (right)), one experiment representative of eight experiments (**c**), one experiment representative of three (**d**, top), or three independent experiments with technical triplicates (**d**, bottom; mean ± s.e.m. of $n = 45$ (1:1 ratio, iT$_{reg}$-CTR cells from five subjects), $n = 36$ (1:1 ratio, iT$_{reg}$-2DG cells from four subjects), $n = 27$ (1:1 ratio, iT$_{reg}$-Etx cells from three subjects), or $n = 18$ (all other conditions, cells from two subjects).
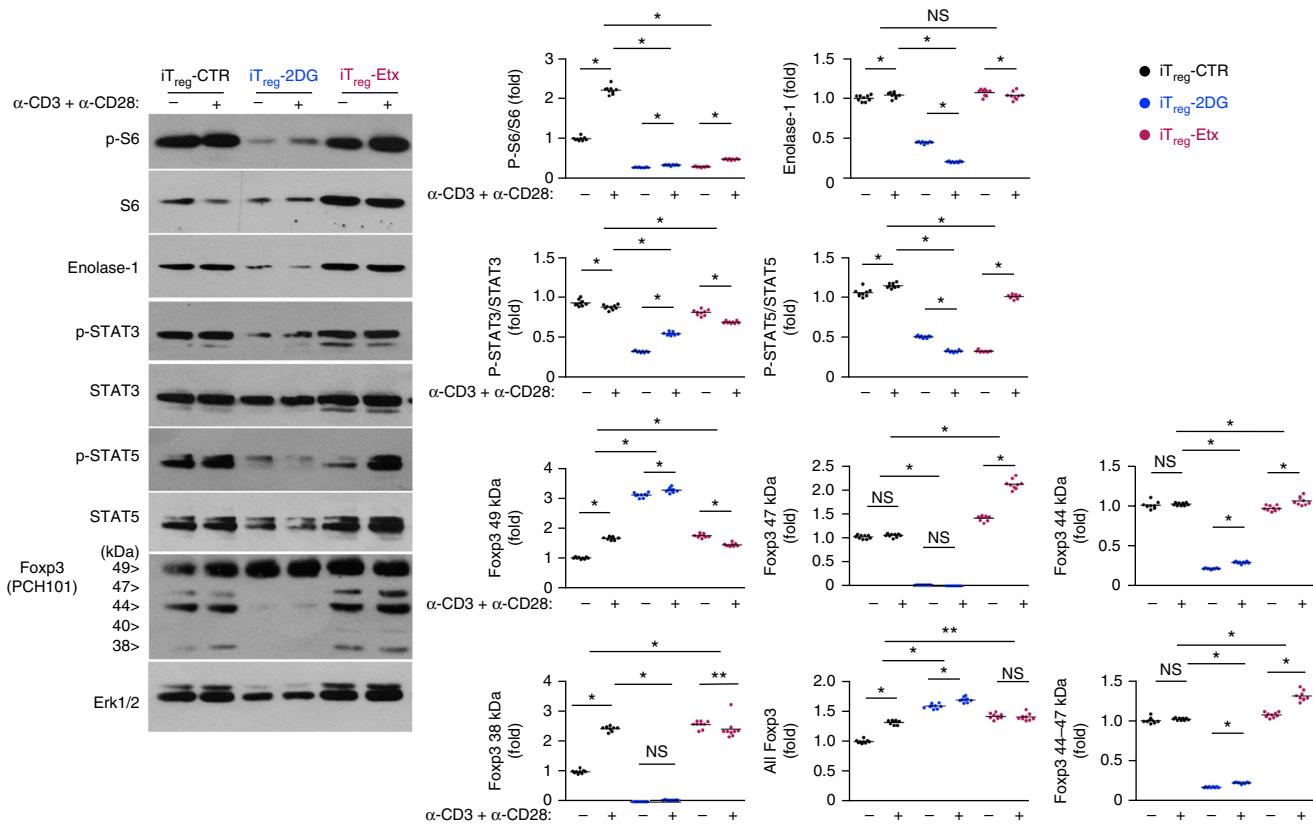
Figure 3 Biochemical analysis of human iT$_{reg}$ cells. Immunoblot analysis (far left) of total and/or phosphorylated (p-) S6, enolase-1, STAT3, STAT5, Foxp3 and Erk1/2 in iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells (above blots) left unstimulated (−) or stimulated (+) for 1 h with mAb to CD3 plus mAb to CD28 (α-CD3 + α-CD28; above lanes); total Foxp3 was assessed with mAb PCH101, which recognizes all splicing variants of Foxp3 (left margin, molecular sizes). Right, densitometry analysis, with each phosphorylated protein normalized to its total form (S6, STAT3 and STAT5), or total protein normalized to total Erk1/2 (enolase-1 and Foxp3), presented relative to results obtained for unstimulated iT$_{reg}$-CTR cells. Each symbol (right) represents an individual data point ($n = 9$); small horizontal lines indicate the mean. *$P < 0.0001$ and **$P < 0.05$ (two-tailed Student's t-test). Data are from one experiment representative of three independent experiments (far left) or three independent experiments with technical triplicates (right).

the proliferation of CD4$^+$ T cells *in vitro*. iT$_{reg}$-2DG cells displayed less suppressive function than that of iT$_{reg}$-CTR cells, whereas iT$_{reg}$-Etx cells had higher suppressive activity than that of iT$_{reg}$-CTR cells (**Fig. 2d**). Together these results indicated that glycolysis was necessary for the generation and suppressive function of human iT$_{reg}$ cells.

Biochemical signature of iT$_{reg}$ cells

To better understand the molecular basis of the findings reported above for iT$_{reg}$ cells, we studied cell metabolism, cytokine-related signaling and Foxp3 expression during specific metabolic perturbations. Immunoblot analysis of iT$_{reg}$ cells stimulated *in vitro* for 1 h with mAb to CD3 plus mAb to CD28 showed that both 2DG and Etx reduced the activation of mTOR in iT$_{reg}$ cells, in terms of the phosphorylation of S6, but via different mechanisms (**Fig. 3**). Indeed, while 2DG abolished the phosphorylation of S6, Etx increased the total amount of S6 without increasing its activity in terms of phosphorylation. Moreover, iT$_{reg}$-2DG cells had lower expression of enolase-1 than that of iT$_{reg}$-CTR cells, whereas iT$_{reg}$-Etx cells had enolase-1 expression similar to that of iT$_{reg}$-CTR cells. Also, iT$_{reg}$-2DG cells had less phosphorylation of the signal transducer STAT3 than that of iT$_{reg}$-CTR cells, whereas iT$_{reg}$-Etx cells had an amount of STAT3 phosphorylation similar to that of iT$_{reg}$-CTR cells. The IL-2–IL-2R–STAT5 signaling pathway has a central role in the induction of Foxp3 and in the generation and homeostasis of T$_{reg}$ cells[25,26]. We found profound impairment in the activation of STAT5 in iT$_{reg}$-2DG cells following TCR stimulation,

whereas the induction of STAT5 phosphorylation was greater in stimulated iT$_{reg}$-Etx cells than in their iT$_{reg}$-CTR counterparts.

Human Foxp3 has various splicing variants; however, despite the importance of Foxp3 in T$_{reg}$ cells, the regulation and functional activity of Foxp3 isoforms remain poorly understood[22–24]. Since we observed that glycolysis was necessary for the induction of human iT$_{reg}$ cells, we evaluated the differential effects of metabolic perturbations on Foxp3 isoforms. Specifically, we found that 2DG hampered induction of the 47-, 44- and 38-kilodalton (kDa) forms of Foxp3 but favored expression of the 49-kDa form (**Fig. 3**). In contrast, Etx increased the expression of the 47-kDa form but did not affect the expression of the 49-, 44- or 38-kDa forms (**Fig. 3**). Of note, the densitometric sum of total Foxp3 forms showed no difference among the three iT$_{reg}$ cell populations (**Fig. 3**), which suggested differential metabolic regulation of the various splicing forms of Foxp3 in iT$_{reg}$ cells. We confirmed these data at the mRNA level by RT-PCR analysis (data not shown). Overall, these analyses revealed that glycolysis controlled the generation of human iT$_{reg}$ cells via a specific biochemical machinery involving the mTOR pathway, IL-2R signaling and the modulation of various splicing variants of Foxp3.

Metabolic programs and phenotypical characteristics of iT$_{reg}$ cells

Next we evaluated the phenotype and the metabolic profile of iT$_{reg}$ cells generated in the presence of inhibitors of either glycolysis or FAO. Inhibition of glycolysis with 2DG diminished expression of the
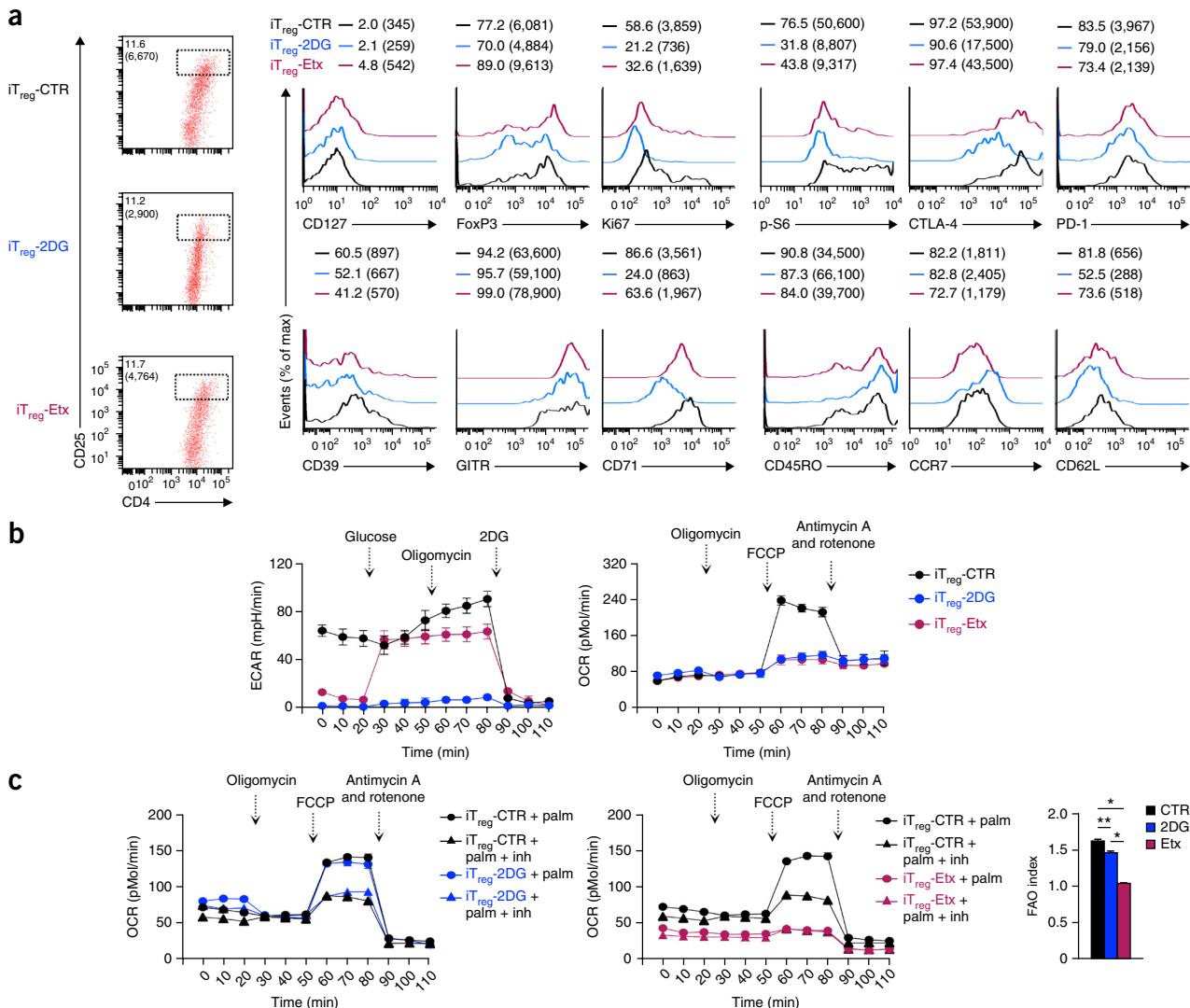
Figure 4 Phenotypical characterization and metabolic programs of iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells. (**a**) Flow cytometry analyzing the expression of CD25 and CD4 in iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells (far left) and of T$_{reg}$ cell–specific markers (horizontal axes) (right). Numbers adjacent to outlined areas (far left) indicate percent CD25^{hi} cells or MFI (in parentheses) of CD25; numbers in plots (right) indicate percent marker-positive cells or MFI (in parentheses) of the marker. (**b**) ECAR (left) and OCR (right) of iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells stimulated for 24 h *in vitro* with mAb to CD3 plus mAb to CD28 (in conditions as in **Fig. 1a**; above plots). (**c**) OCR, quantifying FAO (left), of iT$_{reg}$-CTR and iT$_{reg}$-2DG cells (left) or of iT$_{reg}$-CTR and iT$_{reg}$-Etx cells (middle), stimulated for 24 h *in vitro* (as in **Fig. 1a**, right), in the presence of palmitate alone or palmitate plus inhibitor (as in **Fig. 1b**), and FAO index (right; calculated as in **Fig. 1b**). *$P < 0.0001$ and **$P < 0.005$ (paired two-tailed Student's *t*-test). Data are from one experiment representative of five experiments (**a**), one experiment representative of two experiments with technical duplicates (**b**; mean ± s.e.m.) or three independent experiments with technical duplicates (**c**; mean ± s.e.m. $n = 18$ replicates).

T$_{reg}$ cell markers CTLA-4, PD-1, GITR, CD39 and CD71, as well as the phosphorylation of S6, whereas Etx had an effect similar to that of 2DG on the expression of PD-1 and CD39 and phosphorylation of S6, but slightly reduced the expression of CD71 and CCR7, preserved the expression of CTLA-4, and upregulated the expression of GITR, compared with results obtained with no treatment (in iT$_{reg}$-CTR cells) (**Fig. 4a**). The measurement of ECAR (glycolysis) and OCR (FAO and oxidative phosphorylation) in iT$_{reg}$ cells indicated that iT$_{reg}$-2DG cells had impaired glycolysis and mitochondrial respiration relative to that of iT$_{reg}$-CTR cells (**Fig. 4b**), with FAO rates similar to those of iT$_{reg}$-CTR cells (**Fig. 4c**). Etx affected FAO (**Fig. 4c**) and mitochondrial respiration (**Fig. 4b**) but preserved glycolysis (**Fig. 4b**), relative to results obtained with no treatment (in iT$_{reg}$-CTR cells). These effects, secondary to the metabolic perturbations, were not associated with altered survival and/or viability of T cells, as indicated by staining with

propidium iodide and annexin V, which was similar in iT$_{reg}$-2DG, iT$_{reg}$-Etx and iT$_{reg}$-CTR cells (data not shown). Thus, inhibition of glycolysis impaired the generation and function of iT$_{reg}$ cells and reduced the induction of Foxp3 and the expression of T$_{reg}$ cell markers.

Glycolysis controls iT$_{reg}$ cells' suppressive function via Foxp3-E2
To better understand the effects of metabolism on the expression of the splicing variants of Foxp3, as they were related to the regulatory function of iT$_{reg}$ cells, we analyzed, at level of both mRNA and protein, the expression of all *FOXP3* transcripts and that of the two main *FOXP3* spliced forms[22–24,26]: one containing exon 2 (*FOXP3-E2*), and the other lacking it (**Fig. 5a**). Quantitative RT-PCR with primers spanning exons 9–11 (to amplify all transcripts) showed a 37-fold greater abundance of total *FOXP3* mRNA in iT$_{reg}$-CTR cells than in T$_{conv}$ cells (**Fig. 5b**). iT$_{reg}$-2DG cells had 50% less total
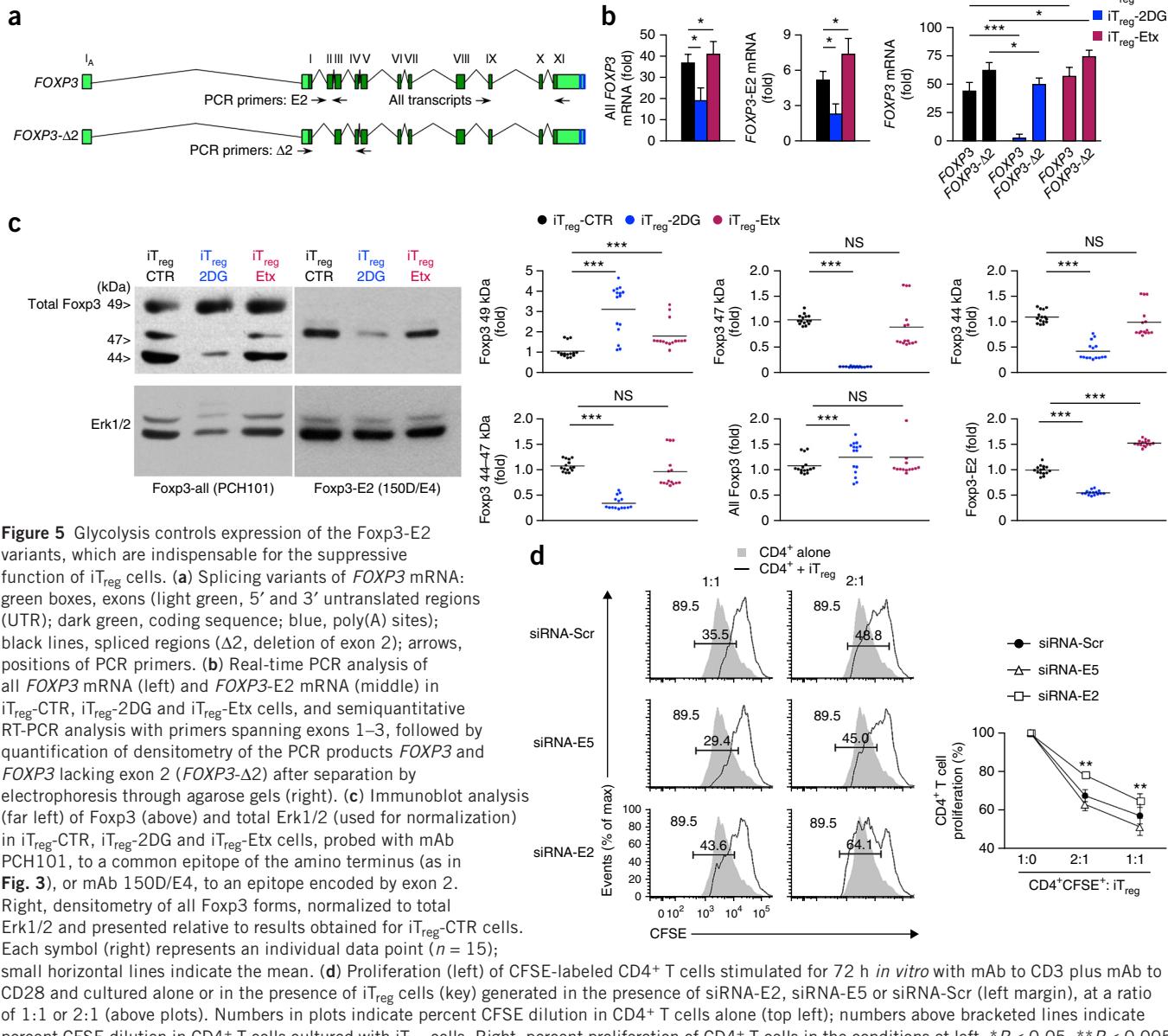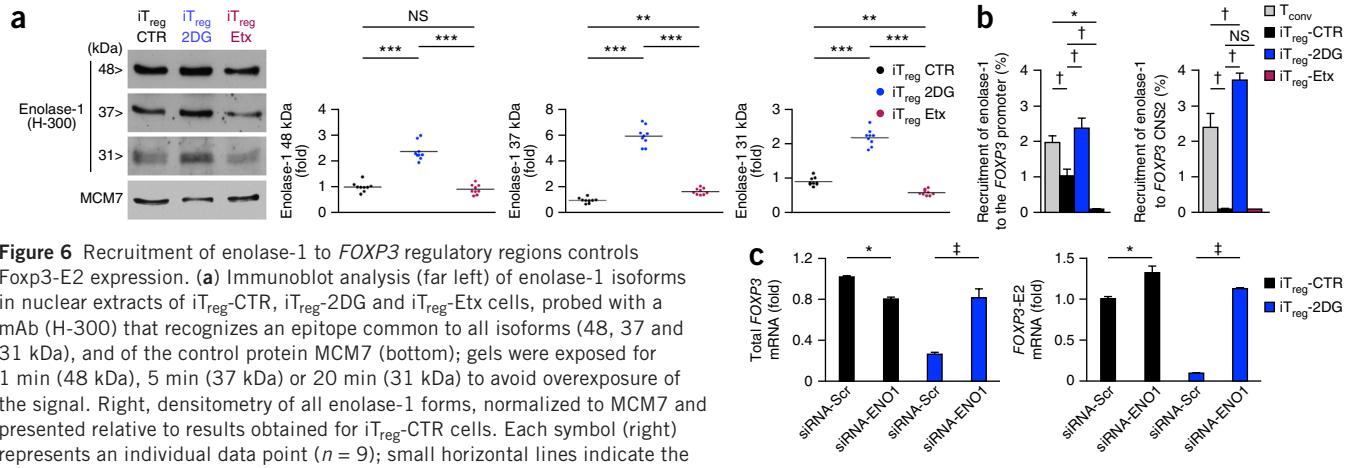
Figure 5 Glycolysis controls expression of the Foxp3-E2 variants, which are indispensable for the suppressive function of iT$_{reg}$ cells. (**a**) Splicing variants of *FOXP3* mRNA: green boxes, exons (light green, 5′ and 3′ untranslated regions (UTR); dark green, coding sequence; blue, poly(A) sites); black lines, spliced regions (Δ2, deletion of exon 2); arrows, positions of PCR primers. (**b**) Real-time PCR analysis of all *FOXP3* mRNA (left) and *FOXP3*-E2 mRNA (middle) in iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells, and semiquantitative RT-PCR analysis with primers spanning exons 1–3, followed by quantification of densitometry of the PCR products *FOXP3* and *FOXP3* lacking exon 2 (*FOXP3*-Δ2) after separation by electrophoresis through agarose gels (right). (**c**) Immunoblot analysis (far left) of Foxp3 (above) and total Erk1/2 (used for normalization) in iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells, probed with mAb PCH101, to a common epitope of the amino terminus (as in **Fig. 3**), or mAb 150D/E4, to an epitope encoded by exon 2. Right, densitometry of all Foxp3 forms, normalized to total Erk1/2 and presented relative to results obtained for iT$_{reg}$-CTR cells. Each symbol (right) represents an individual data point (*n* = 15); small horizontal lines indicate the mean. (**d**) Proliferation (left) of CFSE-labeled CD4+ T cells stimulated for 72 h *in vitro* with mAb to CD3 plus mAb to CD28 and cultured alone or in the presence of iT$_{reg}$ cells (key) generated in the presence of siRNA-E2, siRNA-E5 or siRNA-Scr (left margin), at a ratio of 1:1 or 2:1 (above plots). Numbers in plots indicate percent CFSE dilution in CD4+ T cells alone (top left); numbers above bracketed lines indicate percent CFSE dilution in CD4+ T cells cultured with iT$_{reg}$ cells. Right, percent proliferation of CD4+ T cells in the conditions at left. *$P < 0.05$, **$P < 0.005$ and ***$P < 0.001$, siRNA-E2 versus siRNA-Scr or siRNA-E5 versus siRNA-Scr in **d** (Wilcoxon test (**b**), two-tailed Student's *t*-test (**c**) or Wilcoxon matched-pairs test (**d**)). Data are from three independent experiments with technical triplicates (**b**; mean ± s.d. of *n* = 9 replicates), one experiment representative of five experiments (**c**, far left), five independent experiments with technical triplicates (**c**, right), one experiment representative of three experiments (**d**, left) or three independent experiments with six technical replicates in each (**d**, right; mean ± s.e.m. of *n* = 18 replicates).

FOXP3 mRNA, whereas iT$_{reg}$-Etx cells had very slightly but significantly greater levels of total *FOXP3* mRNA, than iT$_{reg}$-CTR cells had (**Fig. 5b**). Quantitative RT-PCR with specific primers for the amplification of transcripts containing *FOXP3* exon 2 (**Fig. 5a**) showed that the expression of *FOXP3*-E2 mRNA was twofold lower in iT$_{reg}$-2DG cells and 1.4-fold higher in iT$_{reg}$-Etx cells than in iT$_{reg}$-CTR cells (**Fig. 5b**).

To further assess whether 2DG and Etx regulated the splicing of exon 2 differentially, we did semiquantitative RT-PCR with primers spanning exons 1–3. 2DG affected only the expression of *FOXP3*-E2, whereas Etx increased the abundance of both *FOXP3*-E2 and *FOXP3* lacking exon 2 compared with results obtained for untreated (iT$_{reg}$-CTR) cells (**Fig. 5b**). Next we sought to determine whether level of mRNA from the *FOXP3* splicing forms correlated with protein expression in iT$_{reg}$ cells generated in the presence of metabolic inhibitors.

For this, we measured all splicing forms of Foxp3 protein by using two specific mAbs: one (PCH101) that recognizes all splicing variants of Foxp3 (through a common epitope of the amino terminus of Foxp3); the other (150D/E4) specific for the variants encoded by *FOXP3*-E2 (recognized through an epitope encoded by sequence present only in exon 2). Immunoblot analysis with mAb PCH101 revealed that the amount of the 49-kDa form of Foxp3 increased after treatment with each metabolic inhibitor, relative to its abundance in untreated (iT$_{reg}$-CTR) cells, with this phenomenon being more evident in iT$_{reg}$-2DG cells (**Fig. 5c**). iT$_{reg}$-2DG cells had substantially less of both the 47-kDa splicing form and 44-kDa splicing form than iT$_{reg}$-CTR cells had, whereas the level of these forms was not altered in iT$_{reg}$-Etx cells relative to that in iT$_{reg}$-CTR cells (**Fig. 5c**). The densitometric sum of the various Foxp3 splicing forms showed no difference among the three iT$_{reg}$ cell populations (**Fig. 5c**), which

Figure 6 Recruitment of enolase-1 to *FOXP3* regulatory regions controls Foxp3-E2 expression. (**a**) Immunoblot analysis (far left) of enolase-1 isoforms in nuclear extracts of iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells, probed with a mAb (H-300) that recognizes an epitope common to all isoforms (48, 37 and 31 kDa), and of the control protein MCM7 (bottom); gels were exposed for 1 min (48 kDa), 5 min (37 kDa) or 20 min (31 kDa) to avoid overexposure of the signal. Right, densitometry of all enolase-1 forms, normalized to MCM7 and presented relative to results obtained for iT$_{reg}$-CTR cells. Each symbol (right) represents an individual data point ($n = 9$); small horizontal lines indicate the mean. (**b**) Recruitment of enolase-1 to the promoter of *FOXP3* (left) or *FOXP3* CNS2 (right) in T$_{conv}$ cells and iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells (key), assessed by chromatin immunoprecipitation with mAb to enolase-1. (**c**) Quantitative real-time PCR analysis of all *FOXP3* transcripts (left) or *FOXP3*-E2 mRNA (right) in iT$_{reg}$-CTR and iT$_{reg}$-2DG cells generated in the presence of siRNA-Scr or siRNA-ENO1 (horizontal axis); results were normalized to those of control 18S rRNA and are presented relative to those of iT$_{reg}$-CTR cells treated with siRNA-Scr. $*P < 0.05$, $**P < 0.01$, $***P < 0.005$, $†P < 0.001$ and $‡P < 0.0001$ (two-tailed Student's *t*-test). Data are from three independent experiments with technical triplicates (**a**,**c**; mean ± s.e.m. of $n = 9$ in **c**) or three independent experiments (**b** mean ± s.d. of $n = 3$).

indicated that the effects induced by the metabolic inhibitors were specific for defined splicing variants of Foxp3.

To determine whether the lower abundance of the 44- to 47-kDa splicing forms had to be ascribed to a lower abundance of *FOXP3*-E2 (as observed by quantitative RT-PCR), we performed immunoblot analysis of the same cell lysates with mAb 150D/E4 to the variants encoded by *FOXP3*-E2 (Foxp3-E2), which detects a 47-kDa product. iT$_{reg}$-2DG cells had much less Foxp3-E2, whereas iT$_{reg}$-Etx cells had more Foxp3-E2, than iT$_{reg}$-CTR cells had (**Fig. 5c**). Flow cytometry confirmed that treatment with 2DG led to the generation of iT$_{reg}$ cells with reduced expression of Foxp3, both as total Foxp3 and Foxp3-E2, whereas Etx increased both total Foxp3 and Foxp3-E2 in iT$_{reg}$ cells, relative to their expression in untreated (iT$_{reg}$-CTR) cells (**Supplementary Fig. 4a**), and that these effects were dose dependent (**Supplementary Fig. 4b,c**). Thus, glycolysis and FAO controlled the expression of splicing variants of Foxp3 in iT$_{reg}$ cells differentially, at the level of both mRNA and protein.

Since iT$_{reg}$-2DG cells displayed impaired suppressive capacity associated with reduced expression of Foxp3-E2, with the region of Foxp3 encoded by exon 2 being part of the Foxp3 transcriptional repressor domain[22–24], we studied whether Foxp3-E2 was indispensable for the regulatory function of iT$_{reg}$ cells by using small interfering mRNA (siRNA) specific for the *FOXP3*-E2 domain (siRNA-E2). We generated iT$_{reg}$ cells in the presence of siRNA-E2 and assessed their ability to suppress the proliferation of CD4$^+$ T cells *in vitro*. After confirming, by immunoblot analysis, siRNA-induced silencing in T$_{conv}$ cells stimulated with mAb to CD3 plus mAb to CD28 (**Supplementary Fig. 5a**), we found that iT$_{reg}$ cells generated in the presence of siRNA-E2 had less suppressive ability than iT$_{reg}$ cells generated in the presence of control siRNA with a scrambled sequence (siRNA-Scr) (**Fig. 5d**), despite the finding that iT$_{reg}$ cells generated in the presence of siRNA-E2 expressed the other Foxp3 splice variants (44 kDa) (**Supplementary Fig. 5a**). Of note, iT$_{reg}$ cells generated in the presence of siRNA-E2 had an impaired suppressive phenotype (**Fig. 5d**) similar to that of iT$_{reg}$-2DG cells (**Fig. 2d**). As an additional control, we also silenced *FOXP3* exon 5, a common sequence for all *FOXP3* transcripts, with siRNA (siRNA-E5); this reduced the abundance of all the *FOXP3* splicing variants (**Supplementary Fig. 5a**) and markedly hampered the generation of iT$_{reg}$ cells, as indicated

by the substantial reduction in CD25 expression (data not shown). iT$_{reg}$ cells generated in the presence of siRNA-E5, sorted on the basis of their CD25^{hi} phenotype, had a suppressive ability similar to that of iT$_{reg}$ cells generated in the presence of siRNA-Scr (**Fig. 5d**). These phenomena occurred because silencing of *FOXP3* exon 5 reduced the induction of all *FOXP3* splicing variants and, consequently, hampered the upregulation of CD25 expression and the generation of iT$_{reg}$ cells (data not shown). Since the sorting gate was on CD25^{hi} cells, it was likely that CD25^{hi} cells recovered from the silencing of *FOXP3* exon 5 were those in which the knockdown had failed, which had retained their suppressive ability (**Fig. 5d**). Overall, these analyses revealed that glycolysis controlled the expression of Foxp3-E2, which was necessary for the suppressive function of human iT$_{reg}$ cells.

Enolase-1 controls Foxp3-E2 expression

Several isoforms of enolase-1, a key glycolytic enzyme, can bind DNA[27–29]. Among them, the 37-kDa isoform (MBP-1) has been shown to regulate gene expression in various experimental systems[27–29]. Both the glycolytic enzyme enolase-1 and the transcriptional repressor MBP-1 arise from the gene *ENO1* through the use of alternative translation start sites present on *ENO1* mRNA[29,30]. Published evidence suggests that enolase-1 isoforms have multifunctional roles, ranging from glycolytic activity in the cytoplasm to gene regulation in the nucleus[29,30]. For this reason, we evaluated whether enolase-1 isoforms localized in the nuclear fraction of iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells to control *FOXP3* expression. Immunoblot analysis revealed that three different products of the alternative translation start sites of the *ENO1* transcript (48, 37 and 31 kDa in size) were selectively greater in abundance in iT$_{reg}$-2DG cells than in iT$_{reg}$-CTR cells, whereas their amounts were similar in iT$_{reg}$-Etx cells and iT$_{reg}$-CTR cells (**Fig. 6a**). To determine whether enolase-1 was involved in the regulation of *FOXP3* in iT$_{reg}$ cells, we first evaluated the binding of enolase-1 to the *FOXP3* regulatory elements and then assessed the effect of the silencing of *ENO1* on *FOXP3* mRNA expression. Chromatin-immunoprecipitation analysis revealed that the recruitment of enolase-1 to both the *FOXP3* promoter and its CNS2 regions was greater in iT$_{reg}$-2DG cells than in iT$_{reg}$-CTR cells (**Fig. 6b**); in contrast, the association of enolase-1 with the *FOXP3* promoter was much lower in iT$_{reg}$-Etx cells than in iT$_{reg}$-CTR cells, but its association with *FOXP3* CNS2 was similar in
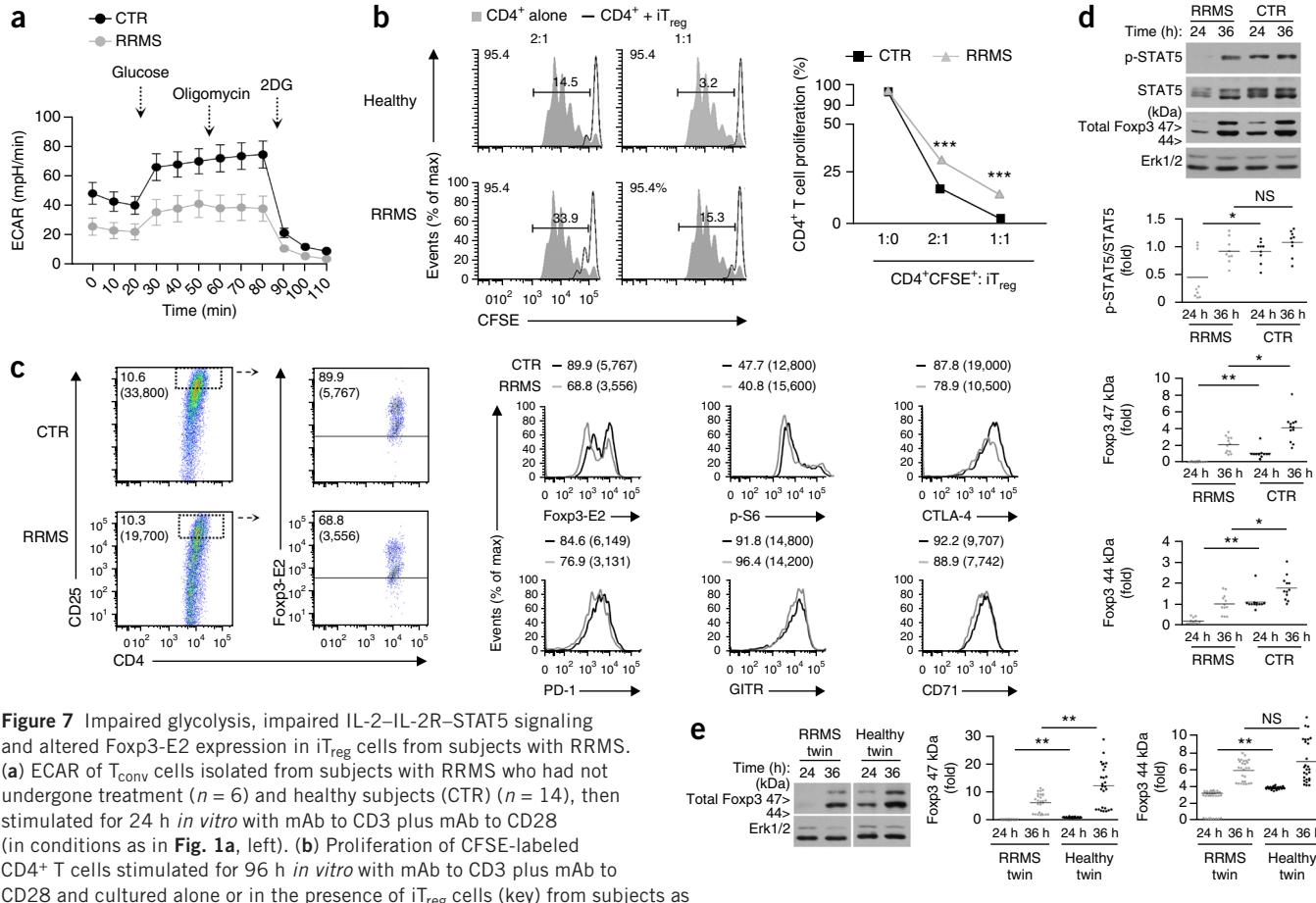
Figure 7 Impaired glycolysis, impaired IL-2–IL-2R–STAT5 signaling and altered Foxp3-E2 expression in iT$_{reg}$ cells from subjects with RRMS. (**a**) ECAR of T$_{conv}$ cells isolated from subjects with RRMS who had not undergone treatment ($n = 6$) and healthy subjects (CTR) ($n = 14$), then stimulated for 24 h *in vitro* with mAb to CD3 plus mAb to CD28 (in conditions as in **Fig. 1a**, left). (**b**) Proliferation of CFSE-labeled CD4$^+$ T cells stimulated for 96 h *in vitro* with mAb to CD3 plus mAb to CD28 and cultured alone or in the presence of iT$_{reg}$ cells from subjects as in **a** (left margin), at a ratio of 1:1 or 1:2 (above plots). Right, percent proliferation of CD4$^+$ T cells in those conditions (far right). Numbers in plots (left) indicate percent CFSE dilution in CD4$^+$ T cells cultured alone (top left); numbers above bracketed lines indicate percent of CFSE dilution in CD4$^+$ T cells cultured with iT$_{reg}$ cells. (**c**) Flow cytometry analyzing the expression of CD25, Foxp3-E2 and CD4 (left) or various T$_{reg}$ cell markers (right) in iT$_{reg}$ cells from subjects as in **a**. Numbers in top left corner indicate percent CD25^{hi} cells or MFI (in parentheses) of CD25 (left) or percent Foxp3-E2^{hi} cells or MFI (in parentheses) of Foxp3-E2 (middle); and numbers above plots (right) indicate percent or MFI (in parentheses) of each marker. (**d**) Immunoblot analysis of phosphorylated and total STAT5 (top) and of the 44- and 47-kDa forms of Foxp3 (bottom) in T$_{conv}$ cells obtained from subjects as in **a** and stimulated for 24 or 36 h (above lanes) *in vitro* with mAb to CD3 plus mAb to CD28. Bottom, densitometry of Foxp3 normalized to total Erk1/2, or of phosphorylated STAT5 normalized to total STAT5, presented relative to results for T$_{conv}$ cells from healthy subjects assessed after 24 h of stimulation. Each symbol (right) represents an individual data point; small horizontal lines indicate the mean. (**e**) Immunoblot analysis (left) of the 44- and 47-kDa forms of Foxp3 in T$_{conv}$ cells obtained from pairs of monozygotic twins ($n = 6$) discordant for RRMS and without treatment, then stimulated for 24 or 36 h *in vitro* with mAb to CD3 plus mAb to CD28 (left). Right, densitometry of Foxp3 normalized to total Erk1/2 and presented relative to results obtained for T$_{conv}$ cells from the healthy twin, assessed after 24 h of stimulation. Each symbol (right) represents an individual data point ($n = 27$); small horizontal lines indicate the mean. *$P < 0.05$, ** $P < 0.001$ and ***$P < 0.0001$ (Wilcoxon matched-pairs test (**b**) or Wilcoxon test (**d,e**)). Data are from six (RRMS) or fourteen (CTR) independent experiments (**a**; mean ± s.e.m.), one experiment representative of three experiments (**b**, left), three independent experiments with technical triplicates (**b**, right; mean ± s.e.m. of $n = 9$ replicates), one experiment representative of three experiments (**c**), one experiment representative of four independent experiments (**d**, left), or four independent experiments (Foxp3) with technical triplicates from four subjects ($n = 12$) or three independent experiments (STAT5) with technical triplicates from three subjects ($n = 9$) (**d**, right), one experiment representative of three (**e**, left) or three independent experiments with nine technical replicates from three sets of twins (**e**, right; $n = 27$).

iT$_{reg}$-Etx cells and iT$_{reg}$-CTR cells (**Fig. 6b**). Freshly isolated, unstimulated T$_{conv}$ cells, which did not express Foxp3, had less cytoplasmic enolase-1 than did iT$_{reg}$-CTR cells, and the recruitment of enolase-1 to the *FOXP3* regulatory elements in these cells was similar to that in iT$_{reg}$-2DG cells (**Supplementary Fig. 5b** and **Fig. 6b**).

To address whether enolase-1 directly controlled the splicing of *FOXP3* mRNA, we silenced *ENO1* by using ENO1-specific siRNA (siRNA-ENO1) during the generation of iT$_{reg}$ cells. After confirmation by quantitative RT-PCR that silencing of *ENO1* occurred (**Supplementary Fig. 5c**), we found that silencing of *ENO1* restored expression of *FOXP3* mRNA in iT$_{reg}$-2DG cells (**Fig. 6c**). Specifically, silencing of *ENO1* in iT$_{reg}$-2DG cells led to the recovery of all *FOXP3* transcripts (fourfold greater abundance than in iT$_{reg}$-2DG cells

generated with siRNA-Scr) and an even greater recovery of *FOXP3*-E2 mRNA (13-fold greater abundance than in iT$_{reg}$-2DG cells generated with siRNA-Scr) (**Fig. 6c**). In addition, the level of *FOXP3*-E2 mRNA was significantly higher, whereas the level of total *FOXP3* mRNA was slightly lower, in iT$_{reg}$-CTR cells generated in the presence of siRNA-ENO1 than in those generated with siRNA-Scr (**Fig. 6c**). Together these data established that enolase-1 specifically repressed the expression of *FOXP3*-E2 in iT$_{reg}$ cells.

Properties of CD4$^+$Foxp3-E2$^+$ T cells

To investigate the physiological relevance of our findings reported above, we examined *ex vivo* CD4$^+$Foxp3-E2$^+$ cells and their biological properties. CD4$^+$Foxp3-E2$^+$ cells constituted (on average) 5.94% ± 2.68%

125

Figure 8 Impaired IL-2–IL-2R–STAT5 signaling and altered Foxp3-E2 expression in iT$_{reg}$ cells from subjects with T1D. (**a**) Flow cytometry analyzing the expression of CD25, Foxp3-E2 and CD4 (left) or T$_{reg}$ cell markers (right) in iT$_{reg}$ cells from subjects with early-onset T1D or healthy subjects (CTR) (numbers in plots as in **Fig. 7c**). (**b**) Immunoblot analysis (far left) of phosphorylated and total STAT5 (top) and the 44- and 47-kDa forms of Foxp3 (below) in T$_{conv}$ cells obtained from subjects as in **a** and stimulated for 24 or 36 h in vitro with mAb to CD3 plus mAb to CD28. Right, densitometry of phosphorylated STAT5 normalized to total STAT5 or of all Foxp3 normalized to total Erk1/2, presented relative to results obtained for T$_{conv}$ cells from healthy subjects, assessed after 24 h of stimulation. Each symbol (right) represents an individual data point; small horizontal lines indicate the mean. (**c**) Immunoblot analysis of the 44- and 47-kDa forms of Foxp3 in T$_{conv}$ cells obtained from monozygotic twins ($n = 2$) discordant for T1D, then stimulated for 24 or 36 h in vitro with mAb to CD3 plus mAb to CD28. Right, densitometry of Foxp3 normalized to total Erk1/2, presented relative to results obtained for T$_{conv}$ cells from healthy donors, assessed after 24 h of stimulation. Each symbol (right) represents an individual data point; small horizontal lines indicate the mean. $*P < 0.005$ and $**P < 0.001$ (Wilcoxon test). Data are from one experiment representative of two (**a**), one experiment representative of three independent experiments (**b**, left), three independent experiments with technical triplicates from five healthy subjects ($n = 15$) or seven subjects with T1D ($n = 21$) (**b**, right), one experiment representative of three independent experiments (**c**, left) or three independent experiments with technical triplicates from one set of twins ($n = 9$) (**c**, right).

of PBMCs from healthy human donors, whereas CD4$^+$ cells expressing total Foxp3 represented 10.20% ± 3.97% of these cells (**Supplementary Fig. 6a**). Flow cytometry revealed that CD4$^+$Foxp3-E2$^+$ cells had higher expression of T$_{reg}$ cell–associated markers (CD25, CTLA-4, GITR, CD39 and CD71) than did CD4$^+$ cells that expressed total Foxp3 (**Supplementary Fig. 6b,c**), and they were more proliferative and metabolically active than CD4$^+$ cells that expressed total Foxp3, as indicated by their larger amounts of Ki67 and phosphorylated S6 (**Supplementary Fig. 6d**). These findings suggested that among all the peripheral human CD4$^+$Foxp3$^+$ T cells, those containing the region encoded by exon 2 represented the fraction with the strongest suppressive properties, in agreement with the results we obtained for iT$_{reg}$ cells.

Defective glycolysis and function of iT$_{reg}$ cells in autoimmunity

We next analyzed the relevance of the findings to human autoimmunity, in patients with RRMS who had not undergone treatment and in patients with T1D. Despite the evidence that the proliferation of T$_{conv}$ cells obtained from subjects with RRMS and stimulated in vitro with mAb to CD3 plus mAb to CD28 was similar to that of their counterparts from healthy control subjects (**Supplementary Fig. 7a**), glycolysis was impaired in T$_{conv}$ cells from subjects with RRMS, as indicated by their diminished basal and maximal glycolysis and glycolytic capacity (**Fig. 7a** and **Supplementary Fig. 7b**). iT$_{reg}$ cells generated from T$_{conv}$ cells of subjects with RRMS demonstrated diminished in vitro suppressive ability relative to that of those generated from healthy age- and sex-matched control subjects (**Fig. 7b**). Moreover, iT$_{reg}$ cells from subjects with RRMS had reduced expression of Foxp3-E2, CTLA-4, PD-1 and CD71 and less phosphorylation of S6, and slightly increased expression of GITR, compared with that of
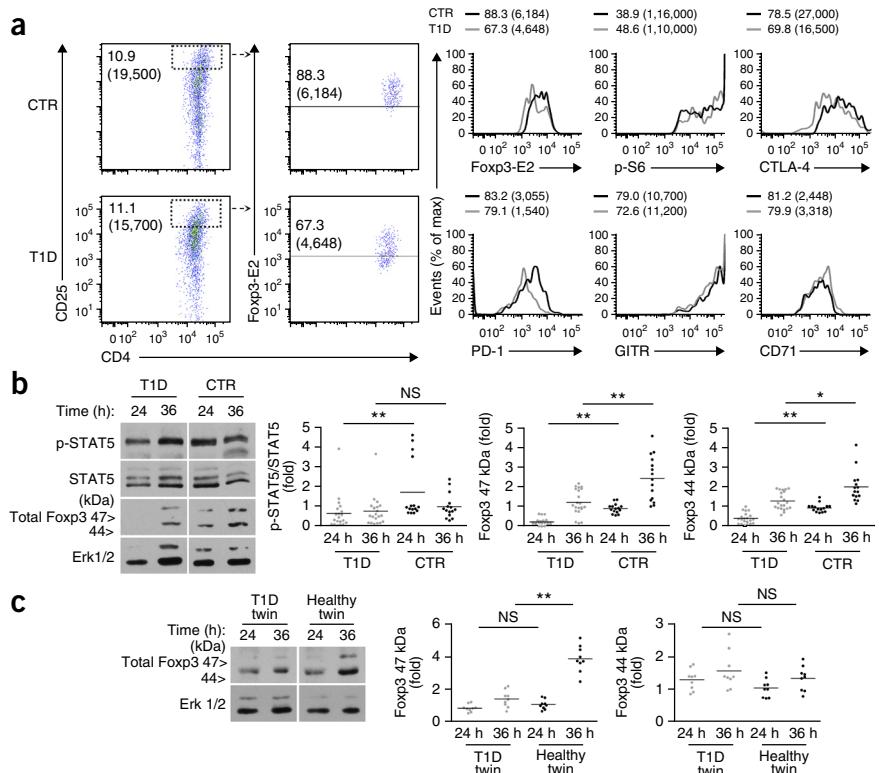
iT$_{reg}$ cells from healthy control subjects (**Fig. 7c**). There was also impaired induction in expression of the 47-kDa and 44-kDa forms of Foxp3 in T$_{conv}$ cells from subjects with RRMS at 24 and 36 h after stimulation with a minimal dose of mAb to CD3 plus mAb to CD28, together with impairment of the IL-2–IL-2R–STAT5 signaling pathway at 24 h (**Fig. 7d**). We also confirmed the specific decrease in Foxp3-E2 expression in T$_{conv}$ cells from subjects with RRMS, by immunoblot analysis with mAb to Foxp3-E2 (**Supplementary Fig. 7c**). Of note, we also confirmed those findings in studies of monozygotic twins discordant for RRMS (one twin affected and one twin healthy) who had also not undergone treatment (**Fig. 7e**), as measured by the induction of Foxp3-E2 in T$_{conv}$ cells stimulated in vitro with mAb to CD3 plus mAb to CD28 (**Supplementary Fig. 1**). Notably, we observed a profound delay in the induction of both the 47-kDa form of Foxp3 and the 44-kDa form of Foxp3 in T$_{conv}$ cells from the twin with RRMS at 24 h, relative to the induction in such cells from the healthy twin (**Fig. 7e**). At 36 h, the level of the 44-kDa form of Foxp3 was not significantly lower than its abundance in T$_{conv}$ cells from the healthy twin, whereas the 47-kDa form of Foxp3 was still profoundly lower in abundance (**Fig. 7e**).

To explore whether altered induction of Foxp3-E2 was a specific feature of RRMS or was instead a common phenomenon in autoimmunity, we also evaluated the induction of Foxp3, T$_{reg}$ cell markers and IL-2–IL-2R–STAT5 signaling in T$_{conv}$ cells obtained from subjects with early-onset T1D and stimulated in vitro with mAb to CD3 plus mAb to CD28 (**Supplementary Fig. 1**). We found that iT$_{reg}$ cells from subjects with T1D had lower expression of Foxp3-E2, CTLA-4, PD-1 and GITR, without differences in the phosphorylation of S6 or expression of CD71, relative to that of iT$_{reg}$ cells from age-matched healthy control subjects, as assessed by flow cytometry (**Fig. 8a**). At both 24 h

and 36 h, there was marked impairment in induction of the 47- and 44-kDa forms of Foxp3 in subjects with T1D relative to their induction in healthy control subjects (**Fig. 8b**). These phenomena correlated with a significant reduction in the phosphorylation of STAT5 observed at 24 h in T_{conv} cells from subjects with T1D, relative to its phosphorylation in such cells from healthy control subjects (**Fig. 8b**). We confirmed, by immunoblot analysis with mAb to Foxp3-E2, the specific decrease in Foxp3-E2 (**Supplementary Fig. 7d**). Of note, we confirmed again the impairment in the induction of Foxp3-E2 in monozygotic twins discordant for T1D. At 24 h, the level of the 47- and 44-kDa forms of Foxp3 in T_{conv} cells from the affected twin was similar to that in cells from the healthy twin, while at 36 h the 47-kDa form was much less abundant in cells from the affected twin than in cells from the healthy twin (**Fig. 8c**), which confirmed our findings for Foxp3-E2 expression in RRMS, another autoimmune disease. Overall, our data suggested that T_{conv} cells from healthy subjects exhibited normal glycolysis during *in vitro* stimulation with mAb to CD3 plus mAb to CD28, which led to the generation of functional iT_{reg} cells that expressed appropriate amounts of Foxp3-E2 (**Supplementary Fig. 8**). In subjects with autoimmunity, impaired glycolysis during T_{conv} cell stimulation led to reduced expression of Foxp3-E2 that would account for the diminished regulatory ability of iT_{reg} cells (**Supplementary Fig. 8**).

DISCUSSION

Here we found that glycolysis was indispensable for the generation of human iT_{reg} cells from T_{conv} cells through localization of the glycolytic enzyme enolase-1 to the nucleus, which directly affected the induction of specific splicing variants of Foxp3-E2 after enolase-1 bound to the *FOXP3* promoter and CNS2. We confirmed our findings in two different human autoimmune diseases, RRMS and T1D. This emphasizes the link between metabolism and immunotolerance in the development of autoimmunity.

Human iT_{reg} cells can be generated anew from non-regulatory T_{conv} cells, following suboptimal stimulation of the TCR and without the requirement for exogenous regulatory-type cytokines or drugs; this provides evidence that TCR signal strength is a critical determinant of the induction of Foxp3, the generation of T_{reg} cells and the engagement of specific metabolic programs during the activation of T_{conv} cells[11,17–19,31]. In our experimental system, weak (0.1 bead per cell) and short (36-hour) activation of T_{conv} cells via the TCR gave rise to iT_{reg} cells that were the highly proliferative and metabolically active fraction of activated T cells (whose generation closely relies on glycolysis). 2DG (a glycolytic inhibitor) affected the induction of Foxp3, reduced the expression of T_{reg} cell markers and impaired the suppressive function of iT_{reg} cells. The inhibition of FAO via Etx increased Foxp3 expression and induced iT_{reg} cells with stronger regulatory properties. These results are in agreement with published reports showing that the engagement of distinct metabolic pathways controls the function and differentiation of T cells[17–19].

Glycolysis fuels the energetic and biosynthetic demands of CD4+ T cell growth and proliferation and represents the metabolic program necessary for appropriate cell activation[17,21,32]. Compelling experimental evidence suggests that during maximal activation of CD4+ T cells, autocrine secretion of IL-2 induces Foxp3 in a small fraction of cells, which leads to conversion into Foxp3+ T_{reg} cells, whose stability over time has been questioned[33]. We observed here that weak stimulation of the TCR generated Foxp3+ iT_{reg} cells that were stable over time. Therefore, we speculate that the activation of T_{conv} cells induces a self-limiting mechanism of activation through the generation of iT_{reg} cells and thereby induces the maintenance of tolerance. This could help to explain the apparent paradox of the association

of T cell immunodeficiency with autoimmunity[34,35]; i.e., a reduced activation rate of T cells (immunodeficiency) might diminish the generation of iT_{reg} cells necessary to replenish the peripheral pool of T_{reg} cells (autoimmunity).

Our study has also demonstrated a crucial role for the glycolytic enzyme enolase-1 and its isoform (i.e., MBP-1) in modulating the expression of specific splicing forms of Foxp3 that were indispensable for the suppressive function of iT_{reg} cells. So far, about twelve different *FOXP3* transcripts have been reported, as well as four to eight different splicing forms of the protein (ranging from 18 kDa to 49 kDa), but their functions are not completely understood[22–24,26]. We found that treatment with 2DG was able to diminish the abundance of the Foxp3-E2 splice forms, which we found were necessary for the regulatory function of iT_{reg} cells (as exemplified by the reduced suppressive ability of both iT_{reg}-2DG cells and cells generated with siRNA-E2, in which *FOXP3*-E2 was silenced). The splice form that accounted for the suppressive activity of iT_{reg} cells was the 47-kDa form, which contains the region encoded by exon 2. The role of that region has been assessed at a molecular level in various experimental systems[36–38]. Foxp3 can block the activity of RORγt, a transcription factor associated with the T_H17 subset of helper T cells, through direct interaction involving the region encoded by exon 2 (ref. 36). Foxp3 interacts also with the transcription factor RORα, and mutations of sequence in exon 2 encoding the 'LxxLL' motif (where 'x' indicates any amino acid) hamper the interaction and repression by Foxp3 (refs. 37,38). In addition, CD4+CD25+ T cells containing spliced forms of *FOXP3* lacking exon 2 are less hypo-responsive and produce more IL-2 than are CD4+CD25+ T cells containing the *FOXP3*-E2 splicing form[24].

The recruitment of different factors to *FOXP3* regulatory elements can influence gene expression or alternative splicing through a combination of physical mechanisms ranging from direct interaction to transient chromatin modification. Among those factors, MBP-1, which is derived from alternative translation of *ENO1* transcripts, has been reported to bind DNA and to suppress gene expression[27–30]. We found that under conditions of inhibition of glycolysis (treatment with 2DG), there was an increase in enolase-1 on both the *FOXP3* promoter and *FOXP3* CNS2, and this was associated with reduced expression of total *FOXP3* and the *FOXP3*-E2 splicing variant. In addition, when glycolysis was inhibited (in iT_{reg}-2DG cells) during RNA-mediated interference of *ENO1*, the observed recovery of total Foxp3 and the Foxp3-E2 isoform depended on the non-glycolytic activity of enolase-1. In contrast, silencing of *ENO1* in the absence of treatment with 2DG (in iT_{reg}-CTR cells) lead to a reduction in Foxp3 expression that could be ascribed to impairment of the glycolytic activity of enolase-1. Our data have identified a previously unknown molecular mechanism that links a specific glycolytic enzyme to the regulation of a 'master gene' (*FOXP3*) whose product is necessary for the induction and function of T_{reg} cells. In resting T_{conv} cells, with low glycolytic demands, enolase-1 was specifically recruited to *FOXP3* regulatory regions, and this correlated with no Foxp3 expression. Activation of the TCR led to an increase in glucose uptake and glycolysis, which supported the energetic needs of T_{conv} cells and drove appropriate Foxp3 expression; impairment of the glycolytic pathway (with 2DG) during the activation of T_{conv} cells hampered the induction of Foxp3-E2 after accumulation of the enolase-1 isoforms in the nucleus on *FOXP3* regulatory regions. This indicated that glycolysis was required for the generation of iT_{reg} cells following the activation of T_{conv} cells. We also found that iT_{reg}-2DG cells had an impairment in the IL-2–IL-2R–STAT5 signaling pathway (which has a pivotal role in the induction of Foxp3 and the generation and homeostasis of T_{reg} cells)[25,39], whereas iT_{reg}-Etx cells exhibited more

activation of this pathway. In T_{conv} cells, a major consequence of IL-2 signaling is the phosphorylation of STAT5, which binds to the *FOXP3* promoter; this leads to *FOXP3* expression and the acquisition of suppressive ability by the cells[40,41]. This is a key event in the generation of T_{reg} cells, as shown by the fact that STAT5- or IL-2Rγ-deficient mice have considerably fewer Foxp3[+] T_{reg} cells than their wild-type counterparts have[42]. We applied those findings to two human autoimmune diseases, including studies of disease-discordant monozygotic twins, in which a reduced suppressive function of iT_{reg} cells was associated specifically with impaired expression of Foxp3-E2, secondary to a glycolytic defect during the activation of T_{conv} cells. In summary, our data suggest that distinct metabolic pathways contribute to the generation of 'waves' of T_{reg} cells during T cell activation to replenish the peripheral T_{reg} cell pool and protect from loss of immunotolerance[34,35]. Targeted manipulation of metabolic disturbances could be instrumental in modulating immunotolerance in autoimmunity.

METHODS
Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
V.D.R., M.G., A.P., A.L.C. and G.M. designed the study, interpreted data and wrote the manuscript; V.D.R., M.G., A.P., A.C., M.S., C.Z., A.R., S.D.S., C.P. and C.L.R. performed the experiments; V.D.R., M.G., A.P. and A.L.C. analyzed the data and interpreted results; V.D.R., M.G., A.P. and C.P. performed statistical analyses; and P.B.C., G.T.M., M.S., M.C.B., A.F. and E.M. obtained human samples from patients and were involved in discussions about data.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Ohkura, N., Kitagawa, Y. & Sakaguchi, S. Development and maintenance of regulatory T cells. *Immunity* **38**, 414–423 (2013).
2. Gavin, M.A. *et al.* Foxp3-dependent programme of regulatory T-cell differentiation. *Nature* **445**, 771–775 (2007).
3. Lin, W. *et al.* Regulatory T cell development in the absence of functional Foxp3. *Nat. Immunol.* **8**, 359–368 (2007).
4. Bilate, A.M. & Lafaille, J.J. Induced CD4+Foxp3+ regulatory T cells in immune tolerance. *Annu. Rev. Immunol.* **30**, 733–758 (2012).
5. Goldstein, J.D. *et al.* Role of cytokines in thymus- versus peripherally derived-regulatory T cell differentiation and function. *Front. Immunol.* **4**, 155 (2013).
6. Chen, W. *et al.* Conversion of peripheral CD4+CD25− naive T cells to CD4+CD25+ regulatory T cells by TGF-β induction of transcription factor Foxp3. *J. Exp. Med.* **198**, 1875–1886 (2003).
7. Zheng, S.G., Gray, J.D., Ohtsuka, K., Yamagiwa, S. & Horwitz, D.A. Generation ex vivo of TGF-β-producing regulatory T cells from CD4+CD25− precursors. *J. Immunol.* **169**, 4183–4189 (2002).
8. Walker, M.R. *et al.* Induction of FoxP3 and acquisition of T regulatory activity by stimulated human CD4+CD25− T cells. *J. Clin. Invest.* **112**, 1437–1443 (2003).
9. Walker, M.R., Carson, B.D., Nepom, G.T., Ziegler, S.F. & Buckner, J.H. De novo generation of antigen-specific CD4+CD25+ regulatory T cells from human CD4+CD25− cells. *Proc. Natl. Acad. Sci. USA* **102**, 4103–4108 (2005).
10. Curotto de Lafaille, M.A., Lino, A.C., Kutchukhidze, N. & Lafaille, J.J. CD25− T cells generate CD25+Foxp3+ regulatory T cells by peripheral expansion. *J. Immunol.* **173**, 7259–7268 (2004).
11. Miskov-Zivanov, N., Turner, M.S., Kane, L.P., Morel, P.A. & Faeder, J.R. The duration of T cell stimulation is a critical determinant of cell fate and plasticity. *Sci. Signal.* **6**, ra97 (2013).
12. Van Panhuys, N., Klauschen, F. & Germain, R.N. T-cell-receptor-dependent signal intensity dominantly controls CD4+ T cell polarization in vivo. *Immunity* **41**, 63–74 (2014).
13. Turner, M.S., Kane, L.P. & Morel, P.A. Dominant role of antigen dose in CD4+Foxp3+ regulatory T cell induction and expansion. *J. Immunol.* **183**, 4895–4903 (2009).
14. Haxhinasto, S., Mathis, D. & Benoist, C. The AKT-mTOR axis regulates de novo differentiation of CD4+Foxp3+ cells. *J. Exp. Med.* **205**, 565–574 (2008).
15. Sauer, S. *et al.* T cell receptor signaling controls Foxp3 expression via PI3K, Akt, and mTOR. *Proc. Natl. Acad. Sci. USA* **105**, 7797–7802 (2008).
16. Liang, S. *et al.* Conversion of CD4+CD25− cells into CD4+CD25+ regulatory T cells in vivo requires B7 costimulation, but not the thymus. *J. Exp. Med.* **201**, 127–137 (2005).
17. Michalek, R.D. *et al.* Cutting edge: distinct glycolytic and lipid oxidative metabolic programs are essential for effector and regulatory CD4+ T cell subsets. *J. Immunol.* **186**, 3299–3303 (2011).
18. Procaccini, C. *et al.* An oscillatory switch in mTOR kinase activity sets regulatory T cell responsiveness. *Immunity* **33**, 929–941 (2010).
19. Pearce, E.L. & Pearce, E.J. Metabolic pathways in immune cell activation and quiescence. *Immunity* **38**, 633–643 (2013).
20. Gerriets, V.A. *et al.* Metabolic programming and PDHK1 control CD4+ T cell subsets and inflammation. *J. Clin. Invest.* **125**, 194–207 (2015).
21. Chang, C.H. *et al.* Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* **153**, 1239–1251 (2013).
22. Kaur, G., Goodall, J.C., Jarvis, L.B. & Hill Gaston, J.S. Characterization of Foxp3 splice variants in human CD4+ and CD8+ T cells–identification of Foxp3Δ7 in human regulatory T cells. *Mol. Immunol.* **48**, 321–332 (2010).
23. Smith, E.L., Finney, H.M., Nesbitt, A.M., Ramsdell, F. & Robinson, M.K. Splice variants of human FOXP3 are functional inhibitors of human CD4+ T-cell activation. *Immunology* **119**, 203–211 (2006).
24. Allan, S.E. *et al.* The role of 2 FOXP3 isoforms in the generation of human CD4+ Tregs. *J. Clin. Invest.* **115**, 3276–3284 (2005).
25. Cheng, G., Yu, A. & Malek, T.R. T-cell tolerance and the multi-functional role of IL-2R signaling in T-regulatory cells. *Immunol. Rev.* **241**, 63–76 (2011).
26. Carbone, F. *et al.* Regulatory T cell proliferative potential is impaired in human autoimmune disease. *Nat. Med.* **20**, 69–74 (2014).
27. Wang, W. *et al.* Identification of alpha-enolase as a nuclear DNA-binding protein in the zona fasciculata but not the zona reticularis of the human adrenal cortex. *J. Endocrinol.* **184**, 85–94 (2005).
28. Hsu, K.W. *et al.* The activated Notch1 receptor cooperates with α-enolase and MBP-1 in modulating c-myc activity. *Mol. Cell. Biol.* **28**, 4829–4842 (2008).
29. Subramanian, A. & Miller, D.M. Structural analysis of alpha-enolase. Mapping the functional domains involved in down-regulation of the c-myc protooncogene. *J. Biol. Chem.* **275**, 5958–5965 (2000).
30. Pancholi, V. Multifunctional alpha-enolase: its role in diseases. *Cell. Mol. Life Sci.* **58**, 902–920 (2001).
31. Gottschalk, R.A., Corse, E. & Allison, J.P. TCR ligand density and affinity determine peripheral induction of Foxp3 in vivo. *J. Exp. Med.* **207**, 1701–1711 (2010).
32. Macintyre, A.N. *et al.* The glucose transporter Glut1 is selectively essential for CD4 T cell activation and effector function. *Cell Metab.* **20**, 61–72 (2014).
33. Malek, T.R. & Castro, I. Interleukin-2 receptor signaling: at the Interface between tolerance and immunity. *Immunity* **33**, 153–165 (2010).
34. Carneiro-Sampaio, M. & Coutinho, A. Early-onset autoimmune disease as a manifestation of primary immunodeficiency. *Front. Immunol.* **6**, 185 (2015).
35. King, C., Ilic, A., Koelsch, K. & Sarvetnick, N. Homeostatic expansion of T cells during immune insufficiency generates autoimmunity. *Cell* **117**, 265–277 (2004).
36. Zhou, L. *et al.* TGF-β-induced Foxp3 inhibits T_H17 cell differentiation by antagonizing RORγt function. *Nature* **453**, 236–240 (2008).
37. Du, J., Huang, C., Zhou, B. & Ziegler, S.F. Isoform-specific inhibition of RORα-mediated transcriptional activation by human FOXP3. *J. Immunol.* **180**, 4785–4792 (2008).
38. Magg, T., Mannert, J., Ellwart, J.W., Schmid, I. & Albert, M.H. Subcellular localization of FOXP3 in human regulatory and nonregulatory T cells. *Eur. J. Immunol.* **42** 1627–1638 (2012).
39. Maruyama, T., Konkel, J.E., Zamarron, B.F. & Chen, W. The molecular mechanisms of Foxp3 gene regulation. *Semin. Immunol.* **23**, 418–423 (2011).
40. D'Cruz, L.M. & Klein, L. Development and function of agonist-induced CD25+Foxp3+ regulatory T cells in the absence of interleukin 2 signaling. *Nat. Immunol.* **6**, 1152–1159 (2005).
41. Passerini, L. *et al.* STAT5-signaling cytokines regulate the expression of FOXP3 in CD4+CD25+ regulatory T cells and CD4+CD25− effector T cells. *Int. Immunol.* **20**, 421–431 (2008).
42. Fontenot, J.D., Rasmussen, J.P., Gavin, M.A. & Rudensky, A.Y. A function for interleukin 2 in Foxp3-expressing regulatory T cells. *Nat. Immunol.* **6**, 1142–1151 (2005).

ONLINE METHODS

Subjects and iT$_{reg}$ cell induction. iT$_{reg}$ cells were induced from CD4$^+$CD25$^-$ cells (T$_{conv}$ cells) from healthy donors and subjects with RR-MS or T1D. The study was approved by the Institutional Review Board of the Università degli Studi di Napoli "Federico II", and peripheral blood was obtained from subjects with RR-MS and healthy control subjects after they signed a written informed consent approved by the Institutional Review Board. Subjects with RR-MS (including two pairs of disease-discordant monozygotic twins) had all not undergone treatment and were 40 ± 20 years of age (at a female/male ratio of 4:1), with relapsing-remitting disease and with a Kurtzke expanded disability status score between 0 and 4 (ref. 43). Healthy donors were matched for age, body mass index and sex with the subjects with RR-MS and had no history of inflammatory, endocrine or autoimmune disease. All blood samples from patients and controls were collected at 9:00 a.m. into heparinized Vacutainers (BD Biosciences) and were processed within the following 4 h.

Subjects with T1D (8 ± 3 years age, at a female/male ratio of 1:1), including the pairs of disease-discordant monozygotic twins, were recruited after resolution of diabetic ketoacidosis, with blood glucose values between 80 mg/dl and 180 mg/dl, after glycemic stabilization through the use of exogenous insulin, achieved in 5 d. Diabetes was defined according to the Global International Diabetes Federation/International Society for Pediatric and Adolescent Diabetes Guidelines for Diabetes in Childhood and Adolescence[44] and included symptoms of diabetes, in addition to a random plasma glucose concentration of ≥11.1 mmol/l (200 mg/dl) or a fasting plasma glucose concentration of ≥7.0 mmol/l (≥126 mg/dl), or a glucose concentration of ≥11.1 mmol/l (≥200 mg/dl) 2 h after glucose loading during an oral glucose-tolerance test, and a glycated hemoglobin (HbA$_{1c}$) of ≥ 6.5 (ref. 44). The following criteria were used for the selection of healthy control subjects for studies of T1D: a fasting blood glucose concentration of <5.5 mmol/l (<100 mg/dl), negative personal and familial history of autoimmune disorders, and negativity for islet autoantibodies at the 99th percentile. The children with T1D and control subjects (matched for age, sex and body mass index) were recruited at the Sezione di Pediatria, Dipartimento di Scienze Mediche Traslazionali, Università di Napoli "Federico II", after the Institutional Review Board of the Università degli Studi di Napoli "Federico II" approved the study and parents provided their written informed consent.

For the generation of iT$_{reg}$ cells, after Ficoll hypaque–gradient centrifugation (GE-Healthcare), T$_{conv}$ cells were isolated from PBMCs by negative selection with a human CD4$^+$CD25$^+$ T cell kit (Invitrogen) (cell purity >98% by flow cytometry), and cells were cultured (2 × 10^6 cells per ml) in six-well plates (Becton-Dickinson; Falcon) with RPMI-1640 medium supplemented with 100 UI/ml penicillin, 100 μg/ml streptomycin (Life Technologies) and either 5% autologous serum or 5% AB human serum (Invitrogen). Cells were stimulated for 36 h with Dynabeads coated with mAb to CD3 plus mAb to CD28 (Invitrogen) at a density of 0.1 bead per cell. T$_{conv}$ cells were then stained with the following mAbs: fluorescein isothiocyanate (FITC)-conjugated anti–human CD4 (RPA-T4; 555346BD; PharMingen), phycoerythrin (PE)–indodicarbocyanine (Cy5)–conjugated anti-human CD25 (M-A251; 555433; BD PharMingen). Cells were sorted by flow cytometry on the basis of their cell-surface expression of CD25 with a MoFlo high-performance cell sorter (Dako/Beckman-Coulter) or a BD FACSJazz (Becton-Dickinson) (**Supplementary Fig. 1**). For the generation of iT$_{reg}$ cells in the presence of inhibitors of specific metabolic pathways, T$_{conv}$ cells were stimulated for 36 h with Dynabeads coated with mAb to CD3 plus mAb to CD28 (0.1 bead per cell) (Invitrogen) in the absence or presence of 1 mM 2-deoxy-D-glucose (2DG) (Sigma-Aldrich) or 200 μM etomoxir (Etx) (Sigma-Aldrich). Cells obtained from each culture were then stained with FITC-conjugated anti–human CD4 (RPA-T4; 555346; BD PharMingen) and PE-Cy5–conjugated anti–human CD25 (M-A251; 555433; BD PharMingen), and iT$_{reg}$ cells from each experimental culture were then sorted by flow cytometry (cell purity >98% by flow cytometry) (**Supplementary Fig. 3**). Cells at 36 h after sorting by flow cytometry are called 'iT$_{reg}$-CTR cells', 'iT$_{reg}$-2DG cells', 'iT$_{reg}$-Etx cells' and 'T$_{conv}$ cells' here.

Molecular signaling and immunoblot analysis. iT$_{reg}$-CTR, iT$_{reg}$-2DG and iT$_{reg}$-Etx cells sorted by flow cytometry were lysed after 1 h of culture with or without stimulation with Dynabeads coated with mAb to CD3 plus mAb to CD28 (0.2 bead per cell) (Invitrogen). T$_{conv}$ cells from subjects with RR-MS or T1D and healthy subjects were lysed 24 and 36 h after stimulation with Dynabeads coated with mAb to CD3 plus mAb to CD28 (0.1 bead per cell) (Invitrogen). Subcellular fractionations were performed with a nuclear-cytosol fractionation kit (NE-PER Nuclear and Cytoplasmic Extraction Reagents; Thermo-Fisher). Total cell lysates were obtained incubation of cells for 20 min at 4 °C in a solution of 50 mM Tris-HCl (pH 7.5), 150 mM NaCl and 1.0% Triton X-100, plus SigmaFast protease inhibitor (S8820; Sigma-Aldrich) and Sigma phosphatase inhibitor (P5726; Sigma-Aldrich), and immunoblot analyses were performed as described[45]. The following mAbs used were used for normalization of nuclear extracts: antibody to S6 phosphorylated at Ser240 and Ser244 (2215), anti-S6 (5G10, 2217), anti-enolase-1 (3810), antibody to STAT3 phosphorylated at Tyr705 (3E2; 9138), anti-STAT3 (79D7; 4904), antibody to STAT5 phosphorylated at Tyr694 (9351), anti-STAT5 (9352) (all 1:1,000 dilution; all from Cell Signaling Technology); anti-Erk1/2 (H72, SC292838) (Santa Cruz Biotechnology); antibody to all Foxp3 (PCH101; 12-4776), antibody to Foxp3-E2 (150D/E4; 14-4774) (all 1:500 dilution; all from eBioscience); anti-enolase-1 (1:500 dilution; H-300; SC15343) and anti-MCM7 (1:3,000 dilution; 141.2; SC9966) (all from Santa Cruz Biotechnology). All filters were quantified by densitometry as described[45]. Results were calculated as the densitometry of each phosphorylated protein normalized to that of its total form (S6, STAT3 and STAT5) or normalized to that of Erk1/2 (enolase-1 and Foxp3) or MCM7 (enolase-1 from nuclear extracts) and are presented relative to results obtained for control samples.

'Seahorse' experiments. Metabolic profiles of T$_{conv}$ and iT$_{reg}$ cells were evaluated in the presence or absence of stimulation with Dynabeads coated with mAb to CD3 plus mAb to CD28 (0.1 bead per cell) (Invitrogen). Real-time measurements of oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) were made with an XF-96 Extracellular Flux Analyzer (Seahorse Bioscience). T$_{conv}$ cells and iT$_{reg}$ cells were plated in XF-96 plates (Seahorse Bioscience) at a density of 2 × 10^5 cells per well and were cultured with RPMI-1640 medium supplemented with 5% AB human serum. OCR was measured in XF medium (non-buffered DMEM medium containing 10 mM glucose, 2 mM L-glutamine and 1 mM sodium pyruvate) under basal conditions and in response to 5 μM oligomycin, 1.5 μM of FCCP (carbonylcyanide-4-(trifluoromethoxy)-phenylhydrazone) and 1 μM of antimycin and rotenone (Sigma-Aldrich). ECAR was measured in XF medium in basal conditions and in response to 10 mM glucose, 5 μM oligomycin and 100 mM of 2DG (all from Sigma-Aldrich). Experiments with the Seahorse system were done with the following assay conditions: 3 min of mixture; 3 min of waiting; and 3 min of measurement. For the analysis of FAO, the XF Palmitate-BSA FAO substrate was used (Seahorse Bioscience). FAO was measured in FAO buffer containing 111 mM NaCl, 4.7 mM KCl, 2 mM MgSO$_4$ and 1.2 mM Na$_2$HPO$_4$, supplemented with 2.5 mM glucose, 0.5 mM carnitine and 5 mM HEPES, pH7.4 (all from Sigma-Aldrich) Etomoxir (final concentration, 40 μM; Sigma-Aldrich) was added 15 min before the 'XF' assay (time 0). At time 0, cells were given 30 μl of 1 mM palmitate conjugated to 0.17 mM BSA. FAO was calculated as the ratio of the FCCP-stimulated OCR in the presence of palmitate to the FCCP-stimulated OCR in the presence of Etx.

Flow cytometry, T cell proliferation and CFSE staining. For staining of PBMCs, T$_{conv}$ or iT$_{reg}$ cells, the following mAbs were used: FITC-conjugated anti–human CD4 (RPA-T4; 555346; BD PharMingen), PE-conjugated anti–human CD25 (M-A251; 555432; BD PharMingen), PE-Cy5-conjugated anti–human CD25 (M-A251; 555433; BD PharMingen), allophycocyanin (APC)-conjugated anti–human CD39 (TU66; 560239; BD PharMingen), APC-conjugated anti–human CD71 (M-A712; 551374; BD PharMingen), APC-conjugated anti–human CD127 (HIL-7R-M21; 555598; BD PharMingen), PE-Cy5-conjugated anti–human CD152 (BNI3 555854; BD PharMingen), PE-conjugated anti–human GITR (DT5D3; 130-092-895; Miltenyi Biotec), PE-Cy5-conjugated anti–human CD45RO (UCHL1 MCA461C; Serotec), FITC-conjugated anti–human CCR7 (150503; FAB197F; R&D Systems), APC-conjugated anti–human CD62L (DREG-56 559772; BD PharMingen) and peridinin chlorophyll protein (PerCP)–cyanine 5.5 (Cy5.5)–conjugated anti–human CD279 (PD-1; EH12.1 581273; BD Pharmingen). Thereafter cells were washed, fixed and permeabilized (fixation-permeabilization buffer; eBio-

science) and were stained with the following mAbs: PE–anti-Foxp3 (PCH101; 12-4776; eBioscience), PE–anti-Foxp3 (150D/E4; 12-4774; eBioscience), Alexa Fluor 647–conjugated antibody to ribosomal protein S6 phosphorylated at Ser235 and Ser236 (D57.2.2E; 4803S; Cell Signaling) and FITC–anti-Ki67 (B56; 556026 BD Biosciences). Analyses were performed with Diva software (BD) and FlowJo software (Tree Star). For T cell–proliferation assays, purified T cells (5×10^4 cells per well) were cultured as described[26]. The fluorescent dye CFSE (5,6-carboxyfluorescein diacetate succinimidyl ester) was used at a concentration of 1 µg/ml (Invitrogen). Flow cytometry analyzing CFSE dilution was performed by gating on $CD4^+CFSE^+$ cells stimulated for 96 h with Dynabeads coated with mAb to CD3 plus mAb to CD28 (0.2 bead per cell) (Invitrogen) alone or cultured with iT_{reg} cells at varous ratios (from 1:1 to 8:1, $CD4^+$ T cells/iT_{reg} cells) in round-bottomed 96-well plates for cell-cell contact experiments or in 24-well plates for Transwell experiments (all from Becton-Dickinson).

RNA extraction, qRT-PCR, qPCR and siRNA. Total RNA was extracted with Triazol (Invitrogen). cDNA was synthesized in a 20-µl reaction volume containing 1 µg of total RNA, 100 units of Superscript III Reverse Transcriptase (Invitrogen) and 1 µl of random primers (200 ng/µl). mRNA was reverse-transcribed for 1 h at 50 °C, and the reaction was heat-inactivated for 15 min at 70 °C. The products were stored at −20 °C until use. Quantitative RT-PCR and quantitative PCR were performed on a 7500 RT-PCR System (Applied Biosystems) with a SYBR Green detection system (FS Universal SYBR Green MasterRox; Roche Applied Science). The *FOXP3* splice variants primers used are in **Supplementary Table 1**. Primers for 18S were used as internal standard control for the reactions (**Supplementary Table 1**).

For knockdown of *FOXP3* and *ENO1* with siRNA, cells were transiently transfected, with a Neon Transfection System, with siRNA at a final concentration of 10 nM (in medium without serum), and incubation was continued for 48 h. RNA with a scrambled sequence, at the same concentration, was used as negative control. siRNA was designed to target *FOXP3* exon 5 (target sequence, 5′-UCGAAGAGCCAGAGGACUU-3′) or exon 2 (target sequence, 5′-GGCCACAUUUCAUGCACCA-3′) or with a scrambled sequence (target sequence, 5′-AUUGAGUGUAGAUUAAGUA-3′). A mixture of siRNA specific for all *ENO1* isoforms was used (target sequences, 5′-CAGUGGUGUCUAUCGAAGA-3′ and 5′-CCGUGACCGAGUCUCU UCA-3′) (4390821; Life Technologies). T_{conv} cells were transfected under the following conditions: pulse voltage, 2,400 V; pulse width, 20 ms; pulse number, 1. Then, cells were stimulated with Dynabeads coated with mAb to CD3 plus mAb to CD28 (0.1 beads/cell) (Invitrogen). After 36 h of stimulation, immunoblot analysis or RT-PCR was performed for confirmation of specific silencing.

Chromatin immunoprecipitation. After the appropriate treatment, cells ($\sim 7 \times 10^5$ for each antibody) were fixed for 10 min at room temperature by the addition of 1 volume of 2% formaldehyde 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 and 0.2% NP-40) containing 1× protease inhibitor cocktail (Roche Applied Science). The lysates were sonicated to produce DNA fragments 300–600 base pairs in length. Sonicated samples were centrifuged and supernatants were diluted twofold in chromatin immunoprecipitation buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA and 1% Triton X-100). An aliquot (1/10) of sheared chromatin was further treated with proteinase K, then underwent extraction with phenol-chloroform and precipitation for measurement of DNA concentration and shearing efficiency (input DNA). Chromatin-immunoprecipitation reactions were set up according to the manufacturer's instructions (Abcam). The sheared chromatin was precleared by incubation for 2 h with 1 µg of non-immune mouse IgG (sc-2025) and rabbit IgG (sc-2027) (both from Santa Cruz Biotechnology) and 20 µl of Protein A/G PLUS-Agarose (Santa Cruz Biotechnology) saturated with salmon sperm (1 mg/ml). Precleared chromatin was divided in aliquots and then was incubated for 16 h at 4 °C with 1 µg of specific antibody (anti-enolase-1: 3810 (Cell Signaling Technology) or sc-15343 (Santa Cruz Biotechnology)) and non-immune IgG (identified above). The immunocomplexes were recovered by incubation for 3 h at 4 °C with 20 µl of protein-A/G agarose, then beads were washed with wash buffers according to the manufacturer's instructions (sc-2003; Santa Cruz Biotechnology), and immunoprecipitated DNA was recovered and analyzed by quantitative PCR (primers, **Supplementary Table 1**).

Methylation experiments. Methyl-sensitive PCR was used for analysis of the methylation status of the CpG island near the CNS2 region of *FOXP3*. Genomic DNA was prepared as described[46]. Genomic DNA (1 µg) isolated from cells was digested overnight at 37 °C with 10 U of HpaII or MspI. Two neighboring regions of the *FOXP3* CpG island have been analyzed by real-time PCR with specific primers (**Supplementary Table 1**). H19 and UEB2B were used as a control for fully methylated and unmethylated regions (primers, **Supplementary Table 1**).

Statistical analysis. All statistical analyses were performed with SPSS software, version 16 for Mac (SPSS), and the GraphPad program (Abacus Concepts). All the comparisons for immunoblot, PCR, proliferation and 'Seahorse' analyses were evaluated with the nonparametric Wilcoxon/Kruskal-Wallis or the paired two-tailed Student's *t*-test. For all analyses, we used two-sided tests, with a P value of <0.05 indicative of statistical significance.

43. Kurtzke, J.F. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* **33**, 1444–1452 (1983).
44. Craig, M.E., Hattersley, A. & Donaghue, K.C. Definition, epidemiology and classification of diabetes in children and adolescents. *Pediatr. Diabetes* **10**, 3–12 (2009).
45. De Rosa, V. *et al.* A key role of leptin in the control of regulatory T cell proliferation. *Immunity* **26**, 241–255 (2007).
46. Cuozzo, C. *et al.* DNA damage, homology-directed repair, and DNA methylation. *Plos. Genet.* **3**, e110 (2007).

Chapter 6: Discussion

Obtened results explain how apparently different mechanisms such as "transcription which induces damage" and "primarily induced DNA damage can condition transcription" are closely connected. This correlation translates into a change of chromatin structure (histone code and DNA looping) which influences transcriptional levels, preserves memory of damage and keeps transcriptional memory too.

First of all, we considered transcriptional activation effects in depending nuclear receptors genes using a model of inducing transcription. Data shows how hormonal stimulation conduce to methylation-demethylation cycles (Shi L. et al., 2011) of K4 and K9 of histone H3 and how these changes are linked to the formation of loops connecting the 5' gene ends, 3'ends and enhancers in examined genes. The formation of the chromatin loops is facilitated by local DNA oxidation following demethylation by LSD1, which presumably releases supercoiling and rigidity of the helix to oxidized bases, in particular oxidized the G to produced 8-oxoG. In addition we assessed the kinetics of repair enzymes recruitment (NER and BER) involved in the repair of oxidative damage, underlining the foundamental role of these enzymes in transcriptional activation. Silencing experiments allowed us to define the timing of recruitment of these enzymes and their role in transcription, in fact their depletion reduces specific mRNA and increases basal transcription (data not shown) but we have not clear instructions about their role in chromatin looping formation. The oxidative burst of Gs, generated by estrogen stimulation, is repaired in asynchronous manner on two strands of different regions of PS2 gene. Our data points out that DNMT3a is essential for the correct repair of oxidized G, in fact it closely works with OGG1 protecting the C point in front of 8-oxoG. We performed pull-down experiments using the deleted forms of DNMT3A (full-lenght, C-terminal, N-terminal) which allowed us to establish that DNMT3A interacts with OGG1 and APE1 (data not shown). Moreover transcriptional activation induces not only the oxidation of G, but also the changes in methylation and hydroxymethylation of CpG, in which DNMT3A and TET1 are involved.

In this first line of research, gene expression regulated by nuclear receptor, we observed that the transcriptional activation induces chromatin remodelling which brings to histone modifications, chromatin 3D structure changes and it uses oxidative damage as start.

Otherwise we evaluated how transcriptional activation and chromatin remodelling respond to primarily induced DNA damage. We explored the changes of histone code that match DNA methylation at the site of DNA repair in a DR-GFP model. Early after DSB formation,

chromatin near the lesion becomes enriched with H3K9m2/3 in two recombinant cell populations (Rec L and Rec H cells). This repressive chromatin marker is linked to DNA methylation, in fact treatment of Rec L cells with DNA demethylating drug, 5-azadC, significantly reduces the levels of H3K9m2/3 on GFP chromatin. We evaluated the histone methylation changes in Rec L and Rec H cells at 14 days post-HR and the obtained results revealed an increase of permissive markers levels (H3K4me2/me3) while the DNA region at 3' end of damaged site is marked by di and tri-methylation of K9. These cell populations do not suffer DNA damage but they are damaged cells' daughter, therefore they have a damage memory. This memory is expressed as DNA methylation and histone code changes. Another important aspect that was analyzed is the link between *de novo* methylation of repaired region and the formation of chromatin loops. We identified two loops: loop A links a region that includes a GFP transcription start site driven by the chicken-actin promoter, high in Rec H cells, and loop C includes a more distal 3' region of the GFP coding sequence located downstream to the I-SceI site, high and stable in Rec L cells. The treatment with 5-aza-C shows that the loop C in L cells is stabilized by the local DNA methylation, while treatment with actinomycin D reduces the loop A in both cell populations and the loop C in the HR-L, suggesting that interactions are generated and maintained by the transcription.

After repair, DNA methylation is not static but it is progressively re- modelled. Demethylation of DNA has recently been linked to BER enzymes, which remove mismatches. In mammals, DNA demethylation requires that 5mC is firstly modified by deaminases or by hydroxylation by TET enzymes (Tahiliani et al., 2009), and then, removed by Thymine DNA Glycosylase (TDG). Our data indicates that BER (TDG, APE1 and OGG1) enzymes are recruited on the DSB and remain on the site after repair. Silencing or inactivation of APE1 or OGG1 enhances methylation and reduces expression of GFP. We proposed that BER enzymes, loaded by the transcription machinery (Wu & Zhang, 2010) contribute to demethylation of Rec L cells and their conversion to Rec H cells. An aspect that could be interesting to investigate whether the methylation is symmetrical or not on the two strands of altered genes in tumors. We saw that asymmetric methylation is associated with transcription that induces cycles of the DNA and histone methylation/demethylation, so the asymmetric promoter methylation of tumor suppressor genes could be used as a prognostic parameter in cancer.

Transcriptional activation modifies chromatin structure, in this change damage is included and it leaves a transcriptional memory on DNA, therefore, we wondered if transcriptional memory works as damage memory. We studied the chromatin changes during cell differentiation process to understand the mechanisms of transcriptional memory that induces chromatin changes on gene of regulative region after the acquisition of a new phenotype. For this purpose we studied the relationship between chromatin remodeling of FOXP3 gene and immunophenotypic differentiation of human primary CD4+ T lymphocytes, in particular we evaluated the histone code modifications, chromatin 3D structure changes and the expression of all FOXP3 transcripts in the early phases of T cell activation, in Tconv and nTreg cells, studing the altered immunosuppressive activity in individuals with autoimmune disorders. We analized the altered immunosuppressive activity in individuals with autoimmune disorders in which the glycolisis is compromised. This dysfunction is associated with an alteration of suppressive activity of iTreg cells connected with FOXP3-E2 variants levels. It is note that glycolysis is indispensable for the suppressive function of a-iTreg, so we demonstrated a crucial role for the glycolytic enzyme enolase-1 and its isoform (i.e., MBP-1) in the modulation of the FOXP3-E2 splice, indispensable for the suppressive function of iTreg cells.

References

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

Ahn S. H., Kim M., Buratowski S. (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol. Cell* 13, 67–76.

Alberts B., Bray D., Lewis J., Raff M., Roberts K., Watson J. D. Biologia molecolare della cellula. *Zanichelli*, 1995.

Allis C. D. and Jenuwein T. (2016) The molecular hallmarks of epigenetic control. *Nat. Rev. Genet*. 17, 487-500.

Andrade M.J. Martin R. Juarez S.F., Lopez de L. (2009) Limited terminal transferase in human DNA polymerase mu defines the required balance between accuracy and efficiency in NHEJ. *Proc. Natl. Acad. Sci. U.S.A*. 106, 16203-16208.

Banerjee S., Kumar B. R., Kundu T. K. (2004) General transcriptional coactivator PC4 activates p53 function. *Mol. Cell. Biol*. 24, 2052–2062.

Bannister A. J. , Kouzarides T. (2007) Chromatin modifications and their function. *Cell* 128(4): 693-705.

Baylin S. B. and Ohm J. E. (2006). Epigenetic gene silencing in cancer – a mechanism for early oncogenic pathway addiction? *Nat Rev Cancer* 6, 107–116.

Bohgaki T., Bohgaki M., Hakem R. (2010) DNA double-strand break signaling and human disorders. *Genome Integr*. 1, 15.

Brunkow M. E., Jeffery E. W., Hjerrild K. A. (2001). Disruption of a new fork head/winged-helix protein, scurfin, results in the fatal lympho proliferative disorder of the scurfy mouse. *Nature Genetics* 27(1): 68–73.

Brunkow M. E., Jeffery E. W., Hjerrild K. A., Paeper B., Clark L. B., Yasayko S. A., Wilkinson J. E., Galas D., Ziegler S. F., Ramsdell F. (2001)Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatally phoproliferative disorder of the scurfy mouse. *Nat Genet* 27:68–73

Burdett V., Baitinger C., Viswanathan M., Lovett S. T., Modrich P. (2001) In vivo requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proc Natl Acad Sci USA*; 98, 6765-6770

Caldecott K.W., Aoufouchi S., Johnson P. & Shall S. (1996). XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. *Nucleic Acids Res*. 24, 4387-4394.

Camenisch U., Dip R., Schumacher S. B., Schuler B., Naegeli H. (2006) Recognition of helical kinks by xeroderma pigmentosum group A protein triggers DNA excision repair. *Nat. Struct. Mol. Biol*. 13, 278–284.

Cheng K. C., Cahill D. S., Kasai H., Nishimura S., Loeb L. A. (1992) 8- Hydroxyguanine, an abundant form of oxidative DNA damage, causes G-T and A-C substitutions. *J Biol Chem* 267: 166–172.

Cho Y. H., Yazici H., Wu H. C., Terry M. B., Gonzalez K., Qu M., Dalay N. and Santella R. M. (2010). Aberrant promoter hypermethylation and genomic hypomethylation in tumor,

adjacent normal tissues and blood from breast cancer patients. *Anticancer Res.* 30, 2489-2496.

Coin F., Oksenych V., Egly J. M. (2007) Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol. Cell* 26, 245–256.

Compe E. and Egly J. M. (2012). TFIIH: when transcription met DNA repair. *Nat Rev Mol Cell Biol.* 13, 343-354. Erratum in: *Nat Rev Mol Cell Biol.* 2012; 13, 476.

Cortazar, D., Kunz, C., Saito, Y., Steinacher, R., and Schar, P. (2007). The enigmatic thymine DNA glycosylase. *DNA Repair (Amst.)* 6, 489–504.

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

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

Dantzer F., de La Rubia G., Ménissier-De Murcia J., Hostomsky Z., de Murcia G. & Schreiber V. (2000). Base excision repair is impaired in mammalian cells lacking Poly(ADP-ribose) polymerase-1. *Biochemistry* 39, 7559-7569.

Deaton, A. M. and Bird A., A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).

Devadas S., Zaritskaya L., Rhee S. G., Oberley L., Williams M. S. (2002). Discrete generation of superoxide and hydrogen peroxide by T cell receptor stimulation: selective regulation of mitogenactivated protein kinase activation and fas ligand expression. *J Exp Med*, 10, 59-70.

Dichtl B., Blank D., Ohnacker M., Friedlein A., Roeder D., Langen H. & Keller W. (2002). A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Mol. Cell* 10, 1139–1150.

Diderich K., Alanazi M., Hoeijmakers J. H. (2011) Premature aging and cancer in nucleotide excision repair-disorders. *DNA Repair* (Amst) 10, 772–780.

Egger G., Jeong S., Escobar S. G., Cortez C. C., Li T. W., Saito Y., Yoo C. B., Jones P. A., Liang G. Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. (2006) *Proc Natl Acad Sci* USA. 103(38):14080-5.

Fagbemi A. F., Orelli B., Scharer O. D. (2010) Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair* (Amst) 10, 722–729.

Fagbemi A. F., Orelli B., Scharer O. D. (2011) Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair* (Amst) 10 722–729.

Fontenot J. D., Gavin M. A., Rudensky A. Y. (2003) FOXP3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat Immunol*;4:330–336.

Franchini D. M., Schmitz K. M., and Petersen-Mahrt S. K. (2012). 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annu. Rev. Genet*. 46, 419–441.

Frit P., Kwon K., Coin F., Auriol J., Dubaele S., Salles B., Jean-Marc Egly J. M. (2002)Transcriptional activators stimulate DNA repair. *Mol. Cell* 10, 1391–1401.

Fuda N. J., Ardehali M. B., Lis J. T. (2009) Defining mechanisms that regulate RNA polymerase II transcription in vivo, *Nature* 461, 186–192.

Fuks F., Burgers W. A., Brehm A., Hughes-Davies L., Kouzarides T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet 24: 88–91.

Fuks F., Hurd P. J., Deplus R., Kouzarides T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* 31: 2305–2312.

Geiman T. M., Sankpal U. T., Robertson A. K., Zhao Y., Zhao Y., Robertson K. D. (2004). DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochem Biophys Res Commun* 318: 544–555.

Getts R. C, Stamato T. D. (1994) Absence of a Ku-like DNA end binding activity in the xrs double-strand DNA repair-deficient mutant. *J. Biol. Chem*. 269, 15981-15984.

Gillet L. C., Scharer O. D.. (2006) Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem. Rev*. 106, 253–276.

Gloire G., Legrand-Poels S., Piette J. (2006). NF-kappaB activation by reactive oxygen species: fifteen years later. *Biochem Pharmacol*, 10, 1493-505.

Goto K., Numata M., Komura J. I, Ono T., Bestor T. H., Kondo H. (1994). Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice. *Differentiation* 56: 39–44.

Greer, E. L. and Shi, Y. (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet*. 13 (5), 343-57.

Gu L., Hong Y., McCulloch S., Watanabe H., Li G. M. (1998) ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair. *Nucleic Acids*; 26:1173-1178.

Harper, J. W. and Elledge, S. J. (2007) The DNA damage response: ten years after. Mol. Cell 28, 739–745.

Hegde M. L., Theriot C. A., Das A., Hegde P. M., Guo Z., Gary R. K., Hazra T. K., Shen B., Mitra S. (2008b). Physical and functional interaction between human oxidized base-specific DNA glycosylase NEIL1 and flap endonuclease 1. *J Biol Chem* 283, 27028–27037.

Hermann A., Goyal R., Jeltsch A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* 279: 48350–48359.

Hildeman D. A. (2004). Regulation of T-cell apoptosis by reactive oxygen species. *Free Radic Biol Med*, 10, 1496-504.

Hori S., Nomura T., Sakaguchi S. (2003) Control of regulatory T cell development by the transcription factor FOXP3. *Science*;299:1057–1061.

Ito S., Shen L., Dai D., Wu S. C., Collins L. B., Swenberg J. A., Egli C., Zhang Y. (2011).Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, 333, pp. 1300–1303.

Jasin M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet* 12: 224-228.

Jensen R. B., Carreira A., Kowalczykowski S. C. (2010) Purified human BRCA2 stimulates RAD51-mediated recombination, *Nature* 467, 678-683.

Jones D. P. (2006). Redefining oxidative stress. *Antioxid Redox Signal*, 10, 1865-79.

Junop M. S. , Obmolova G., Rausch K., Hsieh P., Yang W. (2001) Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair. *Mol Cell*; 7:1-12.

Kadyrov F. A., Dzantiev L., Constantin N., Modrich P. (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*; 126, 297-308

Kareta M. S., Botello Z. M., Ennis J. J., Chou C., Chedin F. (2006) Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. *J Biol Chem*. 281(36):25893

Kim N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. *Nat. Rev*. 13, 204–214.

Klungland, A. and Lindahl, T. (1997) Second pathway for completion of human DNA base excision-repair: ReconstituCell. *Mol. Life Sci*. Vol. 66, 2009

Koch C. A., Agyei R., Galicia S., Metalnikov P., O'Donnell P., Starostine A., Weinfeld M., Durocher D. (2004) Xrcc4 physically links DNA end processing by polynucleotide kinase to DNA ligation by DNA ligase IV. *EMBO J*. 23, 3874-3885.

Kraemer K. H., Patronas N. J., Schiffmann R., Brooks B. P., Tamura D., Di Giovanna J. J. (2007) Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype–phenotype relationship. *Neuroscience* 145, 1388–1396.

Krishnamurthy S., He X., Reyes-Reyes M., Moore C. & Hampsey M. (2004). Ssu72 is an RNA polymerase II CTD phosphatase. *Mol. Cell* 14, 387–394.

Kuehner J. N., Pearson E. L., Moore C. (2011) Unravelling themeans to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol*. 12, 283–294.

Laird P. W., Jackson-Grusby L., Fazeli A., Dickinson S. L., Jung W. E., Li E., Weinberg R. A. and Jaenisch R. (1995). Suppression of intestinal neoplasia by DNA hypomethylation. *Cell* 81, 197–205

Larrea A. A., Lujan S. A., Kunkel T. A.(2010) Snap Shot: DNA mismatch repair. *Cell* 141, 730.

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

Leonhardt H., Page A. W., Weier H. U., Bestor T. H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 71: 865–873.

Li B., Carey M., Workman J. L. (2007) The role of chromatin during transcription, Cell 128, 707–719.

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

Li E., Bestor T. H. and Jaenisch R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915–926.

Lieber M. R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79, 181–211.

Lindahl T. et Wood R. D. (1999). Quality control by DNA repair. *Science* 286, 1897-1905.

Lindahl T., Instability and decay of the primary structure of DNA. (1993) *Nature* 362, 709–715.

Ljungman M. (2005) Activation of DNA damage signaling. *Mutat. Res.* 577, 203–216.

Mathieu N., Kaczmarek N., Naegeli H. (2010) Strand- and site-specific DNA lesion demarcation by the xeroderma pigmentosum group D helicase. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17545–17550.

Mellon, I., Spivak, G., and Hanawalt, P. C. (1987). Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* 51, 241–249.

Min J. H., Pavletich N. P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* 449, 570–575.

Missura M., Buterin T., Hindges T., Hubscher U., Kasparkova J., Brabec V., Naegeli H. (2001) Double-check probing of DNA bending and unwinding by XPA-RPA: an architectural function in DNA repair, *EMBO* J. 20, 3554–3564.

Mittler R., Vanderauwera S., Suzuki N., Miller G, Tognetti V. B., Vandepoele K., Gollery M., Shulaev V., Van Breusegem F. (2011). ROS signaling: the new wave? *Trends Plant Sci*, 10, 300-9.

Mocquet V., Laine J. P., Riedl T., Yajin Z., Lee M. Y., Egly M. J. (2008) Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *EMBO J.* 27, 155–167.

Modrich P., Lahue R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu Rev Biochem*; 65, 101-133.

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

Morano A., Angrisano T., Russo G., Landi R., Pezone A., Bartollino S., Zuchegna C., Babbio F., Bonapace I. M., Allen B., Muller M. T ., Chiariotti L., Gottesman M. E., Porcellini A. and Avvedimento E. V.(2014) Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene. *Nucleic Acids Research*, 42, 2

Moser J., Kool H., Giakzidis I., Caldecott K., Mullenders L. H., Fousteri M. I. (2007) Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Mol. Cell* 27, 311–323.

Nag K., Narsinh H. G., Martinson H. G. (2007) The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase, *Nat. Struct. Mol. Biol.* 14, 662-669.

Napolitano G., Amente S., Lavadera M. L., Di Palo G., Ambrosio S., Lania L., Dellino G. I., Pelicci P. G., Majello B. (2013) Sequence-specific double strand breaks trigger P-TEFb-dependent Rpb1-CTD hyperphosphorylation. Mutation Research 749, 21-27.

Orelli B., McClendon T. B., Tsodikov O. V., Ellenberger T., Niedernhofer L. J., Scharer O. D. (2010) The XPA-binding domain of ERCC1 is required for nucleotide excision repair but not other DNA repair pathways, *J. Biol. Chem.* 285, 3705–3712.

Otterlei M., Warbrick E., Nagelhus T. A., Haug T., Slupphaug G., Akbari M., Aas P. A., Steinsbekk K., Bakke O., Krokan H. E. (1999). Post-replicative base excision repair in replication foci. *EMBO J* 18, 3834–3844.

Pappas D. L. Jr and Hampsey M. (2000). Functional interaction between Ssu72 and the Rpb2 subunit of RNA polymerase II in Saccharomyces cerevisiae. *Mol Cell Biol.* 20, 8343-8351.

Pasini D., Malatesta M., Jung H. R., Walfridsson J., Willer A., Olsson L., Skotte J., Wutz A., Porse B., Jensen O. N., Helin K. (2010) Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. *Nucleic Acids Res* 38(15): 4958-4969.

Perillo B., Ombra, M.N., Bertoni A., Cuozzo C., Sacchetti S., Sasso A., Chiariotti L., Malorni A., Abbondanza C., and Avvedimento E. V. (2008). DNA oxidation as triggered by H3K9me2 demethylation drives estrogen induced gene expression. *Science* 319, 202–206.

Petermann, E., Ziegler M. and Oei S. L. (2003) ATP-dependent selection between single nucleotide and long patch base excision repair. *DNA Repair* (Amst) 2, 1101-1114.

Phillips T. (2008). Regulation of transcription and gene expression in eukaryotes. *Nature Education* 1 (1):199.

Pradhan S., Bacolla A., Wells R. D., Roberts R. J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J Biol Chem* 274: 33002–33010.

Ramsahoye B. H., Biniszkiewicz D., Lyko F., Clark V., Bird A. P., Jaenisch R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc Natl Acad Sci* USA 97: 5237–5242.

Rao K.S. (2007). Mechanisms of disease: DNA repair defects and neurological disease. *Nature clinical practice* 3, 162-172.

Rice J. C., Nishioka K., Sarma K., Steward R., Reinberg D., (2002) Mitotic-specific methylation of histone H4 Lys 20 follows increased PR-Set7 expression and its localization to mitotic chromosomes. *Genes Dev* 16(17): 2225-2230.

Riggs A. D., Xiong Z. Methylation and epigenetic fidelity. (2004) *Proc Natl Acad Sci* USA. 2004;101(1)4-5.

Rudensky A. Y. (2011) Regulatory T cells and FOXP3. *Immunol Rev*;241:260–268.

Sakaguchi S., Yamaguchi T., Nomura T., Ono M. (2008) Regulatory T cells and immune tolerance. *Cell* 2008;133:775–787.

Salomon B., Lenschow D. J. Rhee L., (2000) B7/CD28 cosimulation is essential for the homeostasis of the CD4+CD25+ immunoregulatory T cells that control autoimmune diabetes. Immunity 12(4), 431-440.

Santos-Rosa H., Schneider R., Bannister A. J., Sherriff J., Bernstein B. E., Emre N. C., Schreiber S. L., Mellor J., Kouzarides T. (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* 419(6905): 407-411.

Scharer O. D. (2007). Achieving broad substrate specificity in damage recognition by binding accessible nondamaged DNA. *Mol. Cell* 28,184–186.

Sentenac A. (1985). Eukaryotic RNA polymerases. *CRC Crit Rev Biochem.* 18, 31-90.

Shandilya J., Roberts S. G. E. (2012) The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *Biochimica et Biophysica Acta* 1819, 391-400.

Shen L., Wu H., Diep D., Yamaguchi S., D'Alessio A. C., Fung H.L., Zhang K., and Zhang Y. (2013). Genome-wide analysis reveals TET- and TDG dependent 5-methylcytosine oxidation dynamics. *Cell* 153, 692–706.

Shi Y., Lan F., Matson C., Mulligan P., Whetstine J. R., Cole P. A., Casero R. A., Shi Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homologue LSD1. *Cell* 119, 941–953.

Shu Y., Hu Q., Long H., Chang C., Lu Q., Xiao R. (2016) Epigenetic variability of CD4+CD25+ Treg contributes to the pathogenesis of autoimmune disease. *Clinic Rev Allerg Immunol*.

Stracker T. H., Petrini J. H. (2011) The MRE11 complex: starting from the ends. *Nat. Rev. Mol. Cell Biol.* 12, 90-103.

Strahl B. .D, Allis C. D. (2000) The language of covalent histone modifications. *Nature* 403(6765): 41-45.

Sugasawa K., Akagi J., Nishi R., Iwai S., Hanaoka F. (2009) Two-step recognition of DNA damage for mammalian nucleotide excision repair: directional binding of the XPC complex and DNA strand scanning. *Mol. Cell* 36, 642–653.

Tahiliani M., Koh K.P., Shen Y., Pastor W.A., Bandukwala H., Brudno Y., Agarwal S., Iyer L.M., Liu D.R., Aravind L., Rao A. Conversion of 5methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. (2009). Science. 15; 324 (5929) :930-5.

Tahiliani M., Koh K.P., Shen Y., Pastor W.A., Bandukwala H., Brudno Y., Agarwal S., Iyer L.M., Liu D.R., Aravind L., Rao A. Conversion of 5methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. (2009). Science. 15; 324 (5929) :930-5.

Takahashia S., Suetakea I., Engelhardtc J., Tajima S. (2015) A novel method to analyze 5-hydroxymethylcytosine in CpG sequences using maintenance DNA methyltransferase, DNMT1. *FEBS Open Bio*. Volume 5, 2015, Pages 741–747.

Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., Patel, D. J. (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nature Struct. Mol. Biol.* 14, 1025–1040.

Toker A., Huehn J. (2011) To be or not to be a Treg cell: lineage decisions controlled by epigenetic mechanisms. *Sci Signal* 4:pe4

Tsodikov O. V., Ivanov D., Orelli B., Staresincic L., Shoshani I., Oberman R., Scharer O. D., Wagner G., Ellenberger T. (2007) Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA, EMBO J. 26, 4768–4776.

Tsukada Y., Fang J., Erdjument-Bromage H., Warren M. E., Borchers C. H., Tempst P. & Zhang Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. Nature 439, 811–81.

Turner B. M. (1993) Decoding the nucleosome. *Cell* 75(1): 5-8.

Um S., Harbers M., Benecke A., Pierrat B., Losson R., and Chambon P. (1998). Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem.* 273, 20728-20736.

Walker M. R., Kasprowicz D. J., Gersuk V. H., Benard A., Van Landeghen M., Buckner J. H, Ziegler SF (2003) Induction of FOXP3 and acquisition of T regulatory activity by stimulated human CD4+CD25− Tcells. *J Clin Invest* 112:1437–1443.

Wang Y., Fairley J. A., Roberts S. G. (2010) Phosphorylation of TFIIB links transcription initiation and termination. *Curr. Biol.* 20, 548–553.

Wang Z., Zang C., Rosenfeld J. A., Schones D. E., Barski A., Cuddapah S., Cui K., Roh T. Y., Peng W., Zhang M. Q., Zhao K. (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40(7): 897-903.

Whetstine J. R., Nottke A., Lan F., Huarte M., Smolikov S., Chen Z., Spooner E., Li E., Zhang G., Colaiacovo M., Shi Y. (2006). Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* 125, 467–481.

Williamson L. M. and Lees-Miller S. P. (2011) Estrogen receptor a-mediated transcription induces cell cycle-dependent DNA double-strand breaks. *Carcinogenesis* 32 (3) 279–285.

Wilson III D. M., Bohr V. A., (2007) The mechanics of base excision repair and its relationship to aging and disease, DNA Repair (Amst) 6, 544–559.

Winterbourn C. C. (2008) Reconciling the chemistry and biology of reactive oxygen species. *Nat Chem Biol* 4: 278–286.

Wu C. and Morris J. R. (2001). Genes, genetics, and epigenetics: a correspondence. *Science* 293, 1103-1105.

Wu, S. C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* 11, 607–620.

Xu W., Cho H., Evans R. M**.** (2003). Acetylation and methylation in nuclear receptor gene activation. *Methods Enzymol.* 364, 205-23.

Yamaguchi S., Hong K., Liu R., Inoue A., Shen L., Zhang K. and Zhang Y. (2013) Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Research* 23, 329–339.

Yudkovsky N., Ranish J. A. & Hahn S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature* 408, 225-229.

Zhang Y., Jurkowska R., Soeroes S., Rajavelu A., Dhayalan A., Bock I. (2010).Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res* 38: 4246–4253902.

Zhang Y., Yuan F., Presnell S. R. (2005) Reconstitution of 5′-directed human mismatch repair in a purified system. *Cell*; 122, 693- 705.

Zheng Y., Josefowicz S., Chaudhry A. (2010). Role of conserved non-coding DNA elements in the FOXP3 gene in regulatory T-cell fate. *Nature* 463 (7282): 808–812.

List of publications

1. Di Zazzo E., Feola A., Zuchegna C., **Romano A.**, Donini C. F., Bartollino S., Costagliola C., Frunzio R., Laccetti P., Di Domenico M., Porcellini A. *The p85 regulatory subunit of PI3K mediates cAMP-PKA and insulin biological effects on MCF7 cell growth and motility*. The Scientific World Journal 2014: 1-11. ISSN 2356-6140 DOI 10.1155/2014/565839.

2. Zuchegna C., Aceto F., Bertoni A., **Romano A.**, Perillo B., Laccetti P., Gottesman M. E., Avvedimento E. V., Porcellini A. *Mechanism of Retinoic acid-induced transcription: histone code, DNA oxidation and formation of chromatin loops*. Nucleic Acids Research 2015 Jan 1;42(17): 11040-11055. ISSN 1362-4962 DOI 10.1093/nar/gku823.

3. De Rosa V., Galgani M., Porcellini A., Colamatteo A., Santopaolo M., Zuchegna C., **Romano A.**, De Simone S., Procaccini C., La Rocca C., Carrieri P. B., Maniscalco G. T., Salvetti M., Buscarinu M. C., Franzese A., Mozzillo E., La Cava A., Matarese G. *Glycolysis controls the induction of human regulatory T cells by modulating the expression of FOXP3 exon 2 splicing variants*. Nat Immunol. 2015 Nov;16(11):1174-84. DOI 10.1038/ni.3269.

4. Russo G., Landi R., Pezone A., Morano A., Zuchegna C., **Romano A**., Muller M. T., Gottesman M. E., Porcellini A., Avvedimento E. V. *DNA damage and repair modify DNA methylation and chromatin domain of the targeted locus: mechanism of allele methylation polymorphism*. Scientific reports. 2016 Sept; 6:33222. DOI: 10.1038/srep33222.

5. Pezone A., Russo G., Tramontan A., Florio E., Scala G., Landi R., Zuchegna C., **Romano A.**, Chiariotti L., Muller M. T., Gottesman M. E., Porcellini A., Avvedimento E. V. *High-coverage methylation data of a gene model before and after DNA damage and homologous repair*. Sci Data (In Press).

6. *Histone and DNA methylation cycles are coupled by DNA oxidation to induce transcription by nuclear receptors* (Manuscript in preparation).