ALMA MATER STUDIORUM– UNIVERSITA' DI BOLOGNA

DOTTORATO DI RICERCA IN

BIOLOGIA CELLULARE E MOLECOLARE

Ciclo XXVII

Settore Concorsuale di afferenza: 05/E2

Settore Scientifico disciplinare BIO/11- BIOLOGIA MOLECOLARE

Natural compounds Camptothecin and Triptolide: highly specific enzyme inhibitors and tools to dissect transcriptional functions

Presentata da: Stefano Giustino Manzo

Coordinatore Dottorato: **Relatore:**

Davide Zannoni   **Giovanni Capranico**

Esame finale anno 2015

Index

ABSTRACT

The work of this thesis was mainly focused on the transcriptional effects of two strong and highly specific transcription inhibitors: Triptolide and Campthotecin. The two compounds are both natural products derived from Chinese medicinal plants. Although with completely different molecular targets, these two drugs show interesting properties such as anti-proliferative and anti-tumor activity. Additionally, these two drugs, as selective inhibitors of their targets (XPB and Top1 respectively), can be used to dissect the role of these proteins in transcriptional regulation.

Triptolide (TPL) is a diterpene epoxide derived from the Chinese plant *Trypterigium Wilfoordii* Hook F. This compound shows great immunosuppressive, anti-inflammatory and anti-proliferative activities, mainly due to its strong transcriptional inhibitory property. TPL inhibits the ATPase activity of XPB, a subunit of the general transcription factor TFIIH. One of the most relevant consequences of TPL treatments is the degradation of RNA Polymerase II in a dose and time-dependent manner. In this thesis I found that degradation of Rbp1 (the largest subunit of RNA Polymerase II) during TPL treatments, is preceded by an hyperphosphorylation event at serine 5 of the carboxy-terminal domain (CTD) of Rbp1. This event is concomitant with a block of RNA Polymerase II at promoters of active genes. The enzyme responsible for Ser5 hyperphosphorylation event is CDK7. Notably, CDK7 downregulation rescued both Ser 5 hyperphosphorylation and Rbp1 degradation triggered by TPL. Our data therefore clarify novel aspects of the transcriptional role of TFIIH and show how this complex can regulate RNA Polymerase II stability.

Differently from TPL that is currently used in clinical trials for treating cancer, camptothecin (CPT) is an already FDA-approved drug highly effective in the treatment of solid tumors such as ovarian and colon cancers. CPT specifically inhibits topoisomerase 1 (Top1). This enzyme is able to remove torsional stress created by physiological processes such as transcription and replication. Top1

introduces a nick in the DNA molecule and covalently binds it, allowing a controlled rotation of the cut strand around the uncut strand that removes the torsional stress. Top1 is able to relax both overwound and underwound DNA, usually indicated as positive and negative supercoiled DNA, respectively. CPT can interfere with this process by blocking and stabilizing Top1 on DNA leading to a formation of a *Top1 cleavage complex (Top1cc).* In this work, I revealed new important effects that CPT-induced Top1ccs have on transcription. We first found that CPT induced antisense transcription at divergent CpG islands promoter. Notably, this phenomenon is independent from replication, but depends on both Top1 and CDK9 kinase activity. Interestingly, by immunofluorescence experiments, CPT was found to induce a burst of R loop structures (non B-DNA structures composed by DNA/RNA hybrids stabilized by negative supercoils) at highly transcribed regions such as nucleoli and mitochondria. In collaboration with Frederic Chedin's lab, we tried to immunoprecipitate R loop structures after CPT treatments. Unfortunately, after drug treatment, these structures resulted highly unstable and difficult to isolate. We then decided to investigate the role of Top1 in R loop homeostasis through a short interfering RNA approach (RNAi). Using DNA/RNA immunoprecipitation techniques coupled to next generation sequencing I found that Top1 depletion induces an increase of R loops at a genome-wide level. We found that such increase occurs on the entire gene body and involves both a spreading and an increase in frequency of R loops. At a subset of loci R loops resulted particularly stressed after Top1 depletion. Notably these loci were frequently part of really long genes (>50 kb), in agreement with the role that Top1 has in regulating transcription and supercoils of very long genes. Interestingly, some of these genes showed the formation of new R loops structures (gain of R loop peaks), whereas other loci showed a reduction of R loops (loss of peaks). Interestingly we found that gain of peaks usually occurs at tandem or divergent genes in the entire gene body, while loss of R loop peaks seems to be a feature specific of 3' end regions of convergent genes. RNA seq and PolII-ChIP-seq revealed that only loci having loss of R loop peaks also showed an impairment of PolII progression and transcriptional termination. Thus our data revealed new and unexpected transcriptional role for Top1. Moreover we clarified

2

some aspects of role that Top1 has in homeostasis of R loops, showing how Top1 can modulate such structures.

All together, these findings demonstrated that transcriptional inhibitors are exquisite tools to investigate transcriptional mechanisms and disclose further potential approaches to develop novel therapeutics.

Chapter I

INTRODUCTION

Natural products from medicinal plants have always attracted the attention of the scientific community due to their many biological properties that make these compounds effective in treating several diseases such as cancer and inflammatory and auto-immune disorders. Among the thousand of compounds isolated from different plants and organisms transcriptional inhibitors are among the most effective in treating many disorders. Moreover, as the specific target of some of them is known, then those compounds can be used as a powerful tool to investigate transcriptional mechanisms and allow us to understand the physiological and pathological role of the investigated proteins.

1.1 Transcription and transcriptional inhibitors: from classic compounds to Triptolide

1.1.1 RNA Polymerase and transcription.

The transcriptional process is an essential mechanism for every living organism. This process consists of a synthesis of an RNA molecule (a structure with a ribose/phosphate skeleton) starting from a DNA molecule as template. The new synthetized RNA can represent the intermediate between genetic information and biological effectors as proteins, or alternatively can constitute a biological effector itself.

The enzyme able to catalyze the synthesis of RNA from a DNA template is the RNA Polymerase. Particularly the enzyme catalyzes the formation of covalent phopshodiesteric bonds between the 3' OH of a ribose sugar ring and a phosphate group in 5' position of the following ribose, trough the removal of a pyrophosphate

group from the new triphosphate ribonucleoside that is being incorporated into the nascent RNA molecule.

Eukaryotes possess three different RNA Polymerases: I, II and III.

1) **RNA Polymerase I** synthetizes the precursor of ribosomal RNA, that is processed into rRNA 28S, 18S and 5.8S. These RNAS have a structural role in ribosome biogenesis, therefore RNA Polymerase III activity is intimately linked to cell growth and proliferation (1)

2) **RNA Polymerase II** synthetizes the heterogeneous nuclear RNA (hnRNA) precurosor of the protein-coding mRNA, the regulatory short nuclear RNAs (snRNAs) and the long non codings RNA (lncRNAs) (2)

3) **RNA Polymerase III** synthetized tRNAs, rRNA 5S, snRNA U6 (involved in splicing), RNA 7SL (involved in secretion proteins synthesis) and the small RNA 7SK (inhibitor of the P-TEFb elongating factor). (3)

1.1.2 RNA Polymerase II, CTD and transcription phases

Altough RNA Polymerase I and III play essential role for cell proliferation and growth, transcription mediated by RNA Polymerase II produces the major variety of transcripts and its activity is probably the most tightly regulated. For this reason PolII mediated transcription was deeply investigated in the last twenty years.

RNA Polymerase II is a protein complex consisting of 12 subunits (composition valid for both human and yeast homologues) (4)

The most important feature of RNA Polymerase II (PolII) that characterizes and differentiates this polymerase from the other ones is the presence of a long and flexible carboxyl terminal domain (CTD) of the largest subunit RPB1.

This domain consists of multiple heptarepeats ($Y_1S_2P_3T_4S_5P_6S_7$) that are target of different post-translational modifications during the transcriptional cycle. The importance of this component can be understood if we compare its length with complexity of genomes: the larger is the genome the longer is the CTD (26 heptarepeats in *S. cerevisiae*, 32 in *C. elegans*, 45 in *Drosophila*, and 52 in mammals)

(5)(6). These sequences are target of different kinases and phosphatases that regulate CTD and RNA polymerase functions during transcription.

RNA Polymerase II-mediated transcription is a highly regulated process, that consist of multiple crucial steps:

1) ***Pre-initiation complex (PIC) nucleation:*** RNA Polymerase II associates with general transcription factors (GTFs) to form the PIC at the correct position on on promoter. GTFs have been widely studied and characterized. The first actor is TFIID. This complex composed by TBP (*TATA binding protein)* and TAFs (TBP-associated factors) is the core of the PIC and its binding to DNA is essential for a correct nucleation of the complex. TFIID is helped by TFIIA to recognize the promoter sequences. TFIIB stabilizes the interaction between TFIID and PolII-TFIIF (7). The latter is able to associate tightly to the polymerase allowing the interaction with TFIID/IIA/IIB (8). TFIID, once bound to DNA with the other factors of PIC, interacts via TAFs with the transcriptional co-activator Mediator. This important complex recognizes the unphosphorylated CTD of RNA Polymerase II favoring the polymerase loading to the DNA.

2) ***Promoter clearance:*** The Mediator complex, after binding to the unphosphorylated CTD cooperates with TFIIE to strongly stimulate the kinase activity of TFIIH on the CTD itself. The complex is crucial for the promoter clearance of PolII and the transition from the pre-initiation step to the initiation phase (9). TFIIH is a complex composed of ten subunits: one core complex composed by XPD, XPB, p62, p52, p44, p34 and TTDA; and a cyclin dependent sub-complex composed by CDK7, MAT1 and cyclin H. The former is responsible of the melting of promoter and formation of the *open complex* trough the ATPase activity of XPB (10), the latter favors *promoter escaping* by phosphporylating Ser5 on CTD of PolII via CDK7. This modification commits RNA Polymerase II to a more processive synthesis and at same time leads to dissociation of Mediator complex from PolII CTD (11). Particularly, it has been shown that levels of Ser5 remain high during the synthesis of first hundreds of nucleotides and then decline further

downstream the genes (12). During the initiation phase synthesis of the RNA molecule is an inefficient process that can easily be concluded with premature termination and transcription abortion. At this step, there is a sort of equilibrium between premature termination and productive RNA synthesis and by phosphorylating Ser5, TFIIH shifts this equilibrium therefore suppressing this instability. The presence of TFIIF enhances also. This process. For this reason TFIIH and IIF can be considered at the same time both initiation and elongation factors. (6).

3) ***Pausing and elongation:*** Once the RNA chain reaches the length of about 15 nucleotides the initiation complex become stable and the risk of transcriptional abortion dramatically decreases. At this step RNA polymerase II usually does not enter in a productive RNA extension, instead remaining in a *paused conformation.* In this conformation PolII activity is highly regulated by positive and negative elongation factor that can modulate processivity of polymerase. The most important factors that regulate pausing and elongation of PolII are P-TEF-b (Positive Transcriptional Elongation Factor b) and DSIF/NELF complex. P-TEFb is a cyclin-dependent CTD kinase composed of Cdk9 and one of several cyclins including T1, T2 and K (13). This important factor usually phosphorylates the CTD of PolII at level of Ser2 promoting the release of polymerase from pausing site and allowing a productive elongation. CTD phosphorylation by P-TEFb is required to prevent arrest of elongating pol II. DSIF (DRB sensitivity inducing factor) and NELF (negative elongation factor) cooperate to bind PolII-CTD and preventing elongation inducing transcriptional arrest. By phosphorylating Ser2, P-TEFb impedes the association of the negative complex DSIF/NELF to the CTD of RNA Polymerase II, thus favoring the elongation. Additionaly, it has been shown that P-TEFb also catalyzes the phosphorylation of hSpt5, subunit of DSIF promoting the disruption of DSIF/NELF complex (14). In *in vitro* experiments Rna Polymerase II is able to catalyze the synthesis of RNA with a rate of 100-300 nucleotides/minute. The *in vivo* rate is instead dramatically higher: 1500nt/minute. This notable increase is basically due to the presence of

7

different elongation factors that stimulate the processivity of PolII. These proteins can prevent pausing of PolII during transcription or alternatively reactivate and arrested polymerase:

- TFIIF, even if considered as an initiation factor, this protein is able to re-associate to the transcription elongation complex (TEC) when Pol II is stalled (15)

- Elongin complex, it's a complex able to suppress PolII pausing trough realignment of the 3'OH of the nascent transcript misaligned in the catalytic site of an arrested PolII.

- ELL family, similarly to the previous two complexes these proteins can suppress PolII pausing, particularly at really long genes.

- CSB (Cockayne Syndrome B) protein involved in TC-NER able to stimulate and reactivate a stalled Pol II in proximity of a DNA lesion.

- TFIIS, this factor stimulates intrinsic endonucleasic activity of PolII allowing cleavage of the nascent transcript and the creation of a new 3'OH in the RNA chain correctly aligned in the catalytic site (16). This factor contributes to the proof reading activity of PolII.

There are also other proteins that can affect elongation of RNA Polymerase II, even if they are not properly elongation factors. Notably RNA Polymerase II has to deal with a complicated "architectural complex" such as chromatin. As nucleosomes basically act to repress transcription, chromatin remodeling factors and histone modifiers can affect PolII elongation. Among the chromatin remodeling factors we can find: SWI/SNF, CHD1, FACT and hSpt6. Many of them are able to remove nucleosome during PolII elongation. As regards epigenetic modifiers the most important are Elongator, Set2, Set1 and PAF complex. They usually interact with the phosphorylated CTD triggering histonic modifications (acetylation, H3K4 methylation) that induce an open state of chromatin.

4) **Termination:** this process occurs when PolII decreases progressively its productivity, therefore interrupting the synthesis of RNA. Termination usually occurs co-trascriptionally trough recognition of sequences located at

3' end of transcript that can act *in cis*. In yeast we can distinguish two kind of termination:

- *PolyA-dependent termination*: According to *the torpedo model* a polyA signal located at 3' end of gene is first transcribed by PolII. The latter is then paused and the nascent transcript is cleaved. The upstream cleavage product is polyadenilated, while the downstream cleavage product is degraded. The 3' end processing reaction starts when the AAUAAA signal present in the nascent RNA molecule is recognized by the cleavage and polyadenilation factor CPF through the interaction with PolII body, therefore inducing PolII stalling. Immediately later, the cleavage stimulatory factor CF1A recognizes a GU rich element downstream of polyA site, recruits CPF, thus releasing its hold on PolII body. CPF accompanied CF1A to the PolII-CTD and the latter enhance cleavage reaction performed by the former, therefore allowing release of paused PolII. An efficient release of PolII pause can occur only in the presence of a 5'-3' exoribonuclease Rat1 in complex with Rai1 and Rtt103, this complex is able to recognize the unprotected 5' end generated by the cleavage reaction allowing the degradation of the RNA coming out from the exit channel of PolII. Collision of Rat1 with PolII eventually leads to termination. (17)

- *Sen1-dependent termination:* this mechanism regards the processing of 3' end of snRNA (*short nuclear RNA*) and snoRNA *(small nucleolar RNA),* but not mRNA. The molecular actors in this pathway are different: the reaction cleavage is performed by the nuclear exosome TRAMP complex, Nrd1 and Nab3 bind RNA and the putative RNA/DNA helicase Sen1 promote termination by unwinding the RNA/DNA hybrid in the active site of PolII.(18)

As regards mammals, the termination mechanism involves recognition of a polyadenilation signal (PAS). Most part of protein involved in this pathway are the human homologs of the yeast pathway (hCPSF for CPF, hCstF for CF1A, hXRN2 for Rat1). Although for some mammalian genes, a functional PAS

signal is sufficient for an efficient termination, for many others additional sequences are required. In mammals two kind of additional terminators have been identified involving two different termination mechanisms:

- *Pause element terminators:* these regions are located downstream of the PAS and are usually G rich. At level of this region the nascent transcript form RNA/DNA hybrid structures called R loops, probably inducing a slow down of PolII. Senataxin (human homolog of Sen1) subsequently resolves these structures. This allows access of XRN2 at polyA cleavage sites, 3' transcript degradation and PolII termination (19)

- *Co-Transcriptional cleavage terminators (CoTC)*: In this mechanism the cleavage to create an entry site for XRN2 does not occur at PAS but at region called CoTC element. Degradation at 3' end of RNA is performed by XRN2 and this lead to release of PolII from chromatin template with pre-mRNA associated. Subsequently PAS cleavage induces dissociation from PolII. (20)

It is not clear if in mammals there is a PAS indipedent mechanism like that onevalid for snoRNA in yeast.
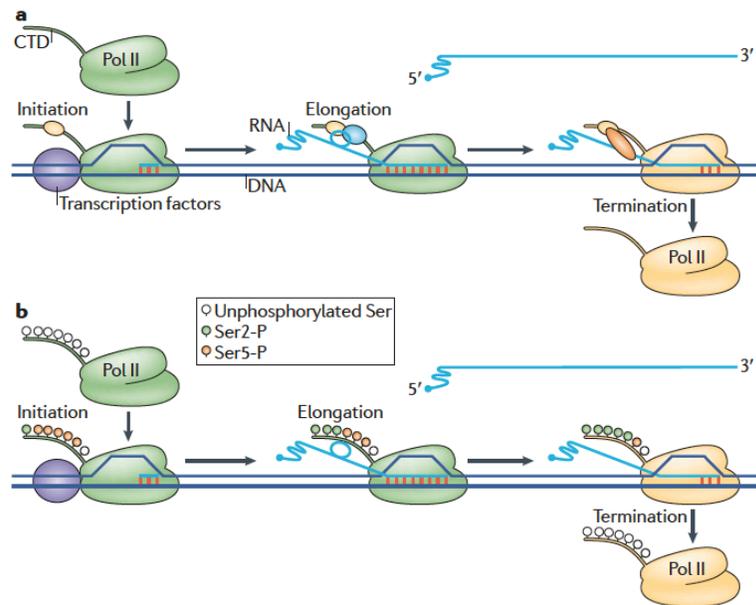

Fig1. Transcriptional steps and regulation of CTD-phosphorylation. Figure from Unravelling the means to an end: RNA polymerase II transcription termination. Jason N. Kuehner, Erika L. Pearson and Claire Moore. Nature Reviews Molecular and Cellular Biology, May 2011

1.1.3 Principal transcriptional inhibitors

The above-mentioned steps of transcription have been deeply investigated and well characterized. One important tool, used to understand the molecular mechanisms underlying the transcriptional process, came from the use of specific transcriptional inhibitors. These compounds, frequently purified as natural products from several plants or fungi, were often used as potent anti proliferative agents. Some of them are used as chemotherapic agents in FDA-approved therapies or they are currently used in clinical trials. Once, their specific targets have been identified these compounds became powerful tools to investigate molecular mechanisms.

a-Amanitin
Alpha-Amanitin is one of the most powerful transcriptional inhibitors. It' s a toxin derived from fungi that are part of gender *Amanita* (*A. phalloides, A. bisporigera*)*.* This compound is able to inhibit both RNA polymerase II and III, even if with different sensitivity. The most sensitive enzyme is RNA Polymerase II that shows an IC_{50} of 0,02 mg/ml, while for RNA Polymerase III this value increase of about 100-folds. Instead, RNA Polymerase I is not sensitive to this drug (21). Alpha-Amanitin binds in a pocket really close to the "bridge helix" responsible for the translocation of RNA Polymerase during the synthesis of the RNA molecule. One of the main molecular effect of a-Amanitin is the degradation of Rbp1, the largest subunit of RNA Polymerase II, without affecting stability of other subunits (Rpb5 and 8). The degradation of Rbp1 could explain the irreversible effect of alpha-amanitin. (22).

Actinomycin D
Actinomycin D is an antibiotic polypeptide produced by bacteria *Streptomyces*. It's a really strong transcriptional inhibitor. It was one of the most common chemotherapic drug, used to treat different kinds of cancer. Now it has been put aside from current chemoterapic protocols, because of its elevated cytotoxicity.

Transcription by all three eukaryotic polymerases is usually affected during Actinomycin treatment, even if with different sensitivities. The most sensitive is RNA Polymerase I mediated transcription, while the least is transcription mediated by RNA Polymerase III. Actinomycin is an intercalating agent that preferentially binds GC rich sequences. It's also able to inhibit DNA topoisomerase-I probably impeding RNA polymerase progression. Actinomycin D generates DNA double-strand breaks marked also by phosphorylation of histone H2AX. Transcriptional inhibition seems to be reversible. Actinomycin induces hyperphosphorylation of CTD of PolII likely trough hyperactivation of P-TEFb.(23)

Flavopiridol and DRB

These two compounds can be mainly classified as CDK9 inhibitors. Several drugs can be included in this category, and a unique mechanism of action is shared among them: competition with ATP for the kinase active site. Since several cyclin-dependent kinases share with CDK9 similarity in the conformation of catalytic site, these compound show activity for other CDKs, even if at lower level compare to CDK9.

Flavopiridol is a flavon chemically synthetized from an alkaloid isolated from leaves of Indian plants such as *Amoora Rohituka and Dysoxylum binectriferum*. Flavopiridol is considered the most efficient CDK9 inhibitor. It binds the ATP binding site and induces a structural change in the kinase. This compound shows activity also for CDK1, CDK4 and CDK8, even if lower affinity. In leaving cells, Flavopiridol efficiently inhibits CDK9 in the range of 100-300 nM (24). Moreover, it is a water-soluble compound and for this reason very useful in molecular biology studies.

DRB (5,6-Dichloro-1-beta-Ribo-furanosyl Benzimidazole) is the second most used CDK9 inhibitors, nevertheless it's poor solubility in water and and the high effective concentration (100 μM). This compound was used to identify the role of P-TEFb and factor DSIF/NELF in regulating PolII pausing and elongation (25). It inhibits CDK7 too, with a 3 fold lower affinity. It's really fast and able to inhibit transcription in the

range of minutes. For this reason this compound can be used to measure transcription rates in run on experiments.

1.1.4 Triptolide: Effects and mechanism of action

Triptolide (TPL) has recently attracted the attention of the scientific community due to its many interesting properties. Triptolide is the major derivative of the Chinese plant *Trypterygium Wilfordi* Hook F., commonly used in traditional Chinese medicine to treat inflammation or inflammatory diseases such as rheumatoid arthritis (26). It's a diterpene triepoxide (Fig 2) with a unique molecular structure, showing several interesting pharmacological properties: anti-inflammation, immunomodulation, anti-angiogenesis, anti tumour and pro-apoptosis.
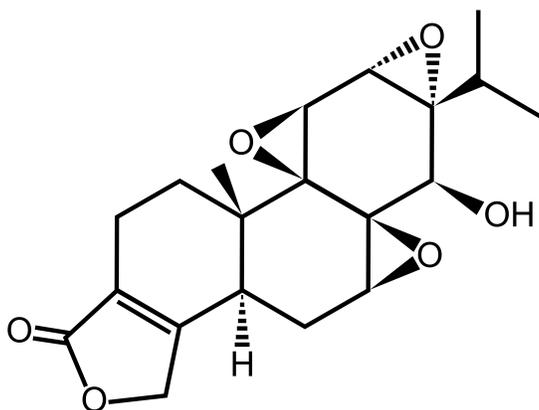
Fig2. Strucure of Tripolide,. *Figure from an online open source.*

Pre-clinical studies show that Tripolide is effective against different kinds of cancer, allograft rejection, and arthritis. Recently, it has entered in human clinical trials for cancer treatment.

Triptolide was first investigated for its anti-inflammatory activity, that was basically explained by the ability of TPL to inhibit the transcriptional pathway controlled by NFkB, responsible for regulation of cytokines important for inflammation, as IL-2,

IL-8 and IFN-γ (27).The antiptroliferative and pro-apoptotic effects of TPL can be partially explained by different properties: induction of caspase response, downregulation of Bcl-2 and upregulation of Bax (28)(29) Recently it has been shown that most part of pharmacological effects of TPL can be explained with the properties that tripolide has as a general transcriptional inhibitor (30). Triptolide inhibits RNA PolII and RNA PolI-mediaetd transcription, but the most rapidly affected is the former (30). Triptolide affects nuclear and nucleolar structure, as a consequence of Pol I and PolII activity inhibition: nucleoli disaggregate after TPL treatments and nuclear speckles (foci with actively transcribing regions, enriched in splicing factor and CDK/cyclin complexes) change in shape.(31) One of the most important molecular effect is the proteasome-mediated degradation of Rbp1, the largest subunit of RNA Polymerase II. This degradation is time and dose-dependent (30). If and how Triptolide affect phosphorylation of PolII CTD, and if this modification affect PolII stability still remains unclear and controversial (31) (32).

Recently, *Titov et al* identified, in an elegant work, the target of Triptolide: XPB, a subunit of TFIIH complex. Triptolide binds covalently XPB, thus inhibiting its ATPase activity. Interestingly, inhibiting ATPase activity of XPB does not block its helicase activity, demonstrating for the first time that these two properties are independent. XPB is part of TFIIH (33). This factor is involved both in transcription initiation and nucleotide excision repair (NER). Triptolide was indeed found to inhibit both transcription and NER. (33).

Even if the TPL target was well characterized, a lot of questions still have to be answered: How does TPL affect PolII dynamics on chromatin? How does Triptolide affect phosphorylation of PolII? Is this important for RNA Polymerase II stability? What are the enzymes involved in TPL induced Rbp1 degradation?

1.2 Top1, Camptothecin and R loops

1.2.1 DNA topology: a critical feature for cellular processes

The most common and stable DNA structure is B-DNA: an antiparallel double-helical structure. This conformation can be intrinsically considered as a double-edged sword: the presence of two strands wound one around the another, on one side, gives incredible stability of the molecule and ensures that genetic information can be maintained and replicated correctly; but, on the other side, processes that require strand separation, such as replication and transcription, lead to DNA over-winding and strand entanglement (34). Additionally, at intracellular level, DNA is not a free and relaxed molecule. First of all, DNA is associated to proteins like histones to form nucleosomes, while several other proteins (transcription factors, polymerases…etc) contribute to constitute a well-structured DNA-protein complex called chromatin. Chromatin is organized in structured levels; each of them is well defined, relatively stable and highly regulated. Chromatin state can change according to the cellular phase or cellular regulation: modulating chromatin packaging regulates accessibility of DNA therefore affecting also transcriptional response and gene expression. Third, DNA is also anchored to nuclear membrane and this interaction is essential in maintaining nuclear structure. On the basis of these considerations, we know that DNA is not a free and relaxed molecule; instead it is highly constrained by several interactions.

In the presence of these constraints cellular processes like replication, transcription, chromosomal segregation and chromatin remodeling create a torsional stress on the DNA molecule leading the DNA double helix to coil around it self, generating a *supercoil.*

According to DNA helix directionality and the type of torsional stress created, a DNA molecule can be over-wound, (or *positively super-coiled)* or under-wound *(negative supercoiling)* (35). With the exception of few hyperthermophilus organisms, for all the other organisms the genomic DNA is slightly negatively supercoiled on average. A negative supercoiled DNA tends to melt easily, therefore facilitating cellular

processes that need the DNA to be open and in a single strand form (such as replication and transcription). Instead, a positively supercoiling hinders these processes, slowing down their progression and making DNA more difficult to melt. For example, using an *in vitro* system, it has been shown that DNA replication machinery can replicate only on negatively supercoiled DNA. The DNA polymerase stalls before replication is complete, and this is probably caused by the accumulation of positive supercoilings in front of the replication machinery. Importantly, the replication can be restored upon the addition of a topoisomerase able to remove positive supercoiling (36).

On the other side, also negative supercoil can be dangerous for cell, and therefore they have to be highly regulated. As told previously, a negative supercoiled DNA tends to melt easily, this can favor the re-annealing of the RNA molecule during the transcription phase leading to the formation of an R loop, a structure that is frequently cause of genomic instability (37). This means that DNA topology can have great influence in cellular process and has to be therefore highly regulated. This is why both prokaryotic and eukaryotic organisms have developed several mechanisms and conserved several enzymes to overcome topological problems. Particularly, to bypass problems generated by the presence of supercoils, knots and interwound, one of the DNA molecule has to be cleaved allowing a strand rotation that relaxes the double helix removing supercoils. Enzymes able to catalyze such reactions are essential for the cell life. Moreover by introducing transient nick in the DNA molecule, these enzymes are potentially deleterious and have to be finely regulated. The enzymes able to modify DNA topology are named topoisomerases.

1.2.2 DNA topoisomerases

DNA topoisomerases are enzymes able to remove or introduce positive and negative supercoilings in the DNA molecule, to catenate/decatanate two molecules of DNA or to knot/unknot DNA. To perform these tasks, DNA topoisomerases have to introduce one or two breaks on DNA and pass one strand of the DNA throughout a break in the other strand or pass a region of the DNA duplex from the same or a

different molecule throughout a double-strand break (DSB). In this reaction DNA cleavage is en essential step, usually accomplished by forming a transient phosphor-tyrosine bond between the tyrosine of the enzyme active site and a 3' or 5' end of the DNA molecule. During this catalysis the enzyme modifies the DNA topology and subsequently religate and release DNA.

In nature there are four classes of DNA topoisomerases, with different properties and structures.

We can mainly distinguish topoisomerases in two groups: Type I and II.

- ***Type I Topoisomerases*** cleave one DNA strand only. This category can additionally be split in two subfamilies: type IA topoisomerases form transient phosphor-tyrosin bond with the 5' phosphates of DNA ends, whereas type IB topoisomerases do with 3' phosphate. *Type IA* subfamily includes: bacterial topoisomerase I and III, archeal reverse gyrase, and eukaryotic topoisomerase III. These enzymes have preference for relaxing highly negative supercoiled DNA molecules. (38). Reverse gyrase has unique properties and it's the only enzyme able to introduce positive supercoils (39). *Typ1 IB* subfamily includes eukaryotic Topoisomerase I, poxvirus topoisomerase I and some homologous in bacteria. (40)

- ***Type II topoisomerase*** cleave both strands of a double helix generating a double strand break. Also this category can be divided in two subfamilies on the basis of their aminoacidic sequence and structure: IIA and IIB. *Type IIA topoisomerases* are really common in nature, while *Type IIB* seem to be exclusive of Archea and Plants. *Type IIA subfamiliy* includes: eukaryotic topoisomerase II (TOP 2), bacterial DNA gyrase and topoisomerase IV, prokaryotic topoisomerase II. These enzymes can resolve both positive and negative supercoils, and, additionally, they can decatenate and unknot DNA. Both type II topoisomerase categories are multisubunit enzymes, and have a similar mechanism of action: double strand breaks performed by the enzyme covalently linked to 5' ends of DNA. Topoisomerase II usually catalyzes topological modifications by passing a second DNA duplex through the

cleavage of another DNA segment. This mechanism requires ATP and Mg^{2+}. (40)

Taken together, all topoisomerases can resolve any topological problem created by main cellular processes:

1) *Replication and Segregation of newly replicated chromosome:* during the replication process, an ongoing replication fork usually generates positive supercoils ahead of itself, on the unreplicated DNA molecule. These supercoils can hinder replication. If the replication machinery is allowed to rotate these positive supercoils can partially dissipate, but as a consequence of the fork rotation, duplicated DNA molecule will intertwine. As regards positive supercoils, these topological problems can be resolved by a Type IB or II, but for the intertwined duplicated molecules, a Type II topoisomerase is essential (41) At the end of replication step, two replication forks usually collide and the problem of decatenating the new synthetized DNA molecules arises. Type IB topoisomerase cannot accomplish to this task. Therefore, a type II topoisomerase activity is required. In general, in eukaryotes Top II seems critical for chromosomal decatenation (41).

2) *Transcription:* according to the *supercoiled twin-domain model* (42), during transcription an RNA Polymerase translocate along DNA generating positive supercoils ahead and negative supercoils behind it. Both kinds of superocoils have to be removed for a correct progression of elongation. Positive supercoils represent a physical barrier to the progression of RNA Polymerase, while negative supercoils can favor R loop formation that also affects polymerase elongation. Type I B topoisomerases are the major class of topoisomerase associated with transcription. However, yeast strains deleted for Top1 do not have an impaired transcription process, which only appear in the presence of double Top1 and Top2 deletions (43)

3) *DNA Recombination:* In this process a Type IA topoisomerase seems to play an essential role and this is probably why all the organisms possess at least one Type IA topoisomerase. Both in bacteria and yeast it has been shown that

18

removal of this kinds of topoisomerase increase genomic instability mediated by proteins involved in recombination (44)

4) *Chromosomal condensation and structure:* topoisomerases can be also involved in these two processes. Chromosomal condensation may generate torsional stress that has to be dissipated. Additionally chromosomal decatenation and condensation are strictly linked. Therefore, topoisomerase II have been found to be essential for these processes. Moreover it has been shown that topoisomerase II are part of a scaffold that constitute mitotic chromosomes.

1.2.3 Human DNA Topoisomerase I: structure and mechanism of action

The human DNA Topoisomerase I (Top1) is probably the most studied and well characterized type IB topoisomerase. It's a protein of about 91 kDa basically composed of 4 domains (Fig 3). The *N-terminal domain* has four NLS, nuclear localization signals. This domain is not important for the catalytic activity of the enzyme; it is the site of interaction with several other proteins (40) and it is poorly conserved. Following the N-term domain we can find the *core domain*, highly conserved, really important for the interaction that Top1 realized with DNA. It includes all the residues that compose the active site, with the exception of the tyrosine catalytic residue (Tyr273) (40). This latter is instead part of the *C terminal domain*. A *linker*, dispensable and poorly conserved, connects core domain and C
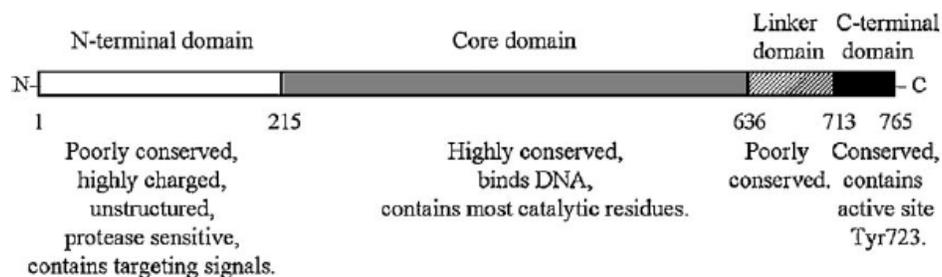

Fig3. Domain structure of human DNA Topoisomerase I. Fig from Champoux, J.J., DNA topoisomerases: structure, function, and mechanism. Annu. Rev. Biochem., (2001). 70: p. 369–413

ter

m domain.

As regards the mechanism of action the enzyme act as a clamp that allocates DNA in the middle. The core domain can be theoretically divided in three sub-domains (I, II, III). Sub-domain I and II form one side of the clamp (called cap), while the other part of the clamp is basically composed by sub-domain III and C term domain. Such structure seems to give to Top1 higher affinity to supercoiled DNA compare to relaxed DNA (45). The mechanism of action of Top1 is as simple as effective: a nucleophilic attack of an oxygen of Tyr273 is realized on the scissile phosphate. This creates a phosphodiesteric bond between Tyr273 and a 3' phosphate, leaving a 5'OH free. At this step Top1 is covalently linked to one strand of DNA and can therefore allow the rotation of this strand around the other to remove supercoil. This is performed using the intrinsic free energy of the DNA molecule and this is why Top1 does not necessitate of ATP. The strand rotation catalyzed by Top1 is considered a "controlled rotation", since some interactions of the linker and the core domain slowed down this rotation (46) (Fig 5). The relegation step needs that the 5' OH is realigned with Tyr273. In normal condition this reaction is thermodynamically favored and this is why the cleavage intermediates are transient and really fast.
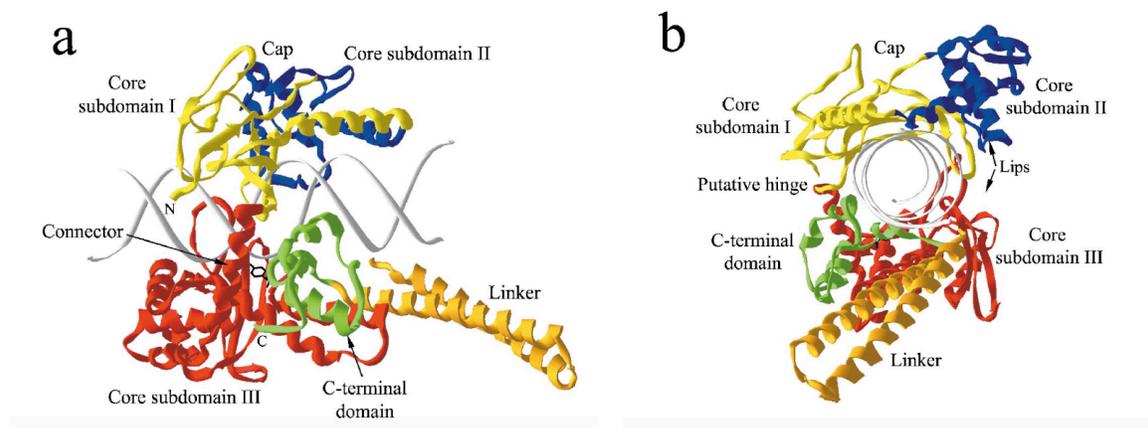
Fig 4. *Two views of structure of human DNA Top1 in the presence of DNA.* A)View from the side with the DNA axis horizontally oriented and B) View looking down the axis of the DNA. *Figure from Champoux, J.J., DNA topoisomerases: structure, function, and mechanism. Annu. Rev. Biochem., (2001). 70: p. 369–413*
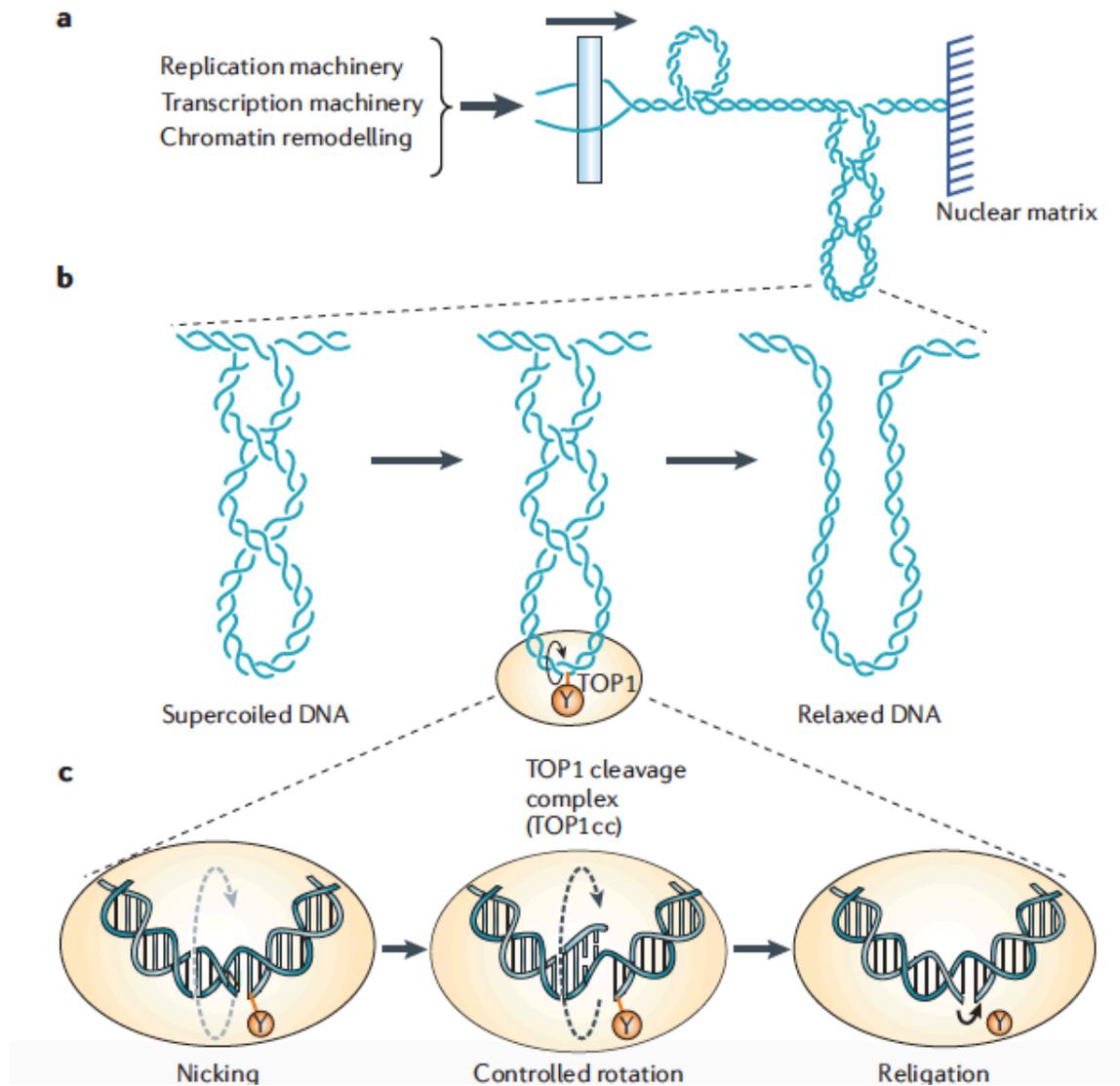
Fig5. Role and Mechanism of action of DNA Top1. *A) Torsianl stress created by physiological processes such as DNA replication, transcription and chromatin remodelling. b) The introduction of DNA single-strand breaks (nicks) by TOP1 provides allowe the rotation of the intact DNA strand around the break and facilitate DNA relaxation. c) An expanded view of DNA relaxation by a TOP1 cleavage complex (TOP1cc). Figure from Pommier, Y., Topoisomerase I inhibitors: camptothecins and beyond. Nat. Rev. Cancer, (2006). 6(10): p. 789-802.*

1.2.4 Transcriptional role of Top1

The link between Top1 and transcription is clear and evident. First, binding sites and occupancy of Top1 has been found enriched at active transcribed regions. (47), (48). In yeast neither Top1 nor Top2 are essential for a correct transcription, but double mutants show great accumulation of negative supercoils on transcribed regions (51). Therefore, the Top1 role in transcription fits well with the supercoiled twin-domain model (42).

Top1 has been shown to act also as a transcriptional activator or repressor. Particularly as regards transcriptional activation, it seems that Top1 favor the binding of TFIID-TFIIA complex on TATA box sequence. Strikingly, also an inactive mutant of Top1 can favor this process (50). A potential explanation for this phenomenon is that the association of TOP 1 in the initiation complex can be necessary for the following elongation phase with the function of relieving transcription-generated supercoils. If Top1 is not loaded, transcription cannot start, as it would be probably affected by the absence of this enzyme.

The best way to characterize the role of Top1 in transcription is to perform loss of function experiments. Unfortunately knockout of Top1 in mammals is not viable. The information that can be obtained with l.o.f. experiments are by short interfering RNA approaches. It has been shown that a cell line stably expressing a shRNA for Top1 show genomic instability and altered transcription for specific genes. Notably, in this cell line Top2 can partially compensate for absence of Top1 (51). In neurons, it has been found that RNA interference of both Top1 and Top2 affect transcription of really long genes, showing an important role for topoisomerases in resolving supercoils on these transcribed units. (52).

1.2.5 Inhibition of Top1 by Camptothecin: mechanism of action and molecular effects

Top1 can be targeted and specifically inhibited by Camptothecin (CPT). This compound is extracted from the Chinese plant *Camptotheca acuminata* and was known for its antitumor activity before the identification of Top1 as its specific target (47).

CPT and its derivatives (irinotecan, topotecan) (Fig 6) are able to interfere with cleavage/ligation reaction catalyzed by Top1. This reaction is usually really fast and undetectable at physiological level. In the presence of CPT the half-life of the cleavage intermediate, defined as *Top1 cleavage complex* (Top1cc), is prolonged and
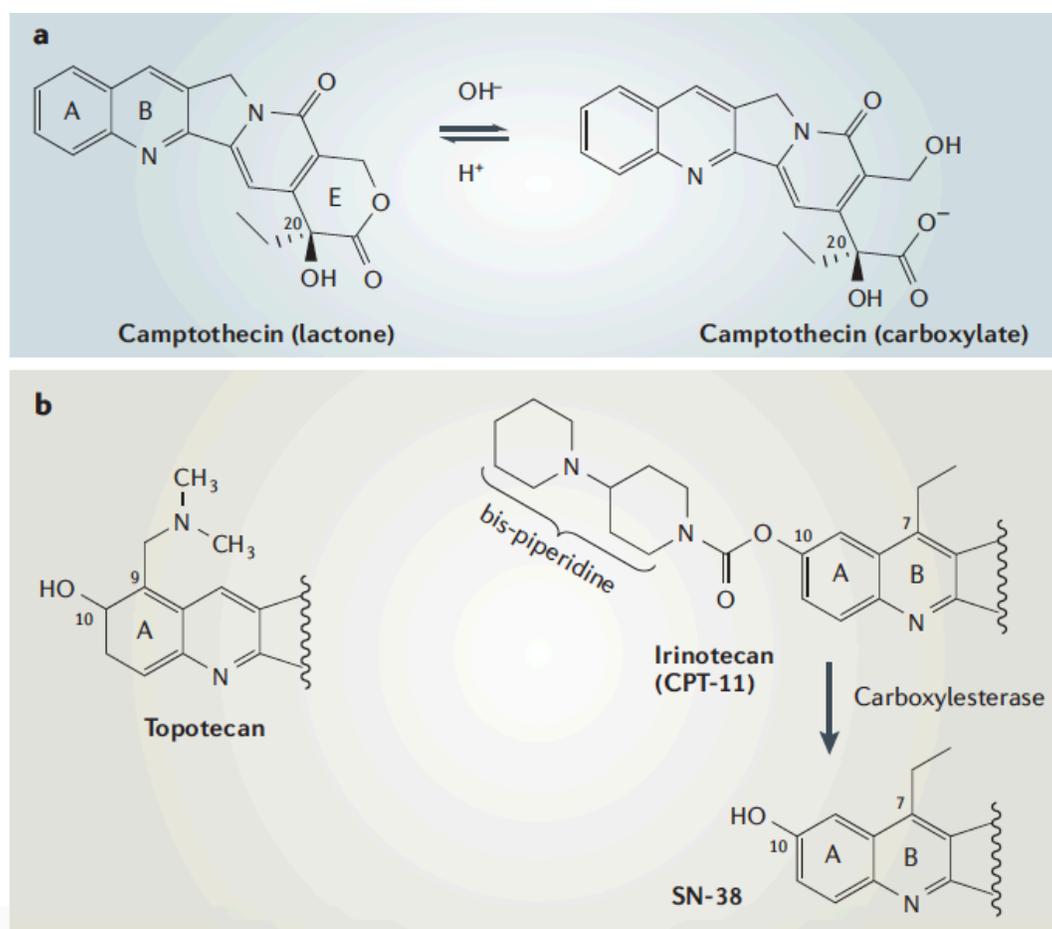

Fig 6. Camptothecin and derivatives. *Figure from* Pommier, Y., *Topoisomerase I inhibitors: camptothecins and beyond.* Nat. Rev. Cancer, (2006). 6(10): p. 789-802.

the Top1cc results stabilized. This kind of inhibition is specific for Top1 and reversible. As CPT can entrap Top1 on DNA in a reversible manner, this drug and its derivatives are considered *Top1 poisons*. The mechanism of action of CPT has been deeply investigated, although not fully understood. First, CPT is active on Top1 bound to DNA and not on the free protein, demonstrating that it is a non-competitive inhibitor of Top1. It is thought that CPT is able to intercalate in the pocket formed between Top1 and DNA, at level of the cut strand and could induce a misalignment of the 5'OH free end that has to attack the phospho-tyrosine bond to complete the religation step (47) (53) Top1-cc are strongly stabilized by CPT, but they can be present also at physiological levels in the presence of DNA adducts, mismatches, pre-existing lesions and ribonucleotides in the template (47).

The antitumor effects of CPT can be explained with the ability of this drug to induce DNA damage. Anyway since Top1cc stabilized by CPT is transient and reversible, it is not the drug itself that introduce a lethal damage of the genome. It is instead the Top1cc that induces a DNA damage by interfering with processes such as replication and transcription. Campthotecin specifically kills cells in S-phase as Top1cc colliding with an ongoing replication fork triggers the formation of a DSB on the template. The DSB is the lethal event that eventually leads to genome instability and/or cell death (54) (Fig 7 a).

The cellular consequences of this event are an S-phase specific cell killing and an arrest at G2 phase. This is obviously regulated by a molecular cascade triggered by the replication-induced double strand breaks: activation of checkpoint kinases ATM, ATR and DNA PK. These key regulators of the cellular cycle checkpoint phosphorylate, among their several targets, CHK2 and CHK1 and the marker of DBS DNA damage g-H2AX. This cascade continues with CDC25 phosphates inhibition, p53 activation and cell cycle arrest (47).

Considering the important role that Top1 has in transcription it is not surprising that CPT has great influence on this process. Synthesis of the 45S precursor rRNA is impaired during CPT treatment (55). Transcriptional inhibition by CPT seems to be an early effect of the drug, and involves both Pol I and Pol II mediated

transcriptions. For both kinds of transcriptions, camptothecin seems to affect the elongation phase (55), (56). Interestingly, Top1 inhibition by CPT has important effects on RNA Polymerase II. During CPT treatments PolII is hyperphosphorilated at ser5 of CTD and this phosphorylation is mediated by CDK7. Interestingly this hyperphosphorylation does not impair PolII stability (57).
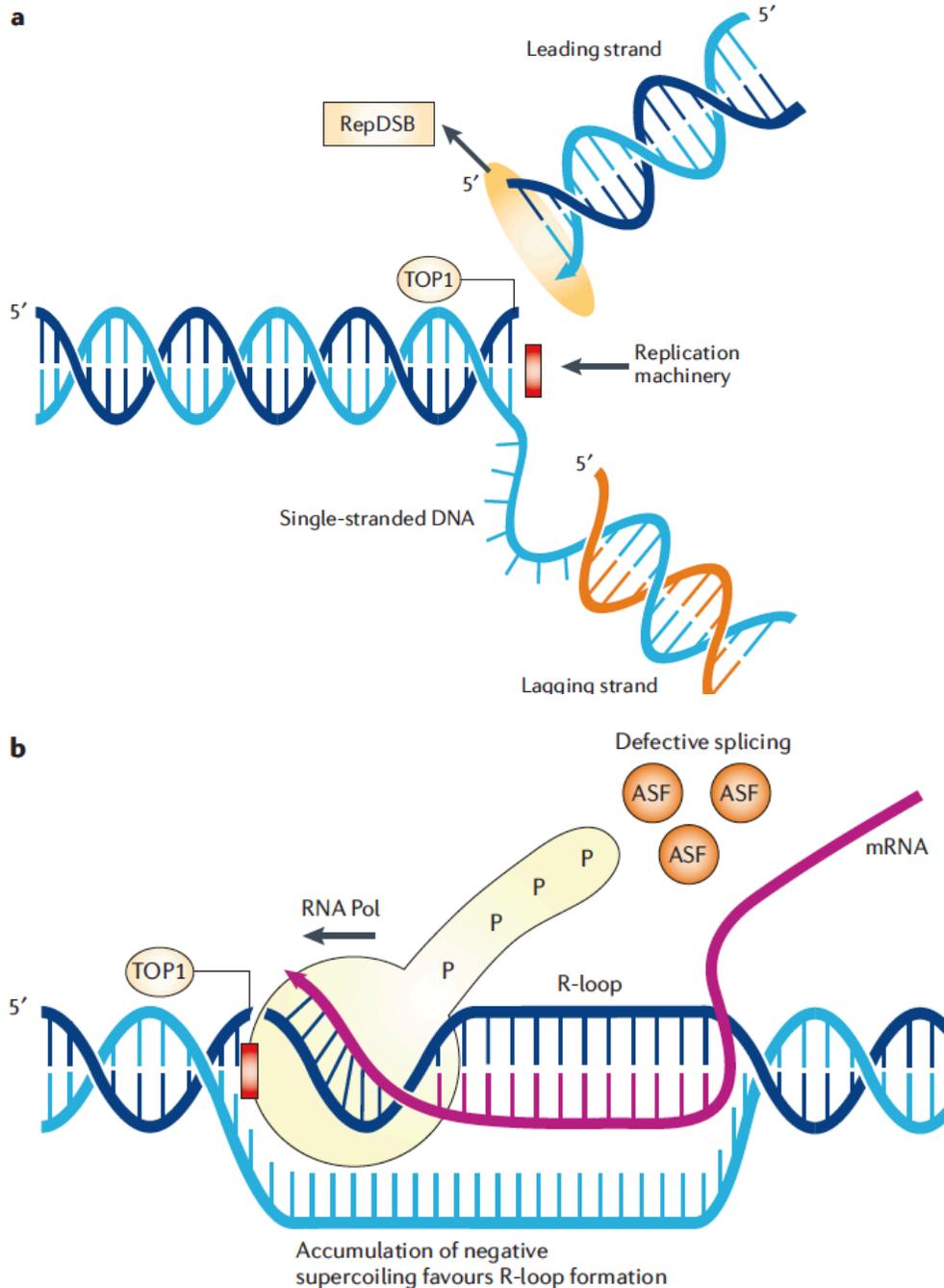

Fig 7. Interference of Top1 cc with a)Replication and B) Transcription. *Figure from* Pommier, Y., *Topoisomerase I inhibitors: camptothecins and beyond.* Nat. Rev. Cancer, (2006). 6(10): p. 789-802.

25

Collision of a transcribing RNA polymerase II on a Top1cc can make this intermediate irreversible (Fig 7 b). How much is the contribution of this phenomenon to DNA damage, genomic instability and therefore drug cytotoxicity in replicating cells is not well understood. Inteserestingly, cells defective in TC-NER (*transcription coupled–nucleotide excision repair*) are hypersensitive to CPT. (58)

One molecular mechanism, strictly related to transcription, is the degradation of Top1 during camptothecin treatment. Once that Top1 is entrapped to DNA within few minutes of CPT treatment, it is modified trough ubiquitination and sumoylation and degraded via 26s Proteasome (59). The ability to degrade Top1 in the presence of CPT correlates with the cell line resistance to the drug. Interestingly, this ubiquitination is strictly dependent to transcription (60) and performed by several such as the tumor suppressor BRCA1 (58).

However the transcriptional effects of CPT are not confined to the above-mentioned. Some of the transcriptional effects of CPT are really fascinating and not fully understood. Top1 can act as a kinase and phosphorylate splicing related-factors (61). So by inhibiting Top1, CPT can impair splicing. (62) However, it is not clear how Top1 can act as kinase, as it does not have a kinase domain. One possibility is that the impaired phosphorylation of SR proteins is an indirect consequence of Top1 inhibition by CPT.

The unbalancing of phosphorylation of PoII during CPT treatment (57) can alter POlII distribution along transcribe genes. Recently, our group demonstrated that Top1 inhibition by CPT increases escape of PolII from pausing sites, via CDKs, and induces antisense transcription at the HIF1a locus (63). If this phenomenon is exclusive of HIF1a or if Top1 regulates genome wide antisense transcription is not clear. A very recent paper brought to light new properties of CPT and topoisomerase inhibitors. In a screening for drugs in an Angelman disease mouse model, Top1 and Top2 inhibitors were found to be the only compounds able to unsilence the imprinted gene Ube3a, responsible for Angelman disease. These novel insights give new light to topoisomerase inhibitors as potential therapeuticals for the treatment of this neurological disorder. Interestingly, CPT and derivatives induce a decrease of the antisense present at the imprinted locus. (64)

26

All these data revealed that Top1 can be a key regulator of transcription and transcription-related process. So the investigation of this enzyme and the use of its specific inhibitor CPT can bring novel insights in the field of regulation of transcription

1.2.6 Top1 and R Loops

Another possible mechanism of DNA damage induced by camptothecin is the formation of R loops. An R loop is three stranded nucleic acid structure composed by a DNA/RNA hybrid and a single strand DNA. R loops are usually form co-transcriptionally when the nascent transcript is able to re-anneal to the DNA template and leave the non-template DNA in single strand form (65). These structures have been always considered "dangerous" since can represent a source of genome stability. Three features can favor R loop formation: 1) an asymmetry in guanine distribution such that the nascent RNA result G rich (GC skew) 2) the presence of a nick on the non-template DNA 3) the presence of negative supercoils. (66). Particularly, as regards the topology, the presence of negative supercoils can lead to the separation of the two strands thus favoring the RNA invasion. So Top1 and R loops are clearly interconnected. First evidence of a possible role of Top1 in R loops homeostasis came with Drolet's studies in *E. coli*: topA mutants complemented growth defects by overexpression of RNAseH, enzyme able to specifically degrade RNA/DNA hybrids (67), (68). In *S. cerevisiae* R loops form at rDNA loci in Top1Δ Top2Δ strains and are enhanced by co-depletion of RNAseH1 (69). In mammals it has been shown that Top1 depletion induces interference between replication and transcription and this effect is reversed by RNAseH1 overexpression (70). Obviously, also CPT can have a role in stabilizing R loops. Sordet et al. demonstrated that CPT is able to induced DSB and DDR response via ATM, in non-replicating cells such as neurons and lymphocytes. Interestingly, this DNA damage is partially reversed in the presence of an overexpression of RNAseH1 (71). R loops can also play a role in the reactivation of the paternal Ube3a allele by CPT, as it has been found that CPT stabilize an R loop in this imprinted locus, which maintain a

decondensed chromatin state, allowing transcription of the paternal Ube3a allele (72).

Top1 is not the only topoisomerase involved in preventing R loops. A really recent paper showed that TOP3b is recruited to chromatin to prevent R loop formation, by interacting with TDRD3 that recognizes di-methylated arginine in histone H4 (73). However, considering the marginal role of TOP3b in transcription, restricted to few loci, the main topological regulator that can affect R loop stability remains, in our knowledge, Top1. All the previously described studies assessed a role of R loops in Top1-driven genomic instability by indirect evidence (usually an overexpression/deletion of RNAseH). However, direct proofs that Top1 depletion or inhibition by CPT stabilize R loops are still missing. Moreover, a genome wide mapping of these structure in the presence/absence of Top1, or in the presence of CPT was not performed yet.

1.2.7 R loops and genomic instability

R loop formation can be considered a dangerous process. Leaving the non template DNA in a single strand form because of DNA/RNA hybridization, can be a source of DNA damage. Anyway how R loop can cause genome instability is not clear yet. The ssDNA can be a template for process such transcription-associated mutagenesis (TAM) and transcription-associated recombination (TAR). First, the single strand DNA can undergo to spontaneous deamination of dC to dU therefore inserting point mutations (74). Another possibility is that specific proteins can recognize the ssDNA or the entire R loops and triggers the mutagenesis process. One possible candidate to this is the activation-induced cytidine deaminase AID for example, involved in Ig heavy chain CSR and hypermutation in B-lymphocytes. This enzyme deaminates cytosines in uracyle on ss DNA (75). However AID expression seems to be restricted to lymphocites and cannot explain alone the mutagenic pattern induced by R loops in every kind of cell. Another potential source of genomic instability triggered by R loops can be the interference with replication (76). Collision between replication fork and a stalled PolII, can trigger TAR, although the exact mechanism is not clear.

A replication fork (RF) can collide with the R loop itself or alternatively R loop can induce stalling an RNA polymerase that will crush with the RF. Anyway what make a break (single or double strand) on the DNA molecule in the presence of an R loop still remains mainly unclear. Apart the above-mentioned Topoisomerase, what other factor can prevent R loop formation?

The first factors that have to be cited are obviously *RNAseH enzymes*. These enzyme are able to specifically cleave the RNA moiety in an RNA/DNA duplex. Most part of organisms have two classes of RNAse H, type 1 and type2, the first monomeric, the second multimeric. RNAseH1 is present both in nucleus and mitochondria. In the first compartment RNAseH1 seems to degrade R loops associated with transcription (77), while at mitochondria level it seems to play a role in replication of the mitochondrial DNA (78). Differently from RNAseH1, RNAseH2, can cleave also a single ribonucleotide misuncorporated in the DNA molecule. It is also responsible for removal of the Okazaki primer composing the RNA primer on the replicating lagging strand (79). Notably RHAseH2 is mutated in a severe immunological disorder called *Acardi-Goutieres Syndrome*.

Another important class of proteins able to remove R loops are the *DNA/RNA helicases:* in yeast Pif1, in mammals DXH9 and Senataxin, among them the latter is definitively better studied. Senataxin (Sen1 in yeast), in particular, is mutated in two severe neurological diseases: *ataxia oculomotor apraxia 2* (AOA2) *and amyotrophic lateral sclerosis type 4* (ALS4). Senataxin has been shown to protect replication fork along RNA Polymerase II transcribed genes by impeding formation of R loops (80)

Another important process that, if altered, can affect R loops formation is the *mRNA packaging and RNA export pathway*. In mammals these two processes are strictly linked and accomplished respectively by THO and TREX complex. The THO/TREX complex is responsible for packaging of pre-mRNA with RNA-binding proteins. These proteins, when mutated in yeast give a hyperrecombination phenotype mediated by R loops (81). The role of these proteins in R loop formation can be explained with their property to bind co-trascriptionally the nascent RNA bringing it toward the nuclear pore. In the absence of these key proteins, RNA is free and may invade the DNA duplex behind a transcribing RNA polymerase.

The last but not the least class of proteins that can prevent R loop formation are *the splicing factors.* From a screening of siRNA libraries aimed to identify factors involved in genome instability, factors involved in splicing resulted the most represented class (82). Among them ASF/SF2 was the first characterized: when depleted this factor lead to R loop formation and genomic instability (83). We can find two rational explanations for the importance that splicing factors have in R loop formation. First, in a manner similar to THO/TREX, co-transcriptional binding of the nascent RNA by splicesome sequesters RNA from invading the DNA duplex. Second, and maybe more importantly, by removing intronic sequence, splicing factors reduce the possibility of a nascent RNA to anneal perfectly to the DNA sequence.

All the considered factors have been shown to induce genomic instability when depleted. The involvement of R loops in this process was always shown trough overexpression of RNAseH that reverse the instability phenotype. However direct evidence of an increased R loops formation and localization of these altered structures in the genome are still missing. One potential mechanism that could explain why R loops increased by depletion of these factors lead to breakage on DNA is explained in a really recent work published by Cimprich's Lab. The group demonstrated in human cells that R-loops induced by the absence of different factors, including the RNA/DNA helicases Aquarius (AQR) and Senataxin (SETX) and the splicing factor ASF/SF2 or by the inhibition of topoisomerase I by CPT, are actively processed into DSBs by the nucleotide excision repair endonucleases XPF and XPG. The entire TC-NER pathway (and not the GG-NER) was shown to be involved in this process (84). This is the first study that shows the possibility of a unique mechanism underlying the R loop-driven instability.

1.2.8 R loops as regulators of gene expression

In spite of their role in genomic instability, recent studies have brought light to new potential roles that these structure could have in regulating transcription and gene expression. The first physiological role of R loop was uncovered in 2003. Yu et al demonstrated that R loop forms at physiological level at Immunoglobulin heavy

chain locus in activated B lymphocytes and these phenomenon triggers CSR switch class recombination. The length of these R loops can exceed over 1 kb (85). This paper suggested for the first time that the substrate to promote CSR by AID enzyme could be an R loop (74).

Another potential role on gene expression was given by Chedin's lab. They provided evidence for wide-spread R loop formation over 5' regions downstream of CpG Islands promoters. This R loop formation is driven by sequence and they form when the template strand is rich in C. Intriguingly R loop formation at CpG Island promoters seems to protect them against methylation (86) (87).

A third and important role identified for R loops comes from Proudfoot's lab and involves the transcriptional termination process. This group found that R loop forms at termination regions at levels of GC reach regions. This R loop slow down PolII downstream of the PolyA site. Then this structure has to be resolved by senataxin to promote Xrn2-mediated transcript degradation and efficient termination (88). Intriguingly R loops at terminator sites induce repressive chromatin marks via RITS complex, to promote efficient termination (19). Finally, a really recent paper demonstrated that senataxin is recruited to terminator site trough interaction with the tumour suppressor BRCA1. The latter repair the R loop-driven DNA damage at terminator sites (89). These data suggest that physiological and aberrant R loops are strictly connected, and deregulation of one of the involved factors could switch the physiological state into a "dangerous situation".

Chapter II

MATERIAL and METHODS

2.1 Cell lines

The cell lines HCT116, PC3, HEK293, NHDF,Ntera2 were purchased from ATCC (LGC Standards S.r.l., Milan, Italy) and were grown in Dulbecco's modified Eagle's medium (HCT116, HEK293,Ntera2, NHDF) or RPMI (PC3) mediums with 10% fetal bovine serum and glutamine 2mM. HCT116-shRNATop1 cell line was gently provided by Y. Pommier (NCI, Bethesda, MD, USA) and was grown as HCT116 cells but in the presence of 200 µg/ml Hygromycin B. The cell line stably expresses a short hairpin RNA targeting exon 17 in the TOP1 gene.Cells were maintained at 37°C in a humidified incubator containing 20% O_2 and 5% CO_2. Cell line identity was certified with Cell ID System (Promega) by BMR Genomics Srl (Padova, Italy).

2.2 Drugs and Antobodies

Triptolide, MG132, Camptothecin and Aphidicolin were purchased from Sigma Aldrich, dissolved in dimethyl sulfoxide (Sigma) and stored at -20° (except for MG132 stored at-80°) Drug aliquots were thawed immediately before each experiment. RPB1 (CTD repeat), phospho-Ser-2-RPB1, phospho-Ser-5-RPB1 antibodies were from Abcam, CDK7, B-actin, Top1 (c-15), RPB1 (H-224), H1 and gH2AX antibodies were from Santa Cruz Biotechnology. S9.6 antibody was purified from a mouse hybridoma (clone HB8730) from ascetic fluids as previously reported.

2.3 Cell treatments

For Triptolide experiments: Exponentially growing cells(75% of confluence) were

exposed to 10 μM or 100nM of Triptolide for the indicated time before protein extraction or formaldehyde fixation.

For Camptothecin experiments: Exponentially growing cells were exposed to 10 μM CPT. In case of co-treatment, cells were incubated with aphidicolin (5 μM) for 15 min, then CPT (10 μM) was added to the medium for the indicated time.

2.4 Knockdown experiments

CDK7 knockdown: PC-3 cells were plated at 18000 cells/cm2 density. Cells were transfected 24 hours after the plating with the RNAiMax transfection reagent (1:1000) and 20 nmol/L of scramble siRNA or CDK7 siRNA (Invitrogen). CDK7 silencing assessment and triptolide treatments were conducted 72 hours after transfection.

CDK9 and Top1 knockdown: HCT116 cells were transfected 24 h after plating (30% confluence) with RNAiMax Transfection reagent (Life Technologies) and 20 nM of scramble siRNA or CDK9 siRNA or Top1 siRNA (targeting exon 16 of the TOP1 gene; Life Technologies). Proteins silencing and drugs treatment were performed 72 h post-transfection.

Top1 double knockdown: HEK293 cells were plated at 16000 cells/cm2 density. A first round of transfection was performed 24 hours post seeding with 10 nM of scramble or Top1 specific siRNA (targeting exon 16 of the TOP1 gene; Life Technologies). 48 hours after transfection 1/3 of cells were transfected in suspension (reverse transfection) with 10 nm of the same siRNA. Top1 silencing testing and genomic DNA extraction were performed 72 h post-second round of transfection.

2.5 Nuclear protein extract preparation

Cells were washed twice with Ice-cold PBS scraped and resuspend in hypotonic buffer (Hepes 10 mM, NaCl 50 mM, EDTA 1 mM, DTT 1 mM, Aprotinin, Leuppetin,

Pepstatitin 10 mg/ml), PMSF 1mM, NP-40 0.2%) for 30'. Nuclear pellet was resuspended in hypertonic buffer (Hepes 10 mM, NaCl 420 mM, EDTA 1 mM, Glycerol 10%, Aprotinin, Leuppetin, Pepstatitin 10 mg/ml), PMSF 1mM) at 4 degrees in gentle rotation for 30'. Cellular debris and DNA were pelleted by centrifugation at 4 degrees 18000 g. Supernatants were recovered and stored at -80 °C.

2.6 Histones extraction

Cells were washed twice with ice-cold PBS scraped and lysed with Lysis buffer [10 mM Tris-HCl (pH 7.5), 1 mM $MgCl_2$, 0.5% NP-40] containing protease inhibitors [aprotinin, leupeptin, pepstatin (10 mg/ml) and Phosphatase Inhibitor Cocktail II (Sigma)]. Cellular debris were pelleted by centrifugation at 16,000 g for 1 min at 4°C, resuspended and incubated for 15 min with Lysis buffer with 400 mM NaCl. Samples were briefly centrifugated, and pellets were incubated for 10 min at 4°C with 5 volumes of Extraction solution [220 mM H_2SO_4, 20% Glycerol, 10 mg/mL 2-mercaptoethanolammine]. The histone-containing supernatant was obtained bycentrifugation at 16,000 g for 10 min. Histones were pelleted from the supernatant by adding 20% trichloroacetic acid, and centrifugation. Then the pellets were resuspended in 100% ethanol and centrifuged again at 16,000 g for 20 min. The pelleted histones were store at -80 °C.

2.7 Western Blot analysis

Nuclear lysates, corresponding to 20 μg of proteins were separated by 6.5% SDS PAGE. Proteins were then blotted onto a Hybond ECL-nitrocellulose membrane. Histone aliquots, corresponding to 10-15 μg were loaded onto a 12% SDS-PAGE gel, and then transferred to a Hybond ECL-nitrocellulose membrane. Equal loading was checked with anti-beta actin for nuclear extracts and anti-histone H1 antibodies for histones lysate. Specific bands were then detected with ECL Plus Western blot imaging system (GE Healthcare). Horseradish peroxidase-

conjugated mouse and rabbit IgG (1:2,000 and 1:5,000 dilution respectively) were purchased from GE Healthcare. Horseradish peroxidase-conjugated goat (1:40,000 diluition) was from Santa Cruz Biotecnology.

2.8 RNA extraction and RT PCR

After drug treatments, $5x10^7$ cells were washed twice with cold PBS and collected through centrifugation. The pellet was frozen at -80 degrees for at least 1 hour and then resuspended and in 3.6 ml AE buffer [50 mM NaOAc (pH 5.2), 10 mM EDTA], 240 ml of SDS 25% and 3.6 ml of acid phenol (pH 4.5). Samples were then incubated for 10 min at 65°C mixing vigorously every minute. After 5' of incubation on ice, samples were centrifuged for 15 min at 12,000 g. The upper phase was collected; 3.9 ml of chloroform/isoamylic alcohol was added to it, then mixed and centrifuged for 10 min at 1,800 g. The upper phase was precipitated with isopropanol and NaOAc. The pellet was resuspended in TE, and DNA was digested with DNAse I (Thermo Scientific). RNAs were subsequently purified with phenol and precipitated with ethanol and NaOAc. RNA integrity was routinely checked by running 1% agarose denaturating gel electrophoresis. Then, 1 mg of total RNA was used to prepare cDNA using SuperScript III (Invitrogen) with reaction buffers suggested by the manufacturer, for 5 min at 65°C, 5 min at 25°C, and 60 min at 50°C, followed by alkaline hydrolysis with NaOH. cDNA was precipitated with EtOH precipitation.

2.9 Quantitative Real-Time PCR.

Real-time PCR were performed using LightCycler and FastStart DNA Master SYBR Green I kit (Roche Diagnostics, Mannheim) or Biorad CFX96 Touch instrument and SsoI Universal Master Mix (Biorad). Quantification and melting curve analyses were performed using Roche LightCycler software or Biorad CFX Manager Software as indicated by the supplier. PCR reactions contained 1x Master Mix and 400 nM of

each primer. Specificity of PCR products was routinely controlled by melting curve analysis and agarose gel electrophoresis.

2.10 Chromatin Immunoprecipitation

For chromatin extraction, $1x10^7$cells (PC3 or HEK293) were fixed with 1% formaldehyde for 15 min. The reaction was stopped with 125 mM glycine and cells were washed twice with ice-cold PBS. For triptolide experiments, PC3 cells were then washed with 7 ml of TEET [10 mM Tris-HCl (pH 8.0) 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100], 5 ml of TEEN [10 mM Tris-HCl (pH 8.0) 10 mM EDTA, 0.5 mM EGTA, 200 mM NaCl], and resuspended in 0.5 ml of TEE [10 mM Tris-HCl (pH 8.0) 10 mM EDTA, 0.5 mM EGTA]. For Top1 knockdown experiments, HEK293 were directly lysed in TEE-SDS 1%. Protease inhibitors [aprotinin, leupeptin and pepstatin (Sigma) 10 mg/ml] were added to the buffers immediately before use. Chromatin was then fragmented by sonication using a Bioruptor (Diagenode) to an average DNA fragment size of 300–400 bp. Immunoprecipitations were performed at 4°C in RIPA buffer [50 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.5 mM EGTA, 150 mM NaCl, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS]. Amounts of chromatin, equivalent to 0.4 U.A at 260 nm were taken for each immunoprecipitation. Samples were precleared for 2 hour with 4 µg of non-immune rabbit IgG and 20 µl of 50% suspension of a 1:1 mix of Protein A- and Protein G-Sepharose beads. Then, beads were descarted and chromatin was recovered by centrifugation for 3 min at 1,000g. 20% of the supernatants were saved as input. Supernatants were incubated overnight with 4 µg of specific antibody or nonimmune rabbit IgG (to measure aspecific recovery). ChIP-grade ab for Rbp1 (the largest subunit of RNA Polimerase II against the N term (H224)) was from Santa Cruz Biotechnology (Santa Cruz, CA). Non-immune rabbit IgG were from Jackson (West Grove, PA, USA). Immunocomplexes were recovered by addition of 40 ml of Protein A-/Protein G-Sepharose beads blocked with DNase-free BSA (9.95 mg/ml) and salmon testes DNA (10.5 mg/ml). Then, the beads were washed four times with RIPA buffer; once with

RIPA buffer containing 0.5 M NaCl; once with Li250 buffer [10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.25 M LiCl, 0.5% Na-deoxycholate, and 0.5% NP40]; twice with TE [10 mM Tris-HCl, 1 mM EDTA 47 pH 8.0)]; and finally resuspended in TE. Each wash was performed for 10 min by rocking at 20 rpm followed by 3 min of centrifugation at 1,000g. The pellets were then adjusted to 0.5% SDS and incubated overnight at 65°C to reverse cross-links. Samples were then digested with proteinase K (500 mg/ml Sigma) for 4 h at 52°C and extracted twice with phenol chloroform. DNAs was precipitated with ethanol in the presence of 20 μg of glycogen (Roche Diagnostics, Manheim) and dissolved in TE. Recovered DNA was quantified by real-time PCR (see Table 1 for the list of primers). At least three dilutions of input DNA were run to generate the standard curve. DNA recovery was measured as input DNA fraction.

2.11 Immunofluorescence with s9.6 antibody.

Cells (HCT116, HEK293, Ntera2 D1, NHDF) were seeded at 16000cells/cm2 on coverslip. 24 hours after post-seeding cells were treated for the indicated time and concentration with CPT. Cells were then fixed with ICE-cold methanol at RT for 10'. Then washed one time with ice cold PBS and permeabilized with acetone on ice for 1'. Slides were blocked with blocking buffer (BSA 3%, Tween 0,1%, SSC 4x) for 30' and incubated with s9.6 (1:50 in blocking buffer) for 2 hours at RT. After primary antibody incubation slides were washes three times for 5' in gentle agitation and then incubated with Alexa-fluor 498 anti-mouse fluorescein conjugated antibody (1:1000 in blocking buffer). Then slide were washed three times in SSC4X, once in PBS and then mounted using mounting solution containing DAPI 2 μg/ml for nuclear staining. Images were acquired with a fluorescence microscope Zeiis.

2.12 Genomic extraction for R loops detection.

Cells were washed twice with ice-cold PBS and directly lysed on the plate with TE-SDS 1%. DNA was collected in a 15ml falcon and incubated overnight with Proteinase K (Roche) at 70 µg/ml. After overnight incubation DNA was precipitated with ethanol 70% and sodium acetate 0,3M pH 5.0. DNA was recovered with a cut tip and washed 5 times in ethanol 70%. Once dried, DNA pellet was resuspended in TE and digested with restriction enzymes cocktail (2 units each: HindIII, EcoRI, XbaI, BsrgI, SspI) overnight at 37 degrees. After digestion fragment size was checked on agarose gel and DNA was stored at -80 degrees.

2.13 Dot blot for detection of global levels of RNA:DNA hybrids

6,5 µg of genomic DNA was dissolved in 800 µl of ddH20. Two-fold dilution were prepared in ddH20. 1 µg of diluted DNA was kept apart, precipitated and loaded on agarose gel as Input control. DNA dilutions were spotted on nitrocellulose membrane equilibrated in SSC 2X with a vacuum dot blot apparatus.. Wells were washed twice with SSC 2X and DNA was crosslinked to the membrane with an UV crosslinker at 12,0000 µJ/cm^2 for 10-15 s. Membrane was blocked for 30' with blocking buffer (PBS 1x, BSA 3%, tween 0,1%) and then incubated for two hours with s9.6 antibody (1:1000 in blocking buffer) at RT in gentle shaking. Three washings were performed with PBS-tween 0.1% and then membrane was incubated with Alexa fluor anti-mouse secondary antibody (1:10000 in blocking buffer). After three washing in PBS-tween 0.1% membrane was scanned din Licor scanner.

2.14 DRIP: DNA-RNA Immunprecipitation

4,445 µg of genomic DNA were diluted in 500 µl of TE (Tris 10mM, EDTA 1mM). 445ng of the diluted DNA was used as INPUT while 4 µg were incubated with 10 µg of s9.6 ab in binding buffer (10 mM NaPO4 pH 7.00,14 M NaCl0.05% Triton X-100) overnight at 4 degrees in gentle rotation. 16 hours later 50 µl of Protein A/G

sepharose beads (Pierce) were equilibrated in binding buffer three times and used to capture immunocomplex trough gentle rotation at 4 degrees for two hours. Beads were washed three times with binding buffer for 10 minutes in gentle rotation at RT. Eluition were performed trough incubation with proteinase k (500 μg/ml) for 45' at 55 degrees. Reactions were cleaned up with phenol/chloroform extraction and DNA was precipated with ethanol, sodium acetate and glycogen. R loop recovery and IP efficiency were assessed by Real time PCR.

2.15 DRIPc: DNA-RNA Immunoprecipitation coupled to RT.

DRIPc analysis was performed on immunoprecipitated DNA derived from at least 8 DRIPs. DNA was treated with 4 units of DNAse I (NEB) for 40' at 37 degrees. Reaction was stopped by adding EDTA 50 mM and incubating at 75 degree for 10'. RNA isolated was then precipitated with EtOH, NaOAC ph5.2 and glycongen. RNA was then used for a retrotranscription reaction. Iscript RT (biorad) protocol was used for first strand synthesis according to manufacturer directions. Zymo column were used to clean up the RT reactions. Second strand synthesis was performed using 10 units of DNA polymerase and 1.6 units of RNAseH (NEB) and E. coli DNA ligase. Replacing dTTP with dUTP performs reaction allowed labeling of second strand with uridine. Raection was cleaned up again with Zymo columns.

2.16 Library Preparation for ChIPseq, DRIPseq and DRIPc seq

DNA from ChIP was directly used for library preparation as already sheared at the desired size level (300-500 bp). DNA from DRIP and DRIPc was instead sonicated using Bioruptor from Diagenode using the following protocol: HIGH intensity 10 cycles 15 s ON/ 90 sec OFF. End repair was performed using NEB repair module cocktail in the presence of ATP at RT for 30'. Reaction was cleaned up with Qiagen PCR clean up columns. A-tailing was realized by using Knleow fragment without exonuclease activity in the presence of dATP. And followed by purification with

MinELute Quiagen columns. Illumina Truseq RNA adapters were added by quick ligation (NEB) and size selection was performed to exclude adapters ligation products. qPCR with primers on adapters was performed to understand number of cycles of PCR for a good library amplification. At this point for strand specific DRIPc DNA libraries were treated with UNG glycosidase to degrade U labeled strand. On the basis of qPCR results library were amplified using Primers for Trueseq adapters and Phusion 2X HF master mix with the following protocol; 10 sec at 98 degrees, 30 sec at 60 degrees and 30 sec at 72 degrees (for 13-16 cycles) and 5' at 72 degrees. Size exclusion using Ampure removes fragments less long than 200 bp and longer than 500 bp. Library quality and quantitation was checked with Byoanalyzer (Agilent). Deep sequencing was performed by Computational Genomics Resource Laboratory, University of Berkley.

2.17 RNAseq libraries preparation.

Total RNA was extracted after 72 hous post-second round of transfection using Direct-zol RNA Mini-prep, quantified and quality checked on denaturating agarose gel. Total RNA libraries preparation and Deep sequencing were peformed by Computational Genomics Resource Laboratory, University of Berkley.

2.18 DRIP-seq, DRIPc-seq and PolII ChIP-seq bioinformatic analysis

Sequencing reads were trimmed using FastqMcf and mapped to the hg19 reference human genome using BWA version 0.6.1-r104 (90). For DRIPc, mapped reads were assigned to plus or minus strand using SAMtools (91). To normalize between different samples, equal number of mapped reads for each sample was used for all downstream analysis. Peak calling was performed using a custom 7-state Hidden Markov Model coupled to a Genetic Algorithm (unpublished software). TSS, TTS and gene metaplots were generated using custom Perl and R scripts. All overlap analysis was performed using BEDTools (92).

2.19 RNA-seq bioinformatics analysis

RNA-seq reads were trimmed as above, mapped to hg19 using TopHat2 (93) and normalized as above. Read counts for each gene were calculated using HTSeq and differential gene expression was identified using DESeq (94) using fold change > 2 and FDR < 0.05.

Chapter III

RESULTS

3.1 Effects of Triptolide on Rbp1 stability

3.1.1 Triptolide induces Ser5 hyperphosphorylation prior to RNA Polymerase II decrease.

To investigate how triptolide (TPL) can affect RNA Polymerase II stability we decided to assess levels of total RNA polymerase and levels of phosphorylation of PolII-CTD during triptolide treatment. Figure 1 shows that degradation of Rpb1, the largest subunit of RNA Polymerase II, induced by triptolide is dose and time dependent. Particularly, focusing on time 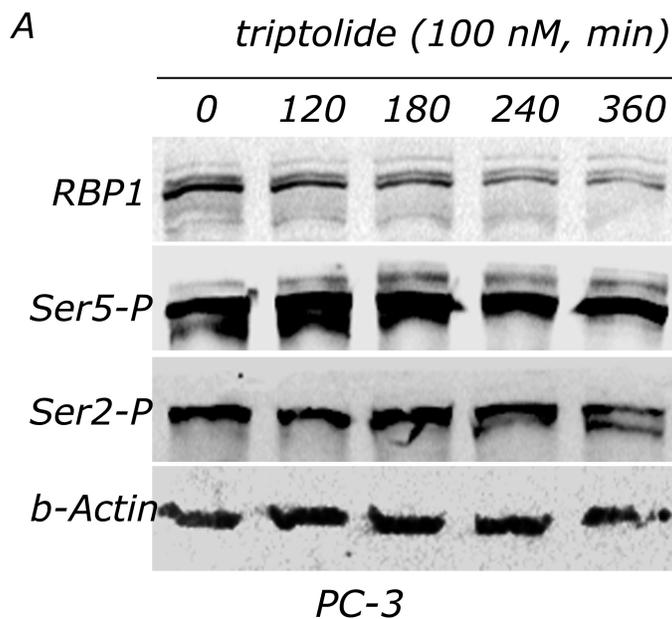course, at early time of treatment (2h) the PolII reduction is minor as compared to control. Rpb1 levels start decreasing at later stages of treatment with TPL 100 nM (3-4 h). Interestingly, western blot analysis using antibodies specific for phosphorylated forms (Ser 2 and Ser5) of the carboxyl terminal domain of PolII reveal that at early times of treatment PolII reduction is preceded by a hyperphoshorylation event of

Fig 1.*Triptolide induces hyperphosphorylation of Ser5 on PolII-CTD preceding RPB1 degradation* Western blot analysis to detect Rpb1 levels and phosphorylation state of CTD of POlII .during triptolide treatment. B actin is used as loading control

Ser5 on CTD. Notably Ser2 levels do not change until 6h of treatment, when lower bands for this residue can be detected, probably due to hypophosphorylating events. This data suggest the possibility that the degradation of RNA Polymerase II induced by triptolide can be an event triggered by an hyperphosphorylation of CTD, particurarly at Ser5 sites.

3.1.2 Triptolide induces block of RNA Polymerase II at promoters of active genes.

Western blot analysis demonstrated that total levels of RNA Polymerase II
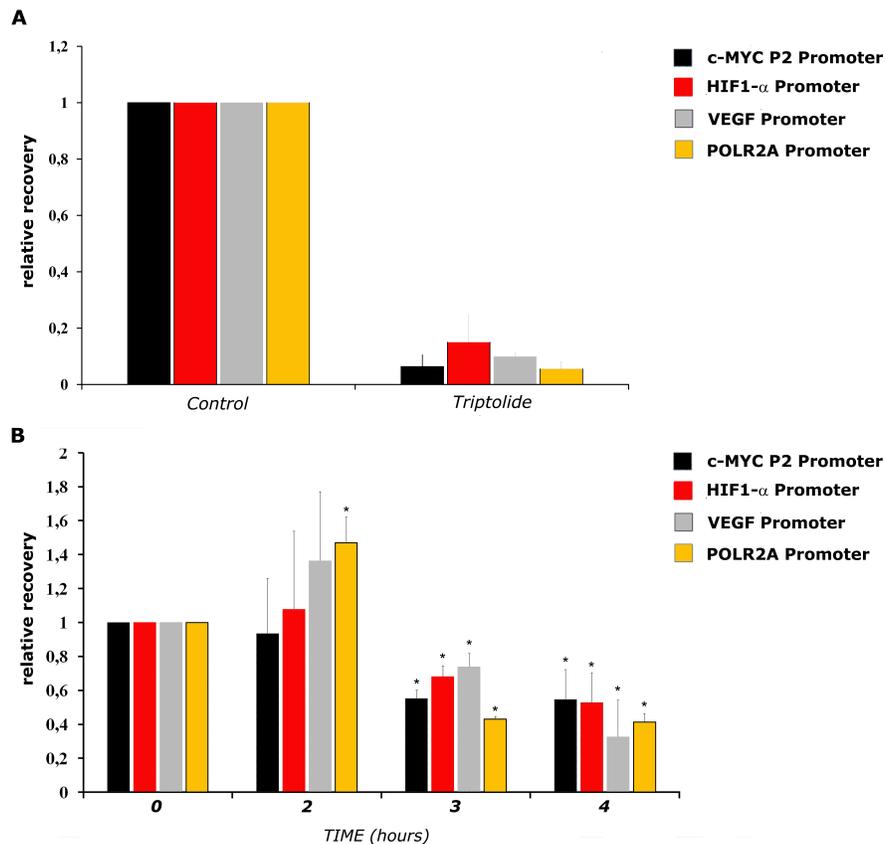

Fig 2. *Triptolide induces a decrease of chromatine bound PolII at promoters of active genes.* ChIP experiments performed using a specific ab against N term of RPB1 with chromatin extracted from PC3 cells control and treated with TPL 1 μM 2 h (A) and 100 nM up to 4 hours (B). Recovery are normalized on control cells. Significance was calculated with t student test with * = P<0,05.

decreased during triptolide treatment in a time dependent manner, and this reduction follows a hyperphosphorylation event at level of Ser5 of CTD. To understand if this reduction involves also chromatin bound fraction of PolII we performed ChIP experiment using an antibody against N terminal of Rpb1 with chromatin extracted from PC3 cells treated with different doses and times TPL of treatment. We first found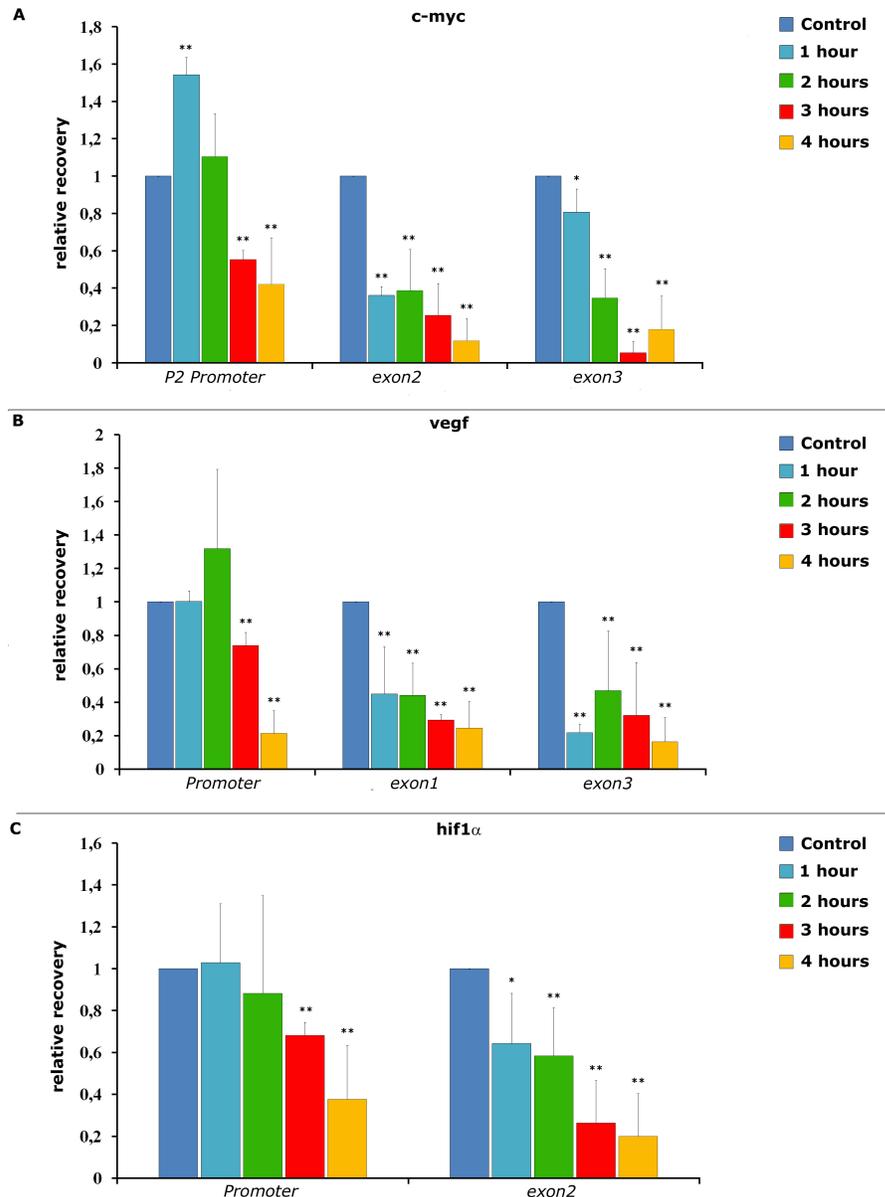 that a treatment of 2 h with 1 µM of TPL is able to deplete more than 90% of chromatin bound PolII at all the tested promoters of active genes (*c-Myc, HIF1a, Polr2a,VEGF*) (Fig 2a).

Time course experiments with lower dose (100 nM) showed reduction at the same loci in a time dependent manner (Fig 2b). Intriguingly short times of treatments (2h) that correspond to hyperphosphorylation at Ser5

Fig 3.*Triptolide induces a block of RNA Polymerase II at promoters of transcribed genes.* ChIP experiments performed using a specific ab against N term of RPB1 with chromatine extracted from PC3 cells control and treated 100 nM up to 4 hours (B). Recovery are normalized on control cells. .a)*c-myc* gene B) *vegf* gene C) *hif1a* gene Significance was calculated with t student test with * = P<0,05 and **= P< 0,01

sites did not show reduction of PolII at promoters and in one case, even an increase of polII occupancy could be detected (*polr2a*). Since a PolII reduction at promoters could be due also to an increased escape of pol II from these regions, as already shown for Top1 inhibitors (63), we decided to check PolII occupancy at exons levels. We found that TPL treatment induces PolII reduction both on exons and promoters levels, but at short times of treatment (1 h 100 nM) PolII occupancy did not decrease at promoters, indeed even an increase for c-myc promoter could be detected. Instead, at the corresponding gene exons a decrease of PolII presence was clearly detected at short times (Fig 3). These data showed that short times of treatment and low doses of TPL induce a block of RNA polymerase II at promoter of active genes.

3.1.3 CDK7 mediates Triptolide induced reduction of RNA Polymerase II

The hyperphosphorylation event seen at short times of TPL treatments could be the key molecular event that induces PolII block at promoters and persistent degradation of Rpb1. The major kinase responsible for Ser 5 phosphorylation is CDK7, a cyclin dependent kinase that is part of the transcriptional factor TFIIH. Intriguingly, the pharmacological target of TPL is XPB, which is part of TFIIH complex. To understand if CDK7 was responsible for the TPL-induced Ser5 hyperphosphorylation and if its kinase activity is important for Rbp1 stability during TPL treatments we performed siRNA knockdown of Cdk7 followed by TPL treatments. We found that downregulation of CDK7 affects hyperphosphorylation of Ser5 of CTD and that the following Rpb1 degradation was partially rescued after the kinase silencing (Fig 4a). Notably, PolII degradation was reversed even with high dose treatment of TPL (1 μM 2 h) (Fig 4b). ChIP experiments showed that CDK7 knockdown protects chromatin bound PolII by TPL induced depletion (Fig 4c).
These data demonstrate that CDK7 is involved in Rpb1 stability during TPL treatments.
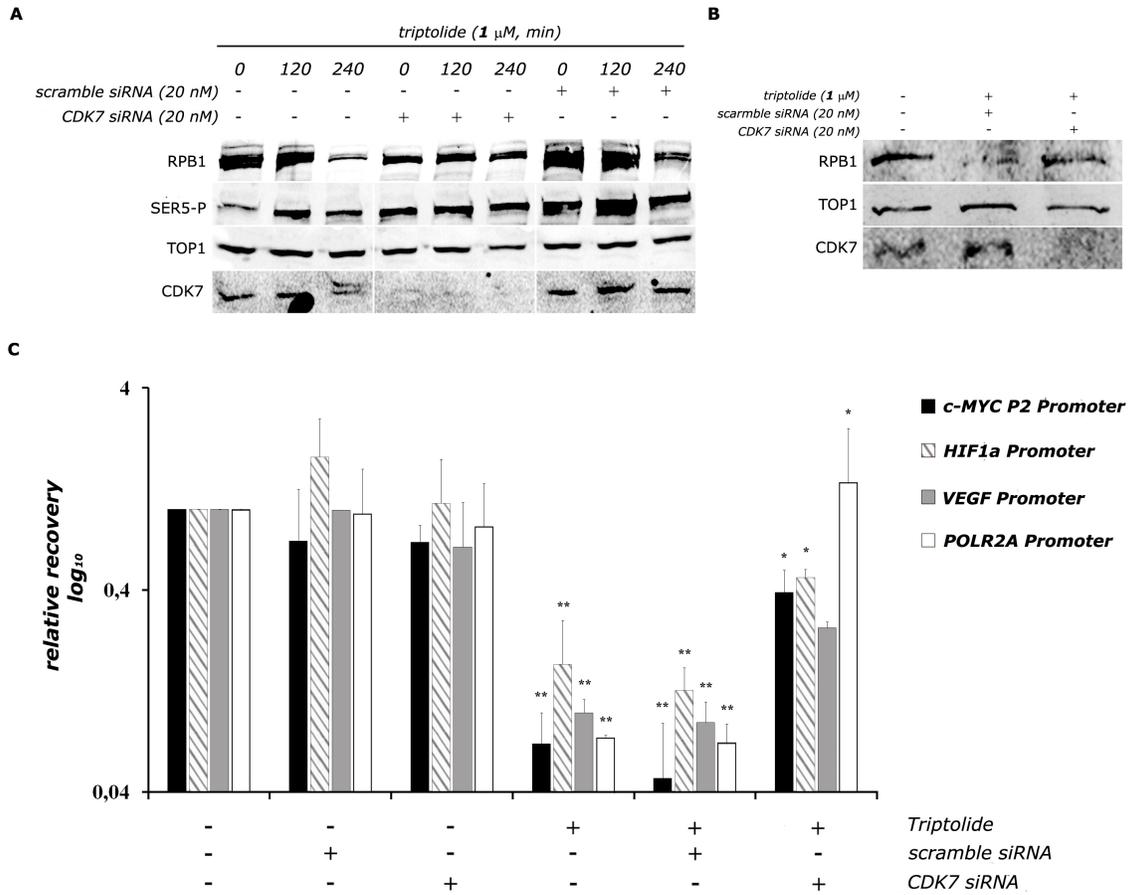
A

*triptolide (**1** μM, min)*

	0	120	240	0	120	240	0	120	240
scramble siRNA (20 nM)	-	-	-	-	-	-	+	+	+
CDK7 siRNA (20 nM)	-	-	-	+	+	+	-	-	-

RPB1

SER5-P

TOP1

CDK7

B

	*triptolide (**1** μM)*	-	+	+
	scarmble siRNA (20 nM)	-	-	-
	CDK7 siRNA (20 nM)	-	-	+

RPB1

TOP1

CDK7

C

■ *c-MYC P2 Promoter*

▨ *HIF1a Promoter*

▧ *VEGF Promoter*

□ *POLR2A Promoter*

relative recovery log₁₀

Triptolide	-	-	-	+	+	+
scramble siRNA	-	+	-	-	+	-
CDK7 siRNA	-	-	+	-	-	+

Fig4. *CDK7 mediates TPL-induced Ser5 hyperphosphorilation and Rbp1 degrdation.* A) Western blot analysis on nuclear extracts from cells treated with TPL 100nm for 2 and 4 hours in the presence/absence of downregulation for CDK7. B) Western blot from cells treated at high dose (1 μM) of TPL for 2hours. C) ChIP experiments for POlII on chromatin extracted from cells treated with TPL 2h 1 μM in the presence/absence of CDK7 downregulation.

3.2 Transcriptional role of Top1: camptothecin-induced Antisense transcription and Top1-mediated R loops stabilization

3.2.1 Camptothecin enhances antisense transcription at divergent promoters.

To better understand if the previously published findings about CPT-induced antisense activation seen at the HIF1a locus was a single gene phenomenon or a more general mechanism, we performed a genome-wide study of CPT-induced variations in transcript levels by NGS techniques. RNAseq data were colected using RNA extracted from
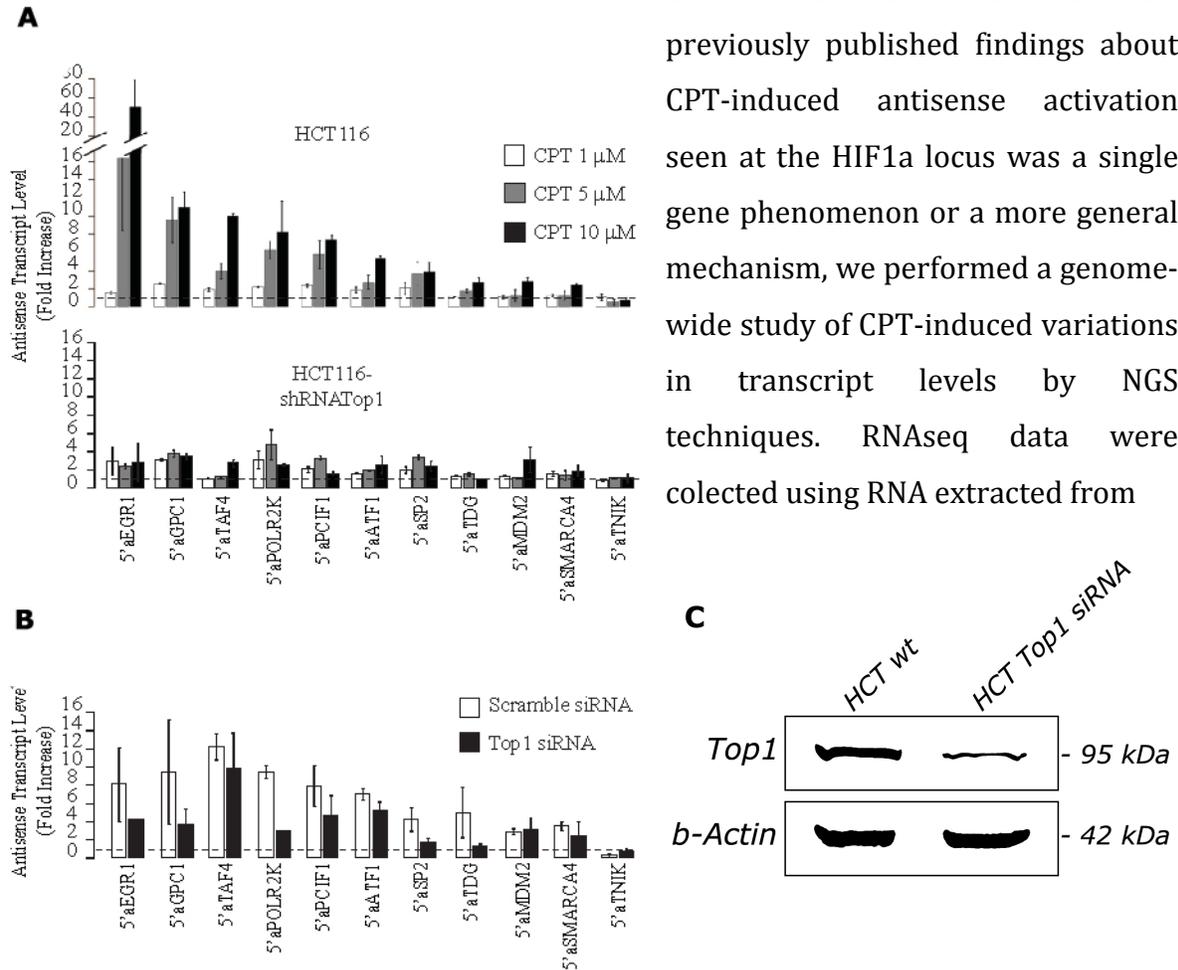
Fig 5. *CPT induces antisense transcription in a Top1 dependent manner.* A) RT-PCR measuring antisense transcript levels in HCT116 and HCT116 Top1 siRNA (stable clone) treated with different doses of CPT. B) RT-PCR measuring antisense transcript levels in HCT116 and HCT116 transiently downregulated for Top1 treated for 4 h with different doses of CPT. C) Western blot analysis to asses Top1 levele in HCT116 and HCT116 Top1shRNA

HCT116 cells control and treated for 4 hours with CPT 10 μM. Before sequencing, RNA was treated with bisulphite to maintain the strand specificity. With this approach we were able to identify 256 antisense transcripts located at CpG Islands

47

promoters. Notably these promoters were almost exclusively bidirectional promoters suggesting that CPT and Top1cc can interfere with directionality of promoter. We decided first to validate and characterize such transcripts. We randomly selected 10 loci to be validated by RT qPCR. Figure shows the fold increase of these transcripts after CPT treatment, demonstrating a good agreement with RNA seq data (FIG 5 A).

Strand information obtained with primer specific RT revealed that these transcripts were effectively antisense to the gene mRNA (data not shown). The antisense activation was definitively reduced in HCT116 cells stably downregulated for Top1 (HCT116 shRNA Top1) (Fig5 A) suggesting that Top1 is necessary for the CPT effect on antisense transcripts at bidirectional CpG-island promoters. Same results were obtained when we performed a transient Top1 knockdown: Fig 5B show that transient knockdown induce a decrease on antisense activation at almost all the tested loci. Notably the simple downregulation of Top1 does not induce antisense
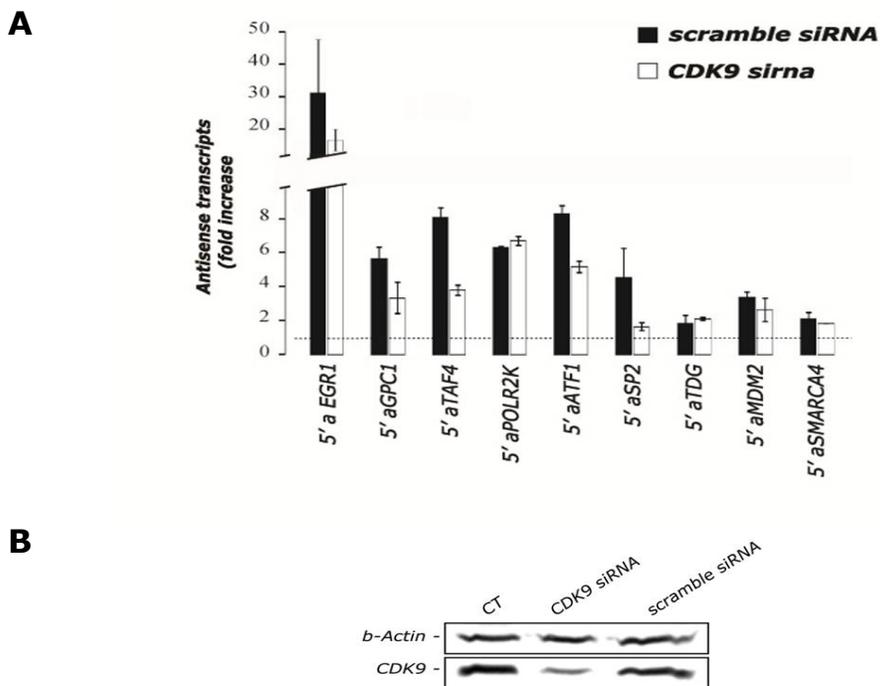

Fig 6. *CDK9 mediates CPT-induced antisense activation.* **A)** RT-qPCR meausuring antisense transcript levels in the presence of downregulation of CDK9 **B)**Western blot analysis to assess CDK9 knockdown

48

enhancement suggesting that Top1cc formation is essential for this mechanism.

Since it has been shown that divergent transcription at bidirectional promoters can be regulated by the transcription elongation factor PTEFb we decided to studyCPT effects on antisense transcription following downregulation of CDK9, the kinase that is part of PTEFb and positively regulates elongation. Fig (6A) shows that transient CDK9 knockdown partially suppressed antisense activation at almost all the tested loci. Suppression was not total, but significant, especially considering the partial knockdown of CDK9 Fig (Fig 6B) .

Top1 activity is crucial for correct transcription, but this enzyme plays important role in replication too. To better understand if replication could affect in some way CPT-induced enhancement of antisense transcripts, we measured by RT qPCR CPT induced antisense transcripts levels in the presence of a DNA polymerase inhibitor aphidicolin, that usually blocks cells in early S phase. Aphidicolin reverses the replication-dependent DNA damage induced by CPT. Fig 7b shows that DNA damage marker γ-H2AX is partially rescued by when cells are co-incubated with CPT and aphidicolin. Fig 7a shows that CPT was able to induce antisense activation in the presence of this replication inhibitor at almost the same extent of control cells. The antisense increase could be detected even with a lower dose of CPT, 1 μM, and aphidicolin was still not able to suppress the phenomenon . These data suggest that Top1 inhibition by CPT perturbs transcription at bidirectional promoters in a replication independent and Top1/CDK9 dependent manner.
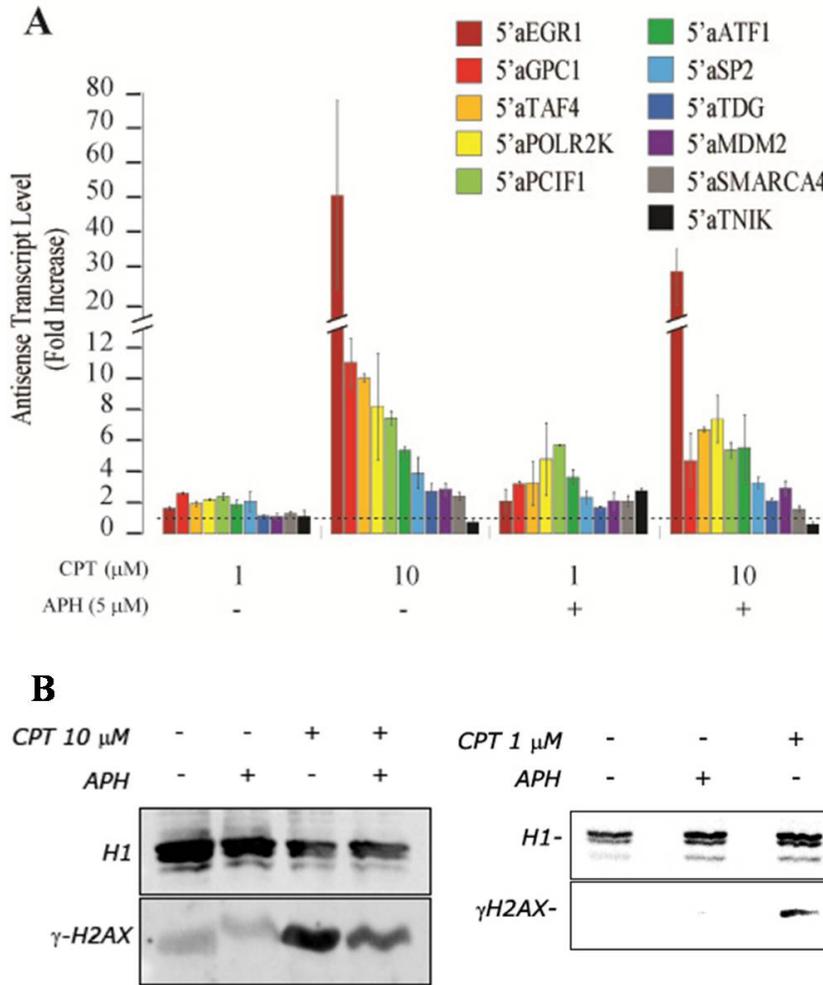
Fig 7. *Antisense activation induced by CPT is a replication independent process.* A)RT-qPCR measuring antisensen transcript levels in cells treated with two different concentration of CPT (1 and 10 μM, 4 hours) in the presence/absence of aphidicolin. B) Western blot analysis with ab against γ-H2AX to detect CPT-induced DNA damage. H1 is used as loading control

3.2.2 Top1cc induced by CPT parallels a transient increase of R loop structures

CPT is highly diffusible molecule. For this reason, it is able to penetrate cells really fast and entrap Top1 on DNA during short times of treatment. Previously published evidence from our lab showed indeed that Top1ccs are formed within 2' of treatment with CPT 10 μM at nuclear chromatin (ref). As CPT-induced Top1cc can increase negative supercoiling at active promoters (REF), and negative supercoils

can favor the formation of non-B DNA structures as R loops, we decided to investigate if R loops structures could be in some ways affected by CPT at promoters
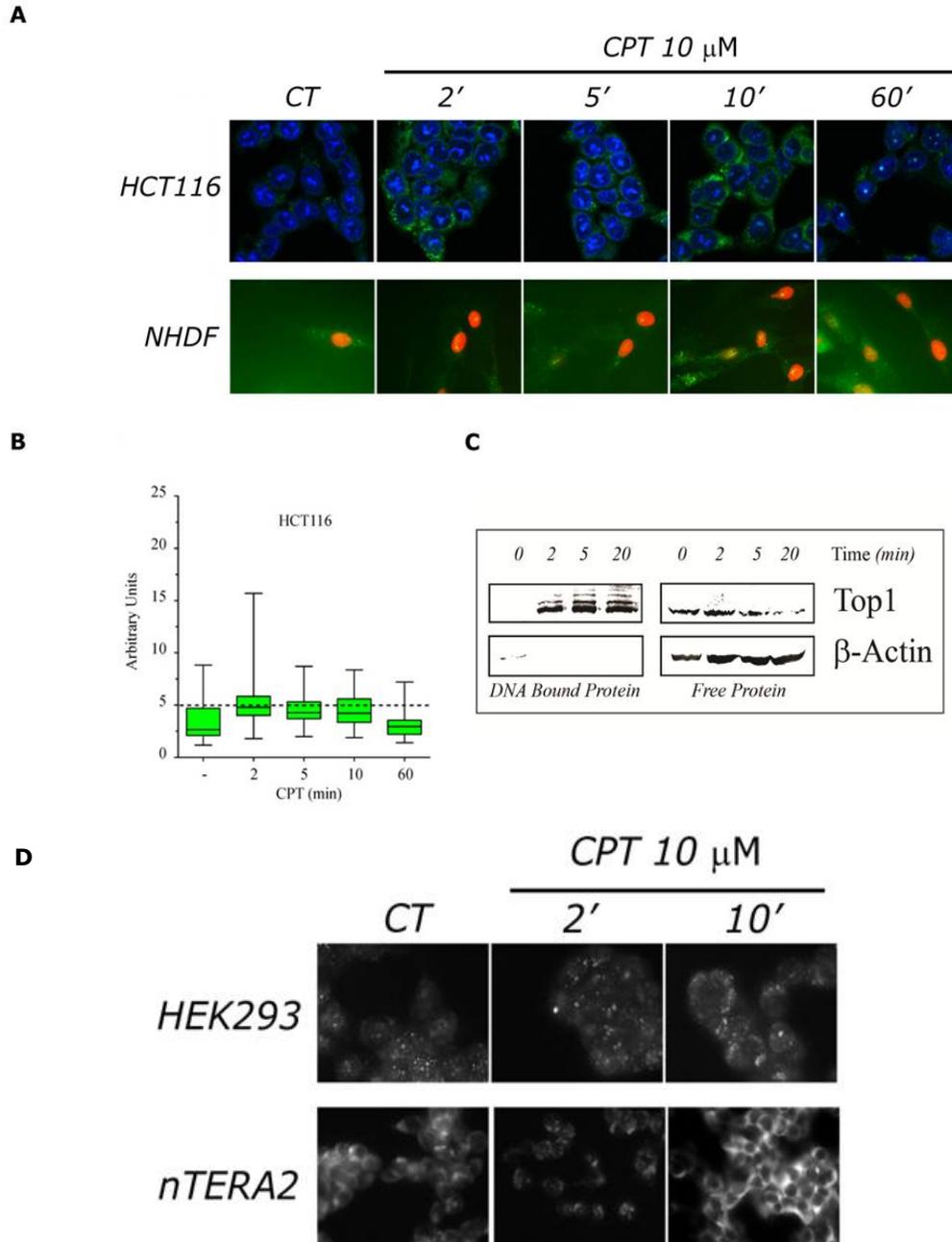

Fig 8. ***Top1 inhibition induced a burst of R loop structure at cellular levels at shorts time of CPT treatments*** A) Immunofluorescence using s9.6 ab on methanol fixed cells (HCT116 and Fibroblast) treated for different times with CPT 10 μM. DAPI was used for nuclear staining. B) Fluorescence intensity for HCT116 in A quantified with Image J. C) Immunofluorescence using s9.6 ab on methanol fixed cells (HEK 293 and Ntera2) treated for short times with CPT 10 μM. D) ICE bioassay showing entrapment of Top1 on DNA during short times of CPT treatments (10 μM).

showing an increase of antisense transcript levels.

First, we performed immunofluorescence investigations using a specific antibody for RNA/DNA hybrids (s9.6 ab). The results are showed in Figure 8 that partially represents our work recently published on *Nucleic Acid Research* (95)*.* We found that short times of treatments (2-5') with 10 μM of CPT induce a burst of R loops structures at cellular level, particularly at highly transcribed regions as nucleoli and mitochondria (Fig 8A).. The R loop increase seems to parallel the Top1cc formation and the post-transductional modifications of Top1, as detected by ICE bioassay, a technique that allow isolating protein covalently bound to DNA. Top1 results entrapped on DNA within few minutes and this phenomenon partially overlaps with burst of R loop structures (Fig 8D). Interestingly, R loops   show a peak in fluorescence at 2'-5' of treatment and then a decrease at longer time (1h)(Fig 8B), while Top1cc continues to increase up 20' even though the experiments were not quantitative (Fig 8c).

Similar dynamics were found for different cell types: HCT116, HEK293, Ntera2, NHDF (Fig 8d). In all cases the increase was detected clearly at nucleoli and mitochondria, in a manner that seems to be general and pervasive. At nuclear level CPT seems to induce the formation of hot R loops foci detected in IF as strong brilliant dots. Altogether, the data suggest that Top1 poisoning increases cellular R loops, but then these structures are rapidly removed.

3.2.3 R loops in genomic DNA from cells treated with CPT are highly unstable and difficult to isolate

Immunofluorescence results showed clearly that poisoning of Top1 by CPT induced a transient increase of R loop structures at cellular level. Thus, we wondered whether such an increase could also be detected at specific gene loci and at bidirectional CpG-island promoters showing an increase of antisense transcripts. To this end, we decided to directly measured R loop formation at selected genomic regions by DRIP technique. In collaboration with Frederic Chedin's Lab we first

performed DRIP (DNA/RNA immunoprecipitation) qPCR on genomic DNA extracted from Ntera2 cells treated for different times with 10 μM CPT.
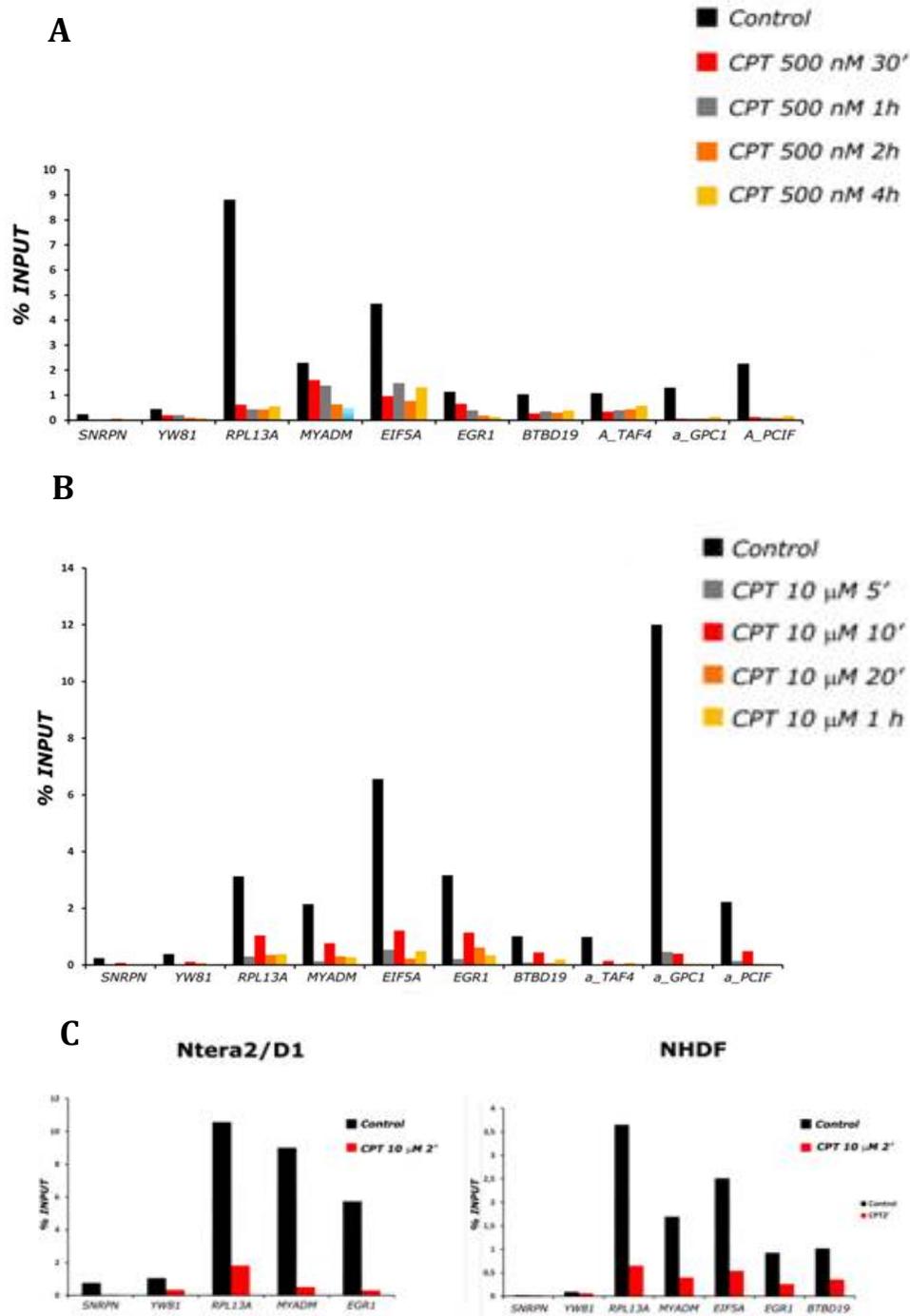

Fig 9. *R loops recovery after CPT treatments is impaired.* DRIP qPCR analysis measuring levels of R loops in control and CPT treated cells.A) Time course with 500 nM CPT up to 4h in Ntera2 cells. B) Time course with 10 μM CPT up to 1 h in Ntera2 cells. C) Effects of 2' of treatments with 10 μM CPT on NTera2 and NHDF cell lines. *SNRPN and YW81* are negative loci for R loop formation

Strikingly we found a great decrease of DNA/RNA hybrids at all the tested loci positive for R loops formation and for all the tested times of treatment (Fig 9).

This data, in clear contrast with IF results, suggested to us the charming hypothesis that Top1 inhibition could lead to a distribution of R loops with a disappearing of classical, well characterized R loops and an arise of new R loops structure. Before performing DRIP analysis coupled to NGS we wanted to be sure that the increase detected in IF could be still detectable on genomic DNA extracted from same cell lines and same CPT treatments. We set up a protocol of dot blot to measure general level of R loop contents in genomic DNA extracted from control cells, using the s9.6 ab to detect RNA/DNA hybrids.

Fig 10 shows that R loop contents in genomic DNA extracted from cells treated with CPT 10 μM for 2' decrease of about 95% compare to control. These results were obtained at all the tested CPT concentrations and times of treatment (data not shown).
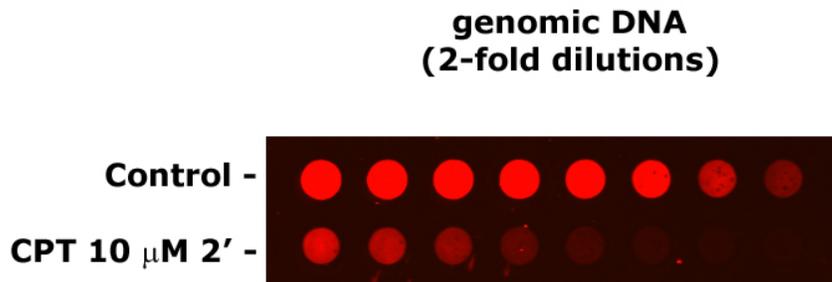

Fig 10. *R loops extraction is overall affected by CPT treatments.* Dot blot analysis using s9.6 ab performed on genomic DNA extracted from Ntera2 cells treated for 2' with 10 μM CPT

This data still in contrast with IF results lead us to consider the possibility that R loop structures after CPT treatment could be really unstable during the extraction step. The type of lysis used (TE-SDS 1%) leaves DNA with many Top1ccs, that after proteinase K become single strand breaks (SSBs). The treatment with proteinase K and the following nucleosome removal and the restriction enzymes activity used in

the DRIP protocol could lead to unwinding of R loop structures from the template DNA.

We tried to stabilize in some ways R loops by using crosslinking agents as PFA or by removing phenol/chloroform extraction, but in any case we failed to preserve R loops strictures following CPT treatment (data not shown).

Our data suggest that current technique to isolate and detect R loops structure cannot be used if cells have been treated with Top1 poisons such as CPT.

3.2.4 Top1 knockdown induces genome wide increase of R loop structures.

As genome wide mapping of CPT-induced R loops looked to be out of our possibility based on current DRIP techniques, we then focused our effort to determine the role of Top1 in R loop stability. With a siRNA approach we decided to understand how downregulation of Top1 may affect R loops distribution and stability. We first decided to check if Top1 depletion induced general increase in R loops cellular contents. Figure 11 shows dot blot analysis on genomic DNA extracted from HEK293 control and double transfected with scramble siRNA or Top1 specific siRNA. Top1 depletion leads to an overall increase of about 2-3 times as compare with scramble-transfected cells. These data are a first evidence that To1 may have an important role in the stability of DNA/RNa hybrids in living cells.
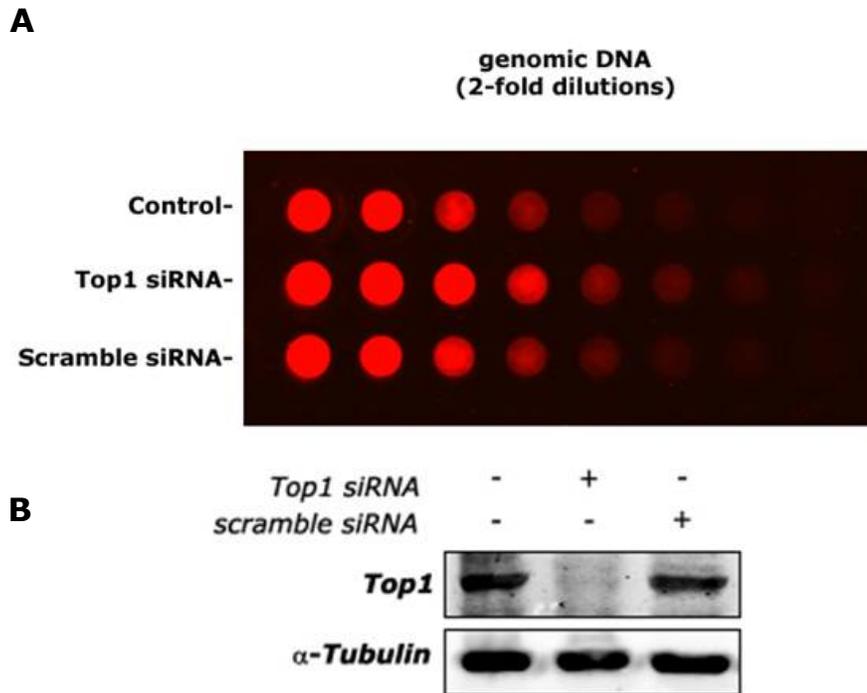
Fig 11. *Double Top1 knockdown induces overall increases of R loops strucutures.* A) Dot blot analysis using s9.6 ab performed on genomic DNA extracted from HEK293 cell control, double-transfected with Top1 specific siRNA or non targeting siRNA .B) Western blot analysis assessing Top1 downregulation

Interestingly only a double knockdown of Top1 increase R loops levels, while a single round of transfection did not change DNA/RNA hybrids abundance, even in the presence of a reduction of Top1 of about 90% (Fig11). Therefore, performing a second round of downregulation on the remaining 10% of Top1 could induce an additional decrease of Top1 protein level with percentage of remaining protein almost undetectable. Therefore downregulated levels of Top1 have to be close to 100% to stress R loops homeostasis. However, the limit of resolution given by WB analysis cannot assess the exact protein levels.

**genomic DNA
(2-fold dilutions)**

Scramble siRNA-

Top1 siRNA-

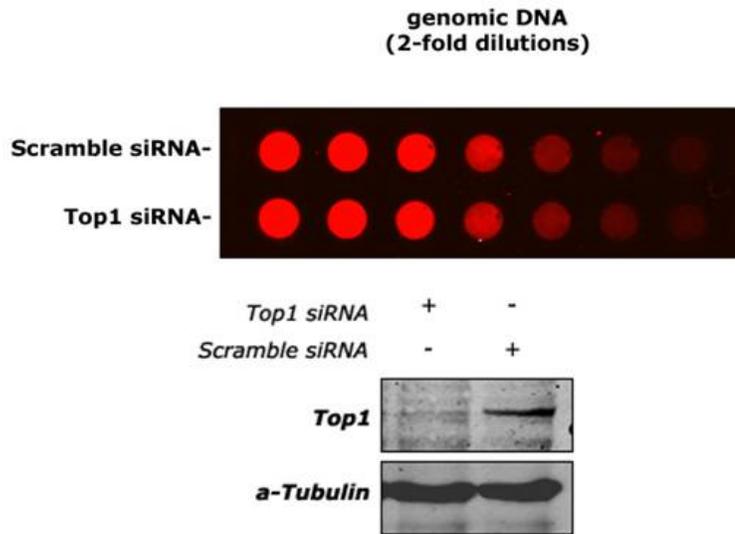|		
Top1 siRNA	+	-
Scramble siRNA	-	+

Top1

a-Tubulin

Fig 11. *Single Top1 knockdown does not induce overall increases of R loops strucutures.* A) Dot blot analysis using s9.6 ab performed on genomic DNA extracted from HEK293 cell control, trasfected with Top1 specific siRNA or non targeting siRNA (one round of transfection). B) Western blot analysis to assess Top1 kd.

Another possibility is that R loop homeostasis could show time-dependence for absence of Top1. Performing a double knockdown lead cells to be depleted for Top1 for about 72 hours, while a single round leave cells with reduced Top1 levels for only 24 hours. Anyway these data show Top1 tightly regulates genomic R loops.

3.2.5 Genome wide mapping of R loop structures after Top1 knockdown

Once we were able to detect a clear and strong effect of Top1 depletion on R loops abundance we decide to perform DRIP-seq and DRIPc-seq analysis to map genome wide these structures and understand which genes were mainly affected by Top1 depletion, in terms of R loop contents. We were also interested on what kind of modification is induced on these structures by Top1 knockdown.

DRIP and DRIPc are similar techniques, with some important differences. The former isolates and sequence the DNA that compose an R loop structure, whereas the latter is a derivative that allow to isolate the RNA molecule that is part of the hybrid resulting in a higher resolution and lower background. Additionally DRIPc

maintains strand specificity and can thus distinguish between sense and antisense R loops.

We first performed DRIP analysis on genomic DNA usedfor dot blot analysis shown in Figure 10a. Immunoprepicitated DNA was then used to build library and was sequenced trough Illumina Solexa Deep sequencing. Table1a shows some statistics about quality of the sequencing :

A	Sample	# Raw reads	# clean reads	# mapped reads	% mapped reads
DRIP-seq	Control	48443728	43264912	41153979	95.12%
	Scramble	53960771	49287510	46257679	93.85%
	Top1 kd	45320776	40466266	38989611	96.35%
B DRIPc seq	Control	47539735	42771663	34583165	80.86%
	Scramble	50107377	45319172	35633207	78.63%
	Top1 kd_1	53789320	48520636	43697302	90.06%
	Top1 kd_2	53090435	48430021	42531238	87.82%

Table1. Number of total, clean and mapped reads for A) DRIP-seq and B) DRIPc-seq data

Through peak calling analysis we were able to identify a comparable number of R loop peaks, with slight differences for Top1 kd sample. Notably, number of peaks was minor in Top1-downregulated cells, and this demonstrates that the increase following Top1 depletion detected by dot blot analysis cannot be explained with raising of new R loop structure.

A	sample	# of peaks	# gain of peaks (Top1 vs Sc)	# loss of peaks (Top1 vs Sc)
DRIP-seq	Control	54934	/	/
	Scramble	57456	/	/
	Top1 kd	42706	2141	2180
B DRIPc seq	Control	64444	/	/
	Scramble	67732	/	/
	Top1 kd_1	65795	8340	5725
	Top1 kd_2	60517	2536	2938

Table 2. Number of peaks in each sample for A) DRIP-seq, B) DRIPc-seq data

Interestingly, Top1 knockdown lead to a redistribution of R loop structures, with a loss of R loop peaks in 2141 loci and appearance of 2180 new r loop peaks (table2A). Notably these re-distributed structures were enriched in particular genic regions: particularly we found that gain of peak mainly occurs in gene body, while loss of peak was a feauture of 3' end of genes (data not shown). We selected two loci that show gain of peak (HIF1A, EEIF1A1 and four loci that show loss of peak (Fig12) and we validated them in DRIP qPCR, All the selected regions showed a good agreement with sequencing data (Fig13).

On the base of these interesting results we wanted to take advantage of the potential of DRIPc technique (higher resolution, minor background and strand information) and we decided to perform DRIPc-seq on two different biological replicates. Peak calling analysis identified a comparable number of peaks for all sample (table 2b) . Top1 knockdown leads to loss of 1798 peaks and appearing of 1658 new peaks (shared for both replicate). Again, in agreement with DRIP-seq data, we found that the increase detected by dot blot analysis in Top1 kd sample cannot be due to rising of new R loop structures in new genomic loci. So how can we explain increase of R loops content in Top1 depleted cells?
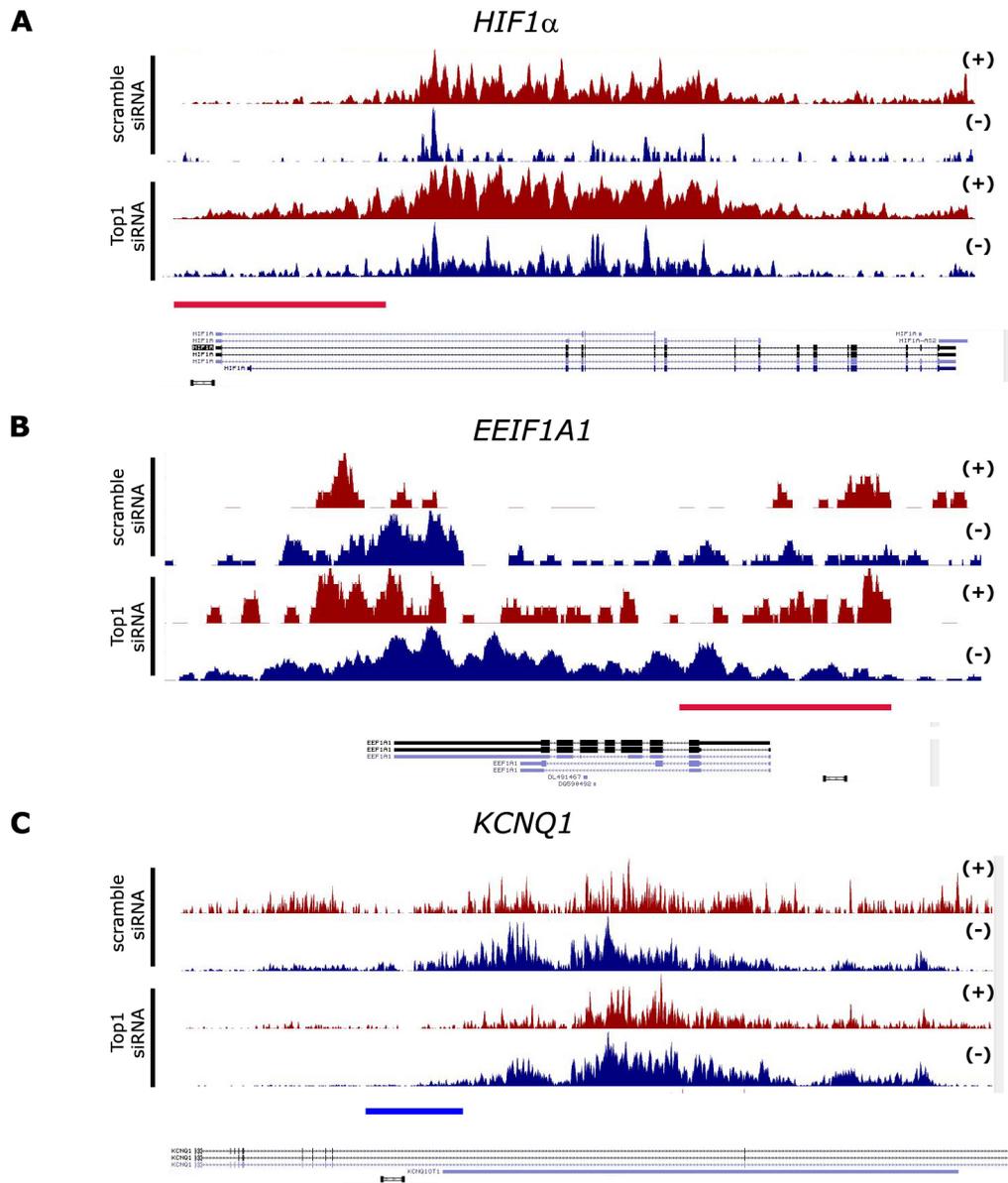
Fig12.*Loci selected for validation of sequencing data by DRIP-qPCR.* *Screenshots from UCSC genome browser of the selected loci: Gain of peaks (A,B) and Loss of Peaks (C,D,E,F, also on following page). The red bars represent gains of peaks recognized by peak calling software, the blue are the losses. Amplicons used for validation in DRIP-qPCR are shown as black and white bars. Red peaks are on positive strand, blue on negative strand.*

D

KCNQ2

E

ATXN10

F

GALNT1

Fig12.*Loci selected for validation of sequencing data by DRIP-qPCR. Screenshots from UCSC genome browser of the selected loci: Gain of peaks (A,B previous page) and Loss of Peaks (C,D,E,F). The red bars represent gains of peaks recognized by peak calling software, the blue are the losses. Amplicons used for validation in DRIP-qPCR are shown as black and white bars. Red peaks are on positive strand, blue on negative strand.*
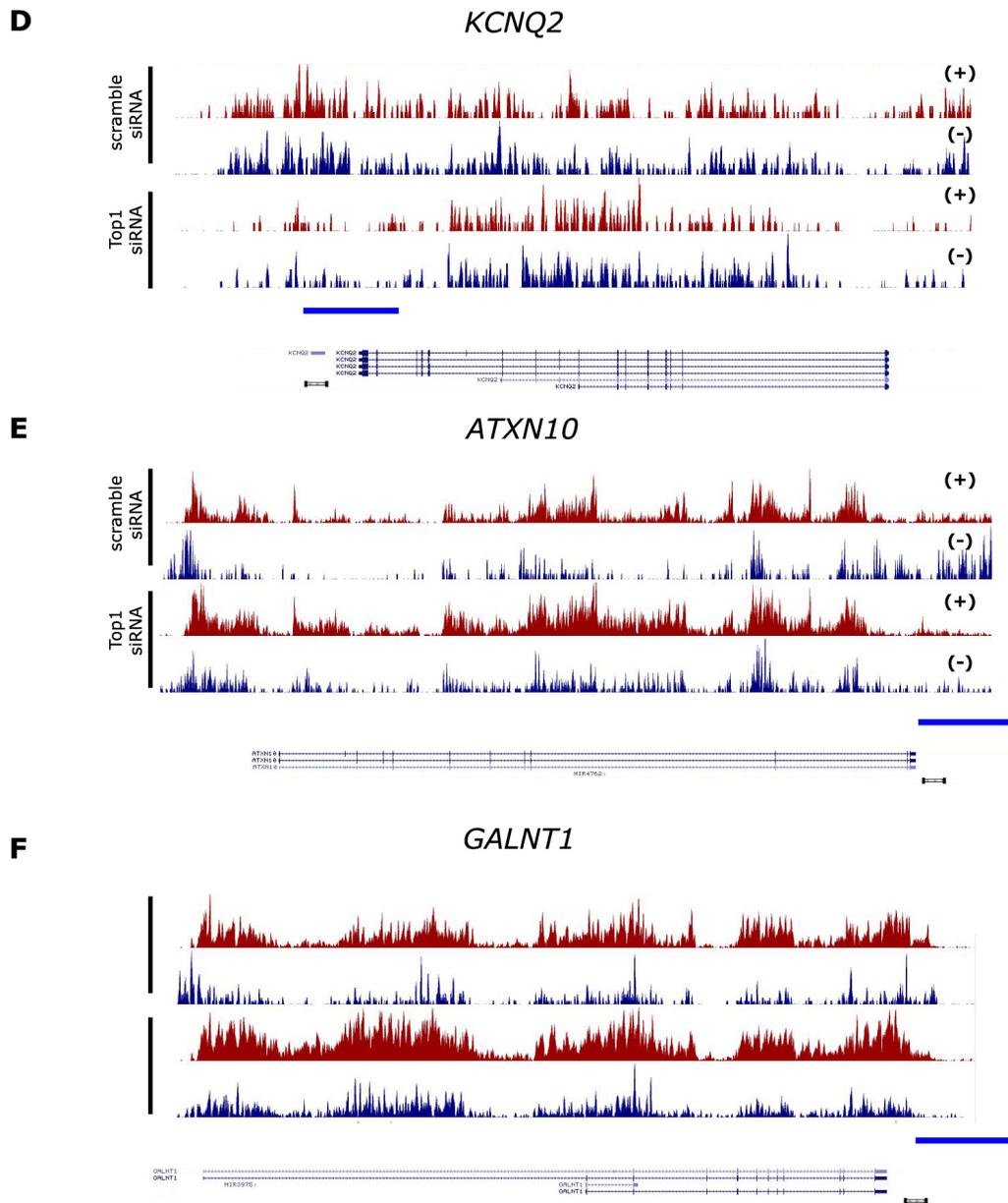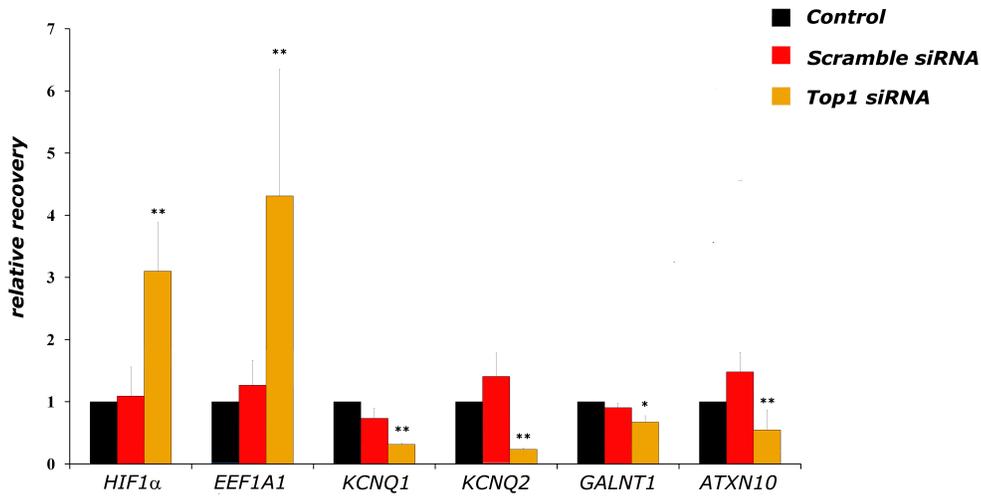
Fig 13. *Validation of gain and loss of R loop peaks after Top1 depletion.* Six loci were selected fromsequencing data and assessed for R loop revovery by DRIp qPCR on genomic DNA extracted from control, scramble trasfected, or Top1 deplted cells
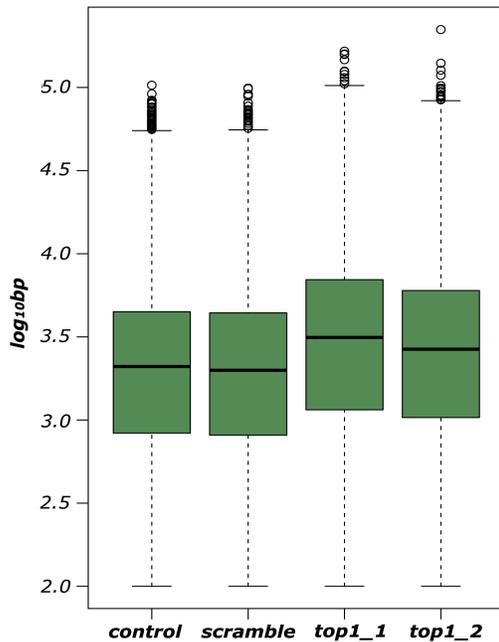

Fig 14.*Top1 depletion leads to a spreading of R loops structure.* Boxplot analysis showing distribution of length od DRIPc peaks in samples Control, Scramble-transfected and Top1 depleted cells (two biological replicates).

R loop can increase both in length and in frequency and these two phenomena could contribute to an overall and general increase. So, we first decided to measure and compare length of R loop total peaks. Box blot analysis showed in Fig 14, revealed that total peaks length distribution is significantly higher in Top1 depleted cells (P < 0,001 for both experiments). By reporting length of peaks relative to frequency formation we found that these longer peaks are also more frequently formed after Top1 depletion (Fig15).
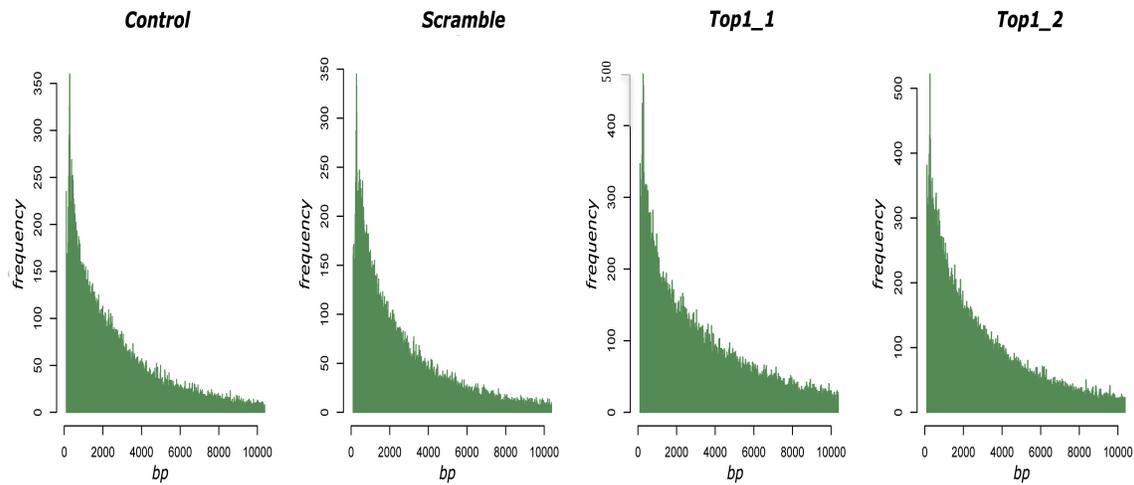
Fig15. *Top1 depletion increases frequency formation of longer R loops.* Distribution of peaks length relative to frequency formation in Control , Scramble, and Top1 depleted cells (two biological replicates)

Length of common DRIPc peaks
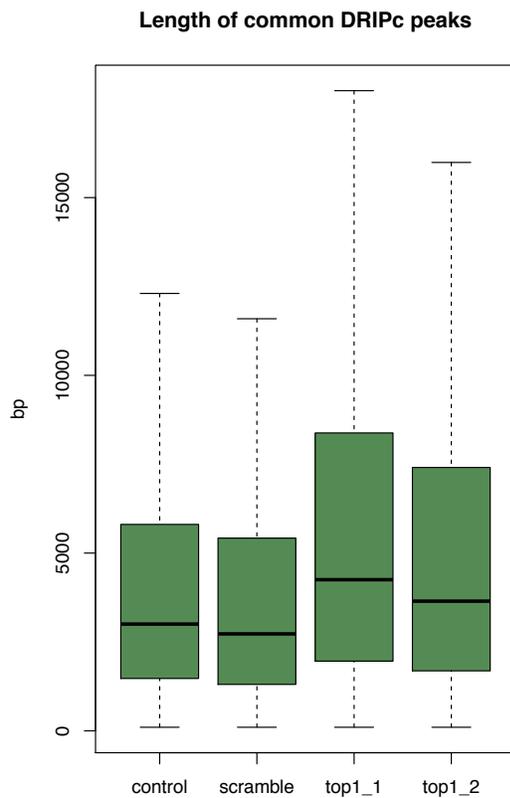
Fig15. *Top1 depletion induces spreading of common R loops.* Boxplot analysis showing averagd length of common peaks in control, scramble transfected cells and Top1 depleted cells.

Notably, difference in peak length can be detected even if we check distribution of peak's length for only common peaks (top1 vs scramble, Fig 16) demonstrating that increase in length is a genome wide effects, caused by change of common and classical R loop structures. We therefore concluded that Top1 depletion lead to spreading of common R loops across the genome.

Considering that increase on average length was not more than 40% as compare to scramble sample, these results could only partially explain the doubling of r loops content detected by dot blot analysis. To understand if a reduced Top1 activity

increases R loop formation frequency and in which part of genes this can occur, we performed metagene analysis. We calculated distribution of DRIPc signal in term of reads, around normalized genes. Every gene was divided in 10000 part to normalize each gene for its length, and DRIPc signal was used to generate a metaplot. Regions upstream and downstream to the gene body were also considered. Fig 17 shows that Top1 depleted cell revealed an higher R loop signal in the entire gene body giving a proof of increased R loop formation frequency.
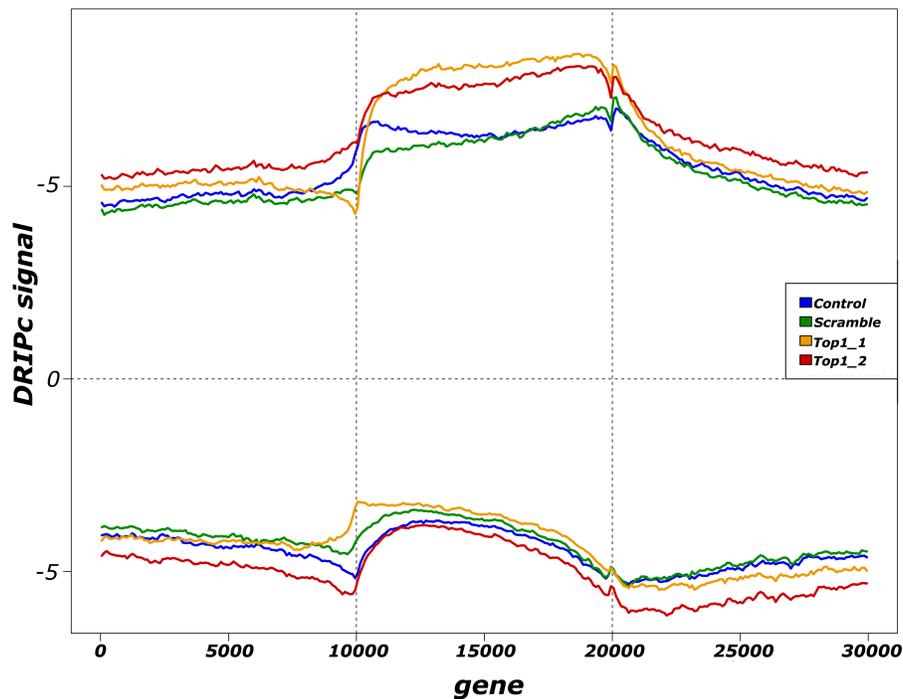
Fig 17. *Top1 depletion induces genome-wide increase of R loop structures in gene body.* Metagene analysis showing DRIPc-seq reads distribution along, upstream and downstream of genes. Each gene, upstream region and downstream region was normalized on 10000 parts.

We then decided to focus on loci that showed gain and loss of R loops peaks after topoisomerase knockdown. We noticed that these genes were particularly stressed in terms of R loops modification after Top1 depletion, with a most pronounced effect . Metagene analysis, performed as prevoulsy shown, but using only genes that show gain of R loop peak, revealed that these new R loops are part of a general and very intense increase that involves the entire gene body (Fig 18 A).

64

Fig 18. *Gain of R loop peaks after Top1 depletion occurs on the entire gene body.* A) Metagene analysis of genes showing gain of R loop peaks after Top1 kd. Each gene, upstream region and downstream region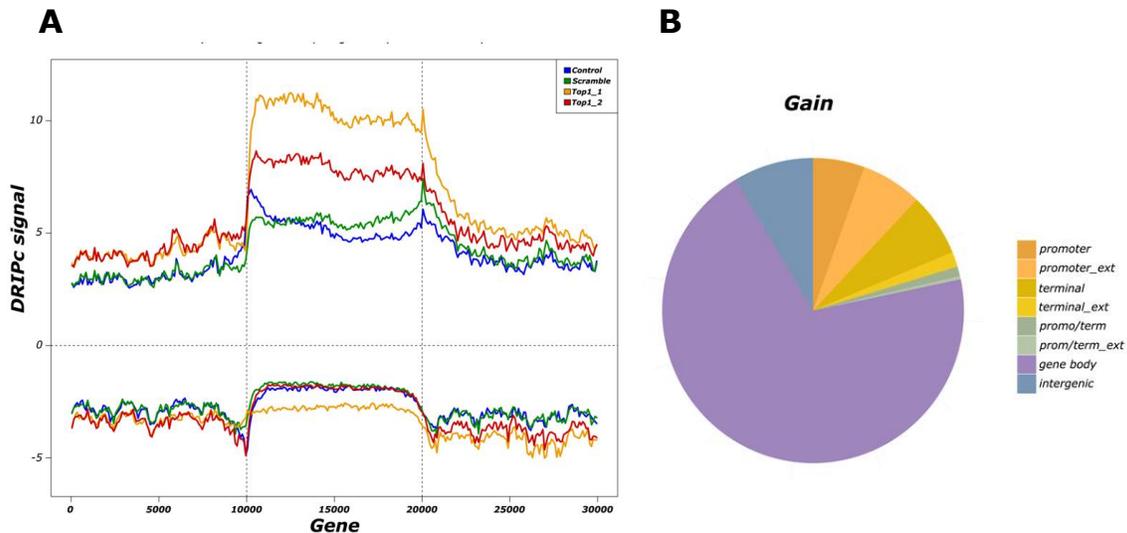 was normalized on 10000 parts. B) Peak location analysis showing genomic distribution of new R loop peaks induced by Top1 depletion.

This data was confirmed also by peak location analysis that showed that gain of R loop peak are not randomly distributed in the genome, but are strongly enriched in gene body (73% of total distribution) (Fig 18b). Interestingly and differently from gain of peaks, when we focused on genes with loss of peaks after Top1 knockdown, metagene analysis revealed that loss of R loop mainly occurs at 3' ends of genes (Fig 19a). Again, peak location analysis confirmed the same result showing an important enrichement at terminator sites (more then 30% of total distribution) (Fig 19b). We were really surprised to notice that the R loop profile on these genes is unusual compare to all the other genes, with an R loop peak really high at level of TTS (Fig 19a). These data suggest that probably loci tha show loss of R loop peaks after Top1 kd are a particular subset of genes with particular but unknown genes.
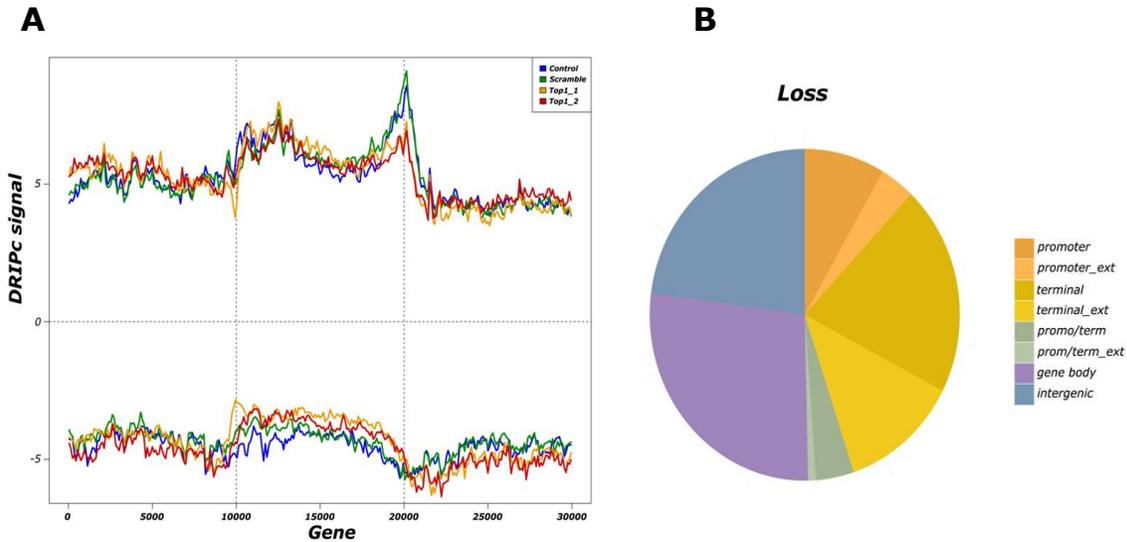
Fig 19. Loss *of R loop peaks after Top1 depletion occurs at 3' and of genes* . A) Metagene analysis of genes showing gain of R loop peaks after Top1 kd. Each gene, upstream region and downstream region was normalized on 10000 parts. B) Peak location analysis showing genomic distribution of R loop peaks reduced by Top1 depletion

We also noticed that genes that showed loss and gain of peaks were particularly long with many long introns. To confirm this hypothesis we checked averaged distribution of lengths of these genes. As controls we decided to analyze median length of all genes and all genes showing DRIPc peaks. As explained by boxplot in Figure 20, we found that genes with gain of peaks are incredibly longer than both all genes and all genes with DRIPc peaks ($p < 2.2e^{-16}$). Same results, but in minor extent were found for genes that show loss of peak ($p < 1.33e^{-13}$). This data is perfectly in agreement with the critical Role that Top1 has in regulating supercoil of really long genes (52).

Since the directionality of transcription can generate different kind of torsional stresses, therefore affecting differently R loops, we
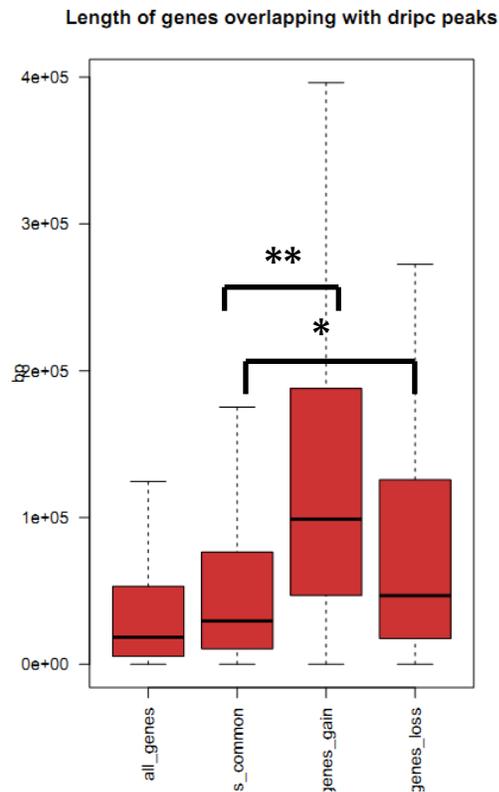

Fig **20.** *Top1 depletion affects R loop homeostasis particularly at level of really long genes.* Boxplot with averaged distribution of length of all genes, all genes with DRIPc signal and gain and loss of R loop peaks.

66

wondered if gain or loss of R loop peaks could show a dependence on gene directionality. First, we checked on the genome browser UCSC the first 250 loci that show loss of peaks and the first 250 showing gain after Top1 knockdown. Then, we assigned them to one of the following categories according to the gene direction and genomic position: we defined two genes convergent when their TTS (transcription termination site) were colliding, unidirectional when their TSS showed same direction, divergent when gene direction and TSS showed opposite direction. We created a fourth category, for centromeric region, since these genomic loci were frequently represented in the group of loss of peaks but directionality cannot be assigned as they are untranscribed regions. Strikingly we found that loss of peaks was particularly enriched for convergent genes (67% of total) (Fig 21). Notably, when we focus on gain of peaks, the percentage of convergent genes drop to less than 30% and most represented class becomes the unidirectional genes (Fig 22).
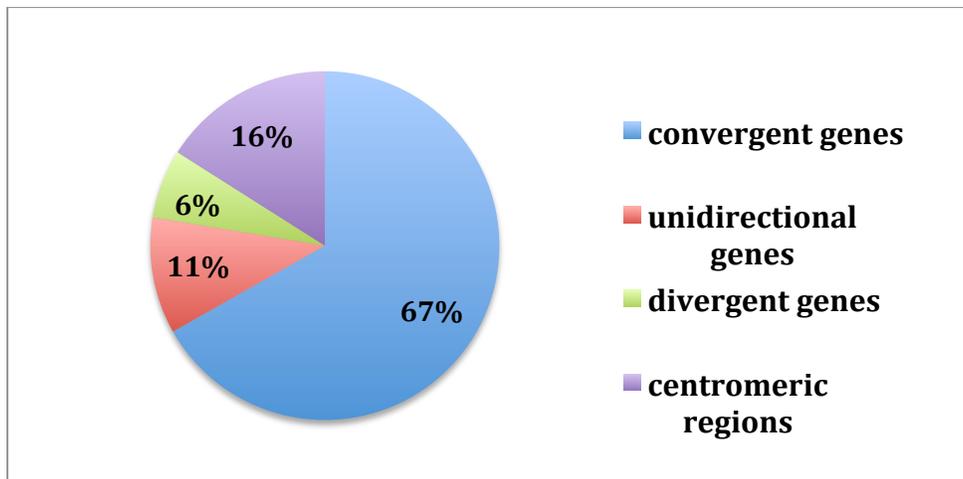

Fig.21. *Loss of peaks mainly occurs at convergent genes. Pie chart of the distribution of 250 loci showing loss of peaks according to gene direction.*
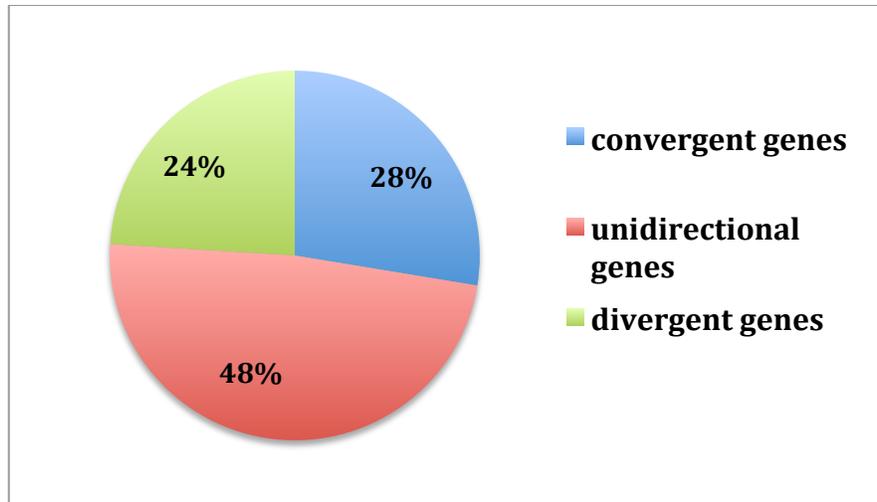
Fig.22. *Gain of peaks after Top1 depletion mainly occurs at unidirectional genes.* *Pie chart of the distribution of 250 loci showing gain of peaks according to gene direction.*

To validate these data bioinformatically and at genome wide level, we decided to calculate genomic distances for these 4 categories:

1) Distance from TSS to upstream TTS (unidirectional genes)

2) Distance from TSS to upstream TTS (divergent genes)

3) Distance from TTS to downstream TSS (unidirectional genes)

4) Distance from TTS to downstream TTS (convergent genes)

We decided to calculate these distances in the following groups:

- All genes

- All genes with DRIPc peaks

- Genes with gain of peaks

- Genes with loss of peaks

The results are shown in Figure 23. First, we found that genes showing gain of R loop peaks were usually "alone genes", very far from potential neighbors. This was seen for all the four analyzed categories (Fig23 A, B, C, D). Strikingly, focusing on loss of peaks, we found that, only in the category of colliding genes, these loci were particularly close to other genes, as compare to all convergent genes and all convergent genes with DRIPc peaks (40 kb for loss, 111 kb for all genes, 68 kb for all genes with DRIPc peaks). On the basis of these data we can concluded that loss of peaks seems to be a feature of convergent genes.

68
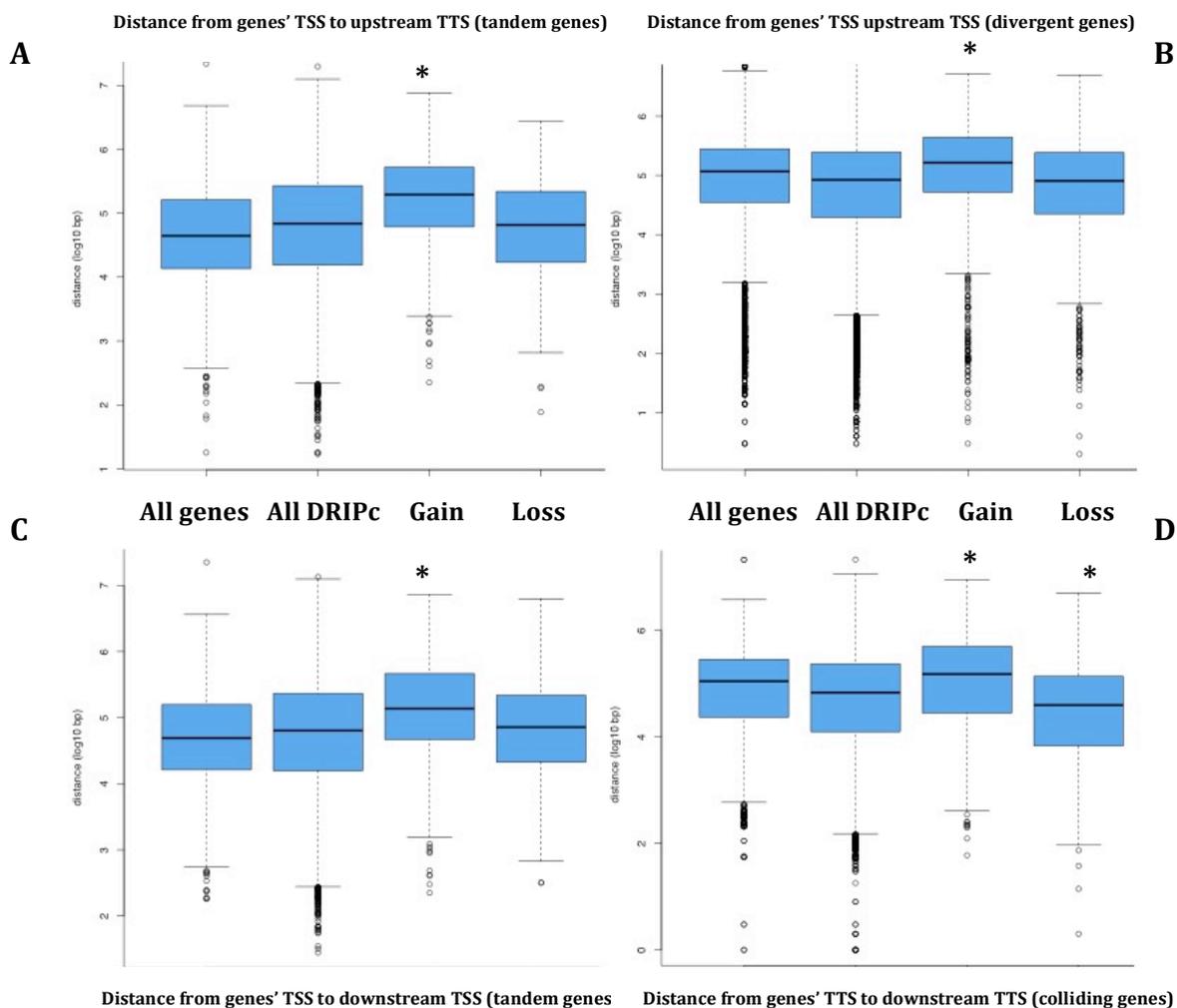
Fig 23. **Genes with loss of peaks tend to be convergent, while genes with gain tend to be isolated genes.** Boxplot analysis showing distribution of distances between genes unidirectional (a and c), divergent (b) and convergent (d). * = at least to P< 0.001

All these data revealed that Top1 could regulate R loop stability in different ways, probably dependently from different and specific genomic features. However, we wanted to better understand what is the effect of these R loop dynamics on gene expression. Additionally, determining gene directionality and therefore a possible positional effect on R loops dynamics needed that our data had to be filter for transcription levels. Moreover it was interesting to understand which class of genes (high, mild or low expression levels) was mainly affected by Top1 knockdown in terms of R loops contents. Finally, we also

wanted to correlate modifications in R loops with change in RNA Polymerase II distribution. To obtain all these information we performed RNAseq and PoII-ChIPseq in cells double transfected with scramble siRNA or depleted for Top1. Bioinformatics analysis on these genomic dataset is not completed yet. However we can still make some interesting conclusion by checking our data on UCSC genome browser. However these observations have to be obviously confirmed on large scale by bioinformatics analysis. As a first step, we focused only on gain and loss of peaks. Strikingly, we found that transcription and PolII distribution were not particularly affected for gene showing gain of R loop peaks (data not shown). Notably the case was different when we focused on genes showing loss of peaks. PolII distribution seemed to be affected mainly at 3' end of these genes, in Top1 depleted cells. In particular, the peak of PolII typically seen at terminator pausing site resulted shifted of few kilobases upstream of TTS. Downstream of this region, in correspondence of the effective loss of peaks and further downstream, PolII seems to reduce progressively. Upstream of terminator site PolII occupancy doesn't seem to decrease and in some cases even an increase can be detected (Fig 24a, b, upper part of the panel). These data suggest that probably, in the presence of a loss of peak, PolII stucks at level of TSS. RNAseq data were at first sight of difficult interpretation. FPKM calculation restricted only to genes with loss of R loops peaks did not reveal great change in expression level. However analysis on mapped reads distribution on the UCSC genome browser showed that specific changes in the RNA molecule happen at 3' end of these genes. Particularly the last exon seems to decrease after Top1 knockdown. Additionally, RNA reads downstream of the 3'UTR always decrease (Fig 24 a.b lower panel). These reads (less in number compare to exon reads) are usually due to a polymerase that is loosing its processivity after transcribing the PolyA site, so could probably suggest potential impairment in the polyadenilation and termination process. However these data have to be confirmed on large scale by bioinformatics analysis.
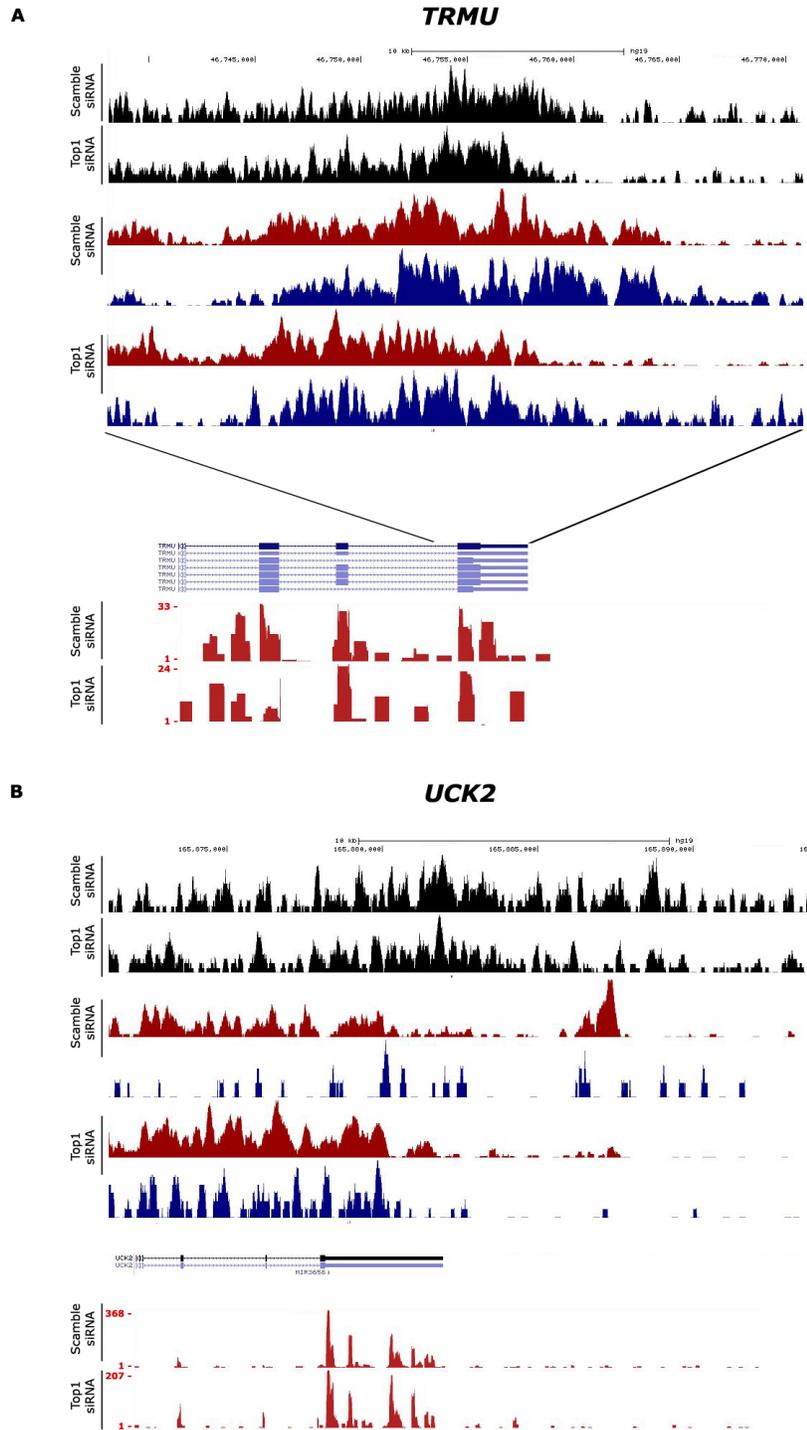
Fig 24. *Loss of peaks after Top1 knockdown seems to correlate with a block of PolII at TTS and an impairment of transcript specifically on last exon and in the termination regions*. Screenshots from UCSC showing PolII distribution, R loop peaks and RNAseq reads. RNAseq data report the y axis to show the decrease in transcript level. A) TRMU gene, B) UCK2 gene

Chapter IV

DISCUSSION

In this work we investigated the role of two natural compounds Triptolide (TPL) and Campthotecin (CPT) and we characterized the effects of inhibition of their specific target (XPB and Top1) at transcriptional level.

Particularly, as regards TPL, we tried to understand the mechanism that underlay the RNA Polymerase II stability during TPL treatments. We identified CDK7 as a new molecular actor in the Rbp1 degradation mechanism induced by TPL.

As regards CPT, we tried to understand what are the consequences of Top1cc formation at actively transcribed regions, particularly at divergent promoters. We aimed our efforts to understand how Top1 and its inhibition by CPT or its depletion affect the stability of R loops structures. Finally, we mapped R loops genome-wide after Top1 depletion and correlate R loop distribution changes with genic organization. The last part, performed in collaboration with Frederic Chedin's lab, will need to be fully analyzed with other data such as expression data and genomic RNA Polymerase II maps.

4.1 Triptolide and TFIIH

Triptolide (TPL) is a strong transcriptional inhibitor whose mechanism of action and effects are not fully understood.

The natural target of TPL is XPB an helicase which is part of the transcriptional factor TFIIH (33). TFIIH is a complex of ten subunits and among these it includes CDK7. TFIIH is a factor involved both in transcriptional initiation and nucleotide excision repair. Therefore, by inhibiting XPB, TPL blocks both transcription and NER pathway (33). Here we provide novel insights regarding RNA polymerase II stability during TPL treatments, showing new molecular aspects of the mechanism of action of the drug. Particularly we found that TPL triggers a CDK7-mediated proteasome-executed Rbp1 degradation. We demonstrated that the time-dependent degradation

is preceded by a biphasic change of Ser5 phosphorylation on CTD of PolII. Particularly, Ser5 phosphorylation levels rise at short time of treatments (100 nM 2 h) and then decline at longer time. This hyperphosphylation is mediated by CDK7 and knockdown of this enzyme reverses both hyperphorylation at Ser5 and Rbp1 degradation induced by the drug. This suggests that Ser5 phosphoprylation is the molecular signal responsible for PolII stability during TPL treatments. Some published reports are controversial about the hyperphosphorylation induced by TPL. Titov et al., for example, did not find any change in phosphorylation level during TPL treatments (33). It is possible that their experimental conditions (200 nM 1 h and 4 hours, A549 cells) are not the optimal ones to detect the biphasic change on Ser5. Wang et al were instead able to detect the hyperphosphorylation events but not to identify either the site or the enzyme responsible for this hyperphosphorylation (32). Our data show that Ser2-P levels do not change until late times of treatments (4-6 hours) when hypophosphorylated levels for this residue can be detected. These data suggested the possibility that the PolII elongation was in some ways affected during TPL treatments. ChIP experiments revealed that at really short times of treatment, PolII levels at active promoters are unaltered while on the corresponding exon a significative decrease can be overall detected. These data suggest that at short times of treatment TPL induces a block of RNA Polymerase II at class II promoters.

IP experiments performed in collaboration with Ze Hong Miao's lab demonstrated that TPL induces also ubiquitination of RNA Polymerase II (96). These data along with CDK7 mediated hyperphosporylation of Ser5 and block of RNA Polymerase II at promoters, lead to important considerations. The degradation of PolII after treatment with transcriptional inhibitors (alpha amanitin) or DNA damage agents (H_2O_2) was already previously showed (22). It seems that hyperphosphorylation of CTD can be a molecular signal to induce PolII degradation. Our data show for the first time that this mechanism can be valid also for inhibition of an helicase as XPB and its complex TFIIH. It is possible that inhibition of ATPase activity of XPB could impede promoter clearance of PolII. A stalled polymerase could be phosphorylated at CTD Ser5 trough an activated CDK7 leading to ubiquitination and degradation of

the PolII largest subunit Rbp1. Alternatively, the hyperphosphorylation of Ser5 residue after TPL inhibition of XPB could induce a PolII stalling and a consequent ubiquitination and degradation. However, additional experiments are required to understand if Ser5 phosphorylation is the cause or the consequence of PolII stalling. TFIIH has also an ubiquitin ligase activity due to the presence of p44 subunit in this complex. So TFIIH potentially possesses all the enzymatic activity (ATPase/helicase for XPB, kinase activity for CDK7 and ubiquitin ligase activity for p44) that can mechanistically explain the molecular effects of TPL. In addition to the standard heptapeptide consensus sequence of $Y_1S_2P_3T_4S_5P_6S_7$, there are several heptapeptide variants including variants containing a lysine residue in position 7 besides the serine residue in position 5. These sites could be the potential targets of the E3 ligase for ubiquitination (Lys7; (97)) and Cdk7 for phosphorylation (Ser-5), respectively (Fig 25). Further investigations are necessary to identify the sites and the enzymes responsible for Rpb1 ubiquitination following TPL treatment.
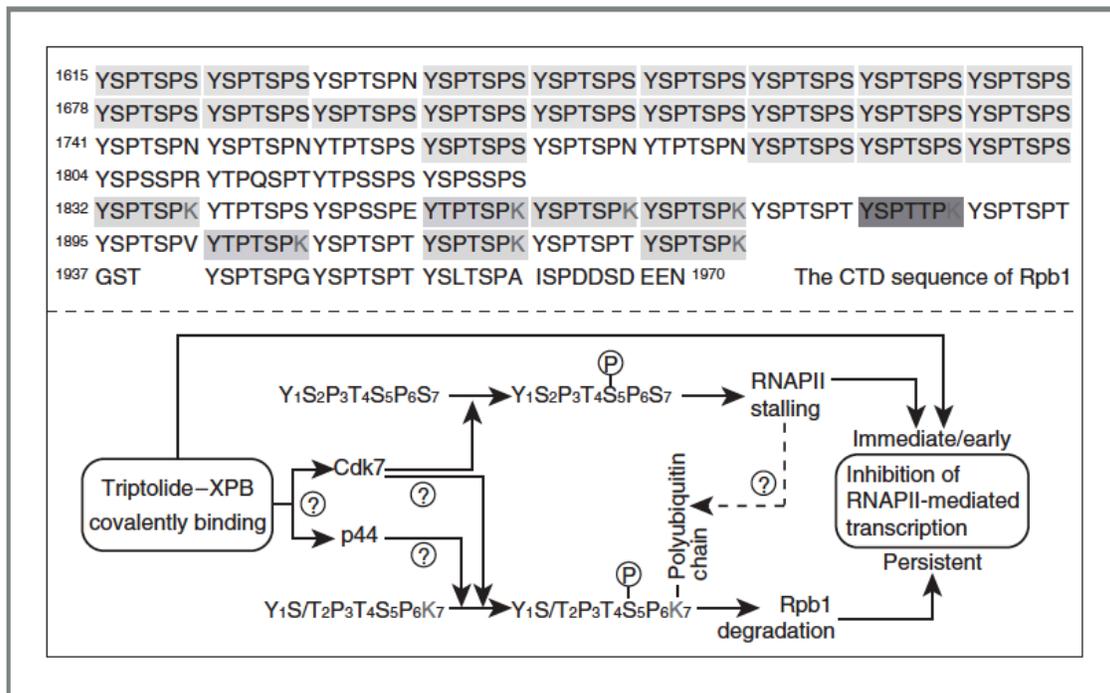
Fig 25. *Model for the mechanism of action of Triptolide. Figure from Manzo SG, Zhou ZL, Wang YQ, Marinello J, He JX, Li YC, Ding J, Capranico G, Miao ZH.Natural product TPL mediates cancer cell death by triggering CDK7-dependent degradation of RNA polymerase II. Cancer Res. 2012 Oct 15;72(20):5363-73.*

Our work show that TPL possesses distinct properties that makes this drug unique compare to all the other already-known transcriptional inhibitors (23). In conclusion TPL could serve as a powerful tool to investigate the regulation and function of TFIIH, and act as a model compound for anticancer drug development specifically targeting initiation factors as TFIIH.

4.2 Camptothecin, Top1 and R loops.

Our data revealed a new and unexpected role of Top1 in regulating transcription. Particularly we found that pharmacological inhibition of Top1 by camptothecin (CPT) deregulates transcription rates at divergent CpG Island promoters favoring the enhancement of antisense transcription. We found that transient and stable knockdown of Top1 reduces the CPT induced-antisense transcription. Interestingly, simple Top1 depletion by RNA interference does not induce antisense activation, suggesting that this phenomenon is mainly caused by Top1cc formation and not simply by reduction of Top1 activity. Surprisingly the phenomenon seems to be replication independent since the co-treatment with DNA Polymerase inhibitor aphidicolin does not suppress antisense transcripts. Divergent transcription has been reported in several eukaryotic cells. For example it has been shown that the majority of TSSs in murine embryonic stem cells are characterized by an active bidirectional transcription (98). Divergent transcription is a tightly regulated process, with the involvement of different factors. The positive transcription elongation factor PTEF-b has been found to play a critical role in regulating antisense transcription (98). Here we found that downregulation of CDK9 reduces CPT-induced antisense transcription.

CDK9 depletion and its pharmacological inhibition by DRB both reduce antisense transcript level, but they do not totally suppress CPT induced-enhancement of them. These data suggest the possibility that antisense transcript could be the result of a degradation impaired by CPT treatment. Interestingly, the reduction of the RNA degradation pathway by downregulating exosome components can increase promoter associated antisense transcription in mammalian cells (99). Further investigations need to be done to assess potential functional interaction between Top1 and exosome pathway. Another possibility is that Top1 inhibition by CPT could impair promoter directionality by interfering with the U1/PAS pathway. It has been shown that biderectionality of promoters is regulated by two important elements: a polyadenilation signal (PAS) upstream of TSS and a U1small nuclear ribonucleoprotein (snRNP) recognition site downstream of TSS (100). The PAS signal upstream induces early termination and cleavage of the antisense transcript, while the U1 binding site promotes correct transcription and suppresses early termination in the "sense" direction. Intriguingly,

76

interference with the U1 snRNP unbalances this mechanism, by inducing early termination in the sense direction and increased antisense transcription upstream of TSS in a manner really similar to what we saw for CPT treatment (95), (100). Strikingly, Top1 inhibition by CPT has been shown to affect phosphorylation and activity of U1 snRNP (101), (102).

The antisense transcription induced by CPT seems to be a specific consequence of Top1ccs. These complexes are rapidly formed during CPT treatment. Our results show that 2' of treatment with high doses of camptothecin are enough to detect clearly Top1cc formation (Fig 8). Strikingly, Top1cc formation parallels an increase of non B DNA structure called R loops at highly transcribed regions such as nucleoli and mitochondria, in all the tested cell lines. Differently from Top1cc that constantly increase during CPT treatments, R loops stabilized by Top1 inhibition show a biphasic change with rising at short time of treatments (2-10') and decreasing at longer times (1 hour). These data suggest that R loops during camptothecin treatments are rapidly removed after an initial stabilization, probably because they are dangerous structures for cell. However, some R loops foci remain relatively high until later stage (Fig 8a) suggesting that cells are not able to remove these structures at all genomic sites. The nature of these foci remains unclear. Moreover, during Top1cc formation due to CPT, Top1 is heavily modified trough sumoylation and ubiquitination (59 and Fig 8). One possibility is that R loops stabilized by CPT decrease following Top1 modifications and degradation, being in this way, a crucial part of Top1cc repair pathway. Howecer, how these modifications and if potential removal of Top1 by proteasome could play a role in R loops resolving has to be elucidated.

R loops could be a link to bidirectionality of promoters too. First, a bidirectional promoter is probably characterized by relative high levels of negative supercoiling, generated by two molecules of RNA polymerases that transcribe in opposite directions. This condition could thermodynamically favor R loop formation. Second, by analyzing DRIPc data we found that bidirectional promoters show presence of antisense R loops upstream of TSS and that some of our antisense transcripts overlap to DNA segments known to form R-loops. Additionally 185 of 246 promoter-associated antisense transcripts overlap with or are within 3000 bp distant from an R loop-prone region (GC

skewed region). Recently it has been shown that R loops promote efficient termination of transcription at 3' ends of genes by inducing repressive epigenetic marks (88). As mentioned above, bidirectional promoters usually show a PAS sequence upstream of TSS to early terminate antisense transcription. Notably U1 binding site sequence is highly GC skewed (Frederic Chedin, personal communication), therefore prone to R loop formation. Top1 inhibition could in some way interfere with this pathway via R loop, therefore inducing antisense transcription. Unfortunately the impossibility to isolate and map R loops following CPT treatments, with current genome wide techniques (DRIP, DRIPc) limits a lot the possibility to clarify and connect Top1cc formation with R loop and antisense transcription.

The difficult to isolate R loops from cells treated with camptothecin merits few considerations. IF results showed clearly the increase of R loops after CPT treatment. Thus, why cannot we isolate the structures following CPT treatment? The first explanation is that the presence of SSBs derived from Top1 cc could make the R loops really unstable. In literature it has been shown that the presence of a nick on the non-template DNA can favor R loop formation leading the RNA (65) to displace the non-template DNA, but how much stable is an R loop in the presence of a partially broken DNA is not understood. Additionally, in transcribed regions Top1 cuts on the DNA template strand with a certain preference (L. Baranello and D. Levens, personal communication). Nobody knows how a nick on template DNA can affect R loops stability. Probably, the action of proteinase K and restriction enzymes used in DRIP protocol could lead to an unwinding of the R loop, driven by processes such as nucleosome removal and DNA endonucleasic cleavage. However, further work needs to be done to define this topic.

Although mapping of R loops structure following CPT treatments resulted out of our possibilities Top1 depletion experiments revealed important findings. First, R loops increase significantly in absence of Top1 (Fig 11). It is interesting that a single round of knockdown does not alter R loops levels, even in the presence of a strong reduction of Top1 (Fig 11). Notably, this effect can be detected only after a double knockdown. However, we probably found the way to critically downregulate Top1 and see an effect on R loop contents. DRIPc-seq analysis revealed really interesting results, some of them

78

totally unexpected. The first remarkable result is that Top1 depletion does not induce arising of new R loops structures in new loci. The number of peaks identified in each sample, which resulted really comparable, can prove this. Considering this data and the high increase detected by dot blot analysis, it looks that the increase of R loops structure after Top1 depletion is mainly a consequence of a stabilization and increase of frequency formation of preexisting R loops (Fig 15-16). Notably, R loops can also increase in length after Top1 depletion suggesting that an increased negative supercoiling state can favor a further reannealing of the nascent transcript therefore extending a pre-formed R loop.

Metagene analysis revealed that Top1 depletion increases R loop frequency at a genome wide level. Surprisingly, this phenomenon regards the entire gene body. These data is completely new and totally unexpected and show a general role for Top1 in regulating R loops. Tuduri et al (70) demonstrated that Top1 depletion lead to interference between transcription and replication fork, resulting in an S-phase specific genomic instability. Notably, these effects are reverted in the presence of an overexpression of RNAseH, suggesting an involvement of R loops structures (70). However, they did not provide direct evidences of an increase in R loop contents. Here we showed for the first time that, overall, Top1 depletion effectively stabilize R loops and that their frequency formation is increased in the entire gene body, whereas R loops levels remain quite similar to controls in promoter and terminator regions (Fig 17). During top1 knockdown experiments, we noticed a reduction in the doubling times for these cells, concordantly with the slowed replication forks and alteration of S-phase reported by Tuduri et al (70). It will be really intriguing to understand what is the consequence of an R loop stabilized by Top1 depletion in cell cycle phases other than S phase.

It is really interesting that R loops stabilized in the presence of CPT are rapidly removed, while Top1 depleted cells show increase of R loop structures for long times (up to 96 hours after second round of transfection, data not shown). We think that Top1 inhibition by CPT and Top1 depletion by siRNA have different effects, with R loops stabilized by CPT being definitively more dangerous and therefore rapidly removed. Further investigations and characterization of R loops stabilized by CPT with genome-wide techniques are needed to establish the exact molecular events and mechanisms.

We were really surprised to find some loci that looked particularly stressed in terms of R

loop contents as we have found regions with "new" or "lost" R loops after Top1 knockdown. However these R loops are not unique and isolated structures that appear or disappear after Top1 knockdown, but they look as a part of a change in length and frequency formation in pre-existing R loops (Fig 12).

Particularly, as regards new R loop peaks, this phenomenon seems to be part of a general increase in frequency and a spreading along the entire gene length and further downstream and upstream (Fig 12A and B). Genes showing gain of R loop peaks resulted the ones with the biggest increase in R loop formation frequency (compare Fig 18 with Fig 17 and 19), so they are probably the most affected by the Top1 depletion in terms of R loops contents. Again, the increase in R loops involves the entire gene length. These genes resulted to be really long compare to average length of genes with DRIPc signal. Top1 has been seen to be critical for transcription of really long genes (52). However, we did not find particular change in PolII distribution and RNA transcript levels for these genes (data not shown). Interestingly, from genomic positional analysis, these genes resulted to be really far from gene neighbors and mainly located in genes-poor regions. Helmrich et al. demonstrated that collision between replication and transcription machinery is inevitable for genes longer that 800 kb (as transcription of these genes spans more than the entire cell cycle) (103). Interestingly, these collisions seem to be the source of common fragile sites (CFS) and involve R loop formation. Genes with "gain" of peaks are not as long as genes studied in Helmrich's works, but could still represent possible genomically unstable loci for the strong stabilization of R loops driven by Top1 depletion. However, we did not identified DNA damage sites and therefore the role of Top1 on R loops and genome instability at these specific loci need further and investigations.

Genes showing the loss of peaks after Top1 depletion resulted the most interesting, with completely unexpected results. Again, as it was for gains of R loops, these sites of R loops suppression seem to be part of modification of preexisting structures that change in shape after Top1 depletion. Loss of peaks usually coincides with a shortening of R loop molecule at 3' end of the gene (see Fig 12 C for a clear example). Interestingly, it seems that upstream of the loss site R loops tend to increase (Fig 12 C, E, F) suggesting a possible contraction of the R loop on the entire gene length. Literature have always

consider Top1 as a "preventing R loop factor" due to its possibility to remove negative supercoils that usually favor R loop formation (65). Here, we show that the story is not as simple and that for a subset of genes this is not the case. Strikingly, the loss of peaks usually occurs at 3' ends of gene (Fig 19), suggesting this phenomenon is specific for this region. The data are in agreement with the role that Top1 has in regulating supercoil during transcription, and may suggest that torsional stress created by transcription machinery accumulates at 3' end of genes and cannot be dissipated due to the absence of Top1. Genes showing a loss of peaks resulted to be long genes, although in a minor extent as compare to genes with gain of peaks (Fig 20). Again, it seems that on this kind of genes Top1 could have a particular role. But why these two opposite effects (gain/loss) on similar genes? The answer probably comes when we focus to gene directions and gene distances. Surprisingly, genes showing loss of peaks resulted to be convergent genes (Fig 21) and interestingly, their TTS are closer compare to average distance between all convergent genes with DRIPc signal (Fig 22). These data bring to light the possibility that suppression of R loops could be the consequence of hyperpositive supercoil accumulated in the convergence region. According to the twin supercoiled model, a transcribing RNA polymerase generates positive supercoil ahead and negative supercoils behind itself (42). In the case of a convergence region the positive supercoils generated by two colliding RNA Polymerase could be probably superimposed leading to a torsional stress particularly strong and difficult to dissipate. This, in case of Top1 depletion, could lead to R loops suppression. A particular consideration has to be done for a small subset of loci showing loss of R loop peaks: centromeric regions (Fig 21). It is not clear why Top1 depletion should lead to reduction of R loops in these loci, especially if we consider that are usually untranscribed regions. R loops have been show to play a role in maintaining eterochromatinic states (104), (105). It would be really interesting to investigate if Top1 plays a role in this process. We also have to consider the possibility that these losses of peaks at centromeric site could be an artifact of the sequencing, as repeated regions can introduce important bias in deep sequencing analysis (106).

However, even excluding centromeric regions, genes with loss of peaks still remain the most interesting category in terms of R loops modulation induced by Top1 depletion. Metagene analysis revealed that these genes have a unique DRIPc profile, with a

distribution of R loops particularly high at TTS (Fig 19). R loops at terminator sites have been shown to play an important role in transcriptional termination process (19), (88), (89), R loop formation at terminator should slow down the elongating polymerase therefore favoring the termination process. Senataxin should then resolve the R loop formed at this region, favoring the Xrn2-mediated degradation o the RNA still attached to PolII (19). The presence of such high levels of R loops at loss of peak sites suggest that probably in these genes termination mechanism via-R loops is particularly reinforced. This correlates quite well with the observation that these genes are close to other convergent genes. Therefore, it seems that a good terminator site is required to avoid collision of convergent transcribing polymerases.

But how Top1 depletion affects RNA polymerase II distribution and transcription at level of R loops loss sites? Although the bioinformatics analysis for our PolII-ChIP-seq and RNA-seq is not completed yet, we found important differences in these two parameters. Loci with loss of R loop at 3' end of gene showed a decrease of PolII density downstream of the R loop loss site. Additionally, at level of TTS PolII paused peak seems to be shifted upstream of few kilobases in absence of Top1 (Fig 24). Therefore Top1 depletion seems to affect Pol II distribution at convergent genes. Notably PolII distribution does not change on genes with gain of R loops that resulted to be really long. So it seems that not only gene length (50), but also gene convergence and gene density can impair PolII progression in absence of Top1. The consequence of this impaired process involves also the efficient production of a transcript. Transcripts at level of R loop loss sites seem to decrease particularly at last exon (Fig 24). Notably even RNA seq signal downstream of the RNA 3' UTR decreases after Top1 depletion. These data can be a possible additional evidence of an impairment of termination and polyadenilation. Additionally, it has to be considered that splicing of last exon and polyadenilation are strictly related (107) Our RNA-seq was performed on total RNA and not mRNA. A Total RNA sequencing is probably not the best approach to study effects on RNA molecules specifically at 3' end of genes. Thus, it will be crucial to understand if polyadenilation is in someway affected in loci showing loss of R loop peaks. However, these results have to be confirmed and validated by further and several analyses.

Another intriguing question arises if we consider the average gene distance between

colliding TTS of genes with loss of peak: 40 kb (Fig 23). Is it possible that torsional stress created by Top1 depletion in convergence region can act on such a long distance? It is hard to answer this question. Two studies have tried to characterize torsional stress linked to Top1 with psoralen photobinding. Both of them used campthotecin to inhibit Top1 (108) (119). Unfortunately these two studies focused their analysis on gene body without considering intergenic regions or gene direction. It will be useful to use these genomic data to see if Top1 inhibition increases positive supercoils in our R loop loss sites.

Finally, we need to consider that all DRIPc data has to be filtered for gene expression levels. This allows us to identify unexpressed genes and determine with higher precision gene direction. Moreover, we will be able to understand how Top1 modulate R loops according to transcription level.

With this work we provide new insights in the transcriptional role of Top1. We found new unexpected properties for its specific inhibitor CPT and we characterized the role Top1 in an unknown and poorly studied field such that one for R loops. Chedin's lab calculated that R loops are probably the most abundant class of non B DNA structure, probably covering the 5% of the genome. Most part of these structures is responsive to Top1 depletion. Therefore, it is unlikely that all of them simply contribute to genomic instability. It's clear that there are dangerous R loops and physiological R loops, R loops that are rapidly removed and R loops well tolerated by cells' control systems. Thus, a lot of work has to be done to correlate these structures to molecular interplayers, biological processes and their deregulation.

Convergent long genes

Top1

Top1

Top1

(-) R loops   (+) R loops   (-) R loops

Block of PolII progression
Impaired Polyadenilation ???

Divergent/Unidirectional   very long genes

Top1

Top1

(-) R loops   (-) R loops

Genomic instability ???
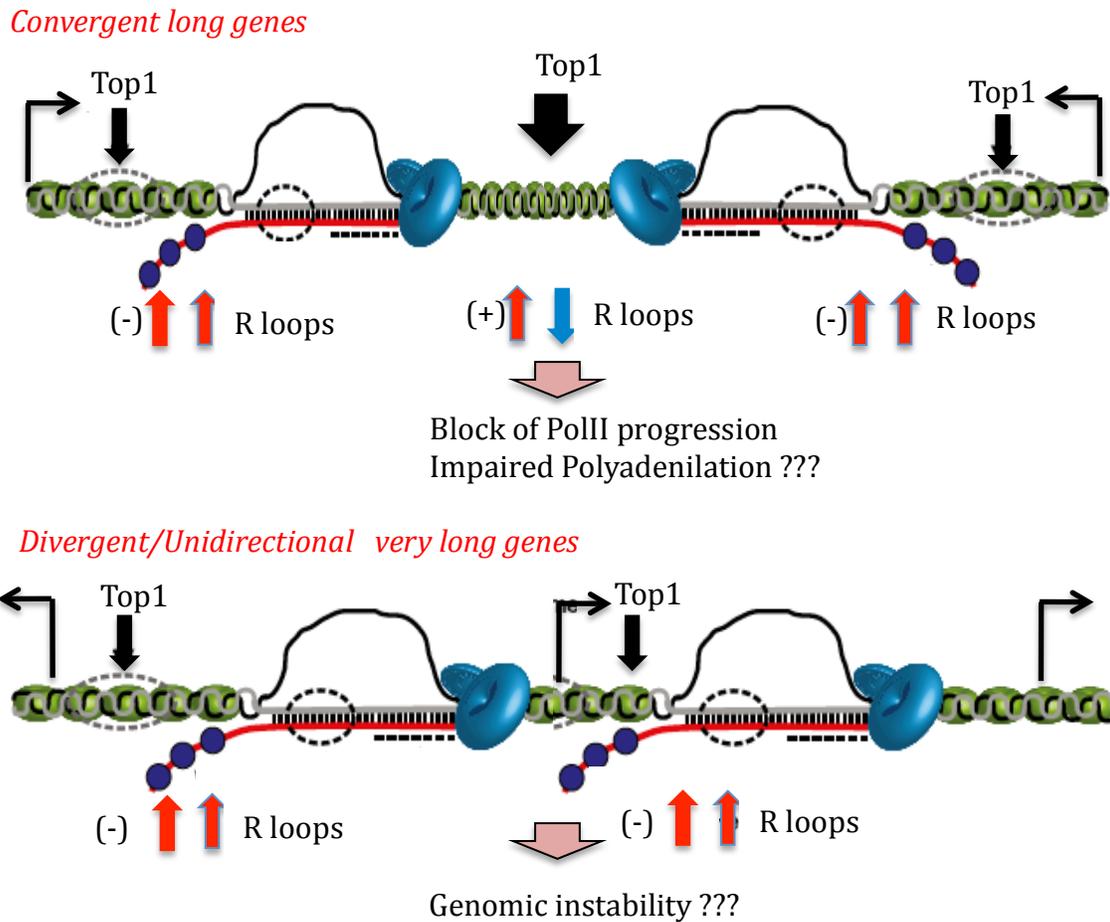
Fig 26. *Role of Top1 in regulating R loops and transcription* *Top1 different regulate R loops and transcription according to gene direction.**t*

In conclusion we found that TPL and CPT are exquisite tool to dissect the transcriptional role of their specific targets XPB and Top1. We disclosed new properties of the drugs and characterize the transcriptional effects of drug treatment, adding new insight on how cell system responds to these drugs. This work will also help to understand the physiological and the pathological role of important enzymes such as XPB and Top1.

Chapter V

BIBLIOGRAPHY

1) Viktorovskaya OV, Schneider DA. Gene. 2015 Feb 1;556(1):19-26. Functional divergence of eukaryotic RNA polymerases: Unique properties of RNA polymerase I suit its cellular role.

2) Svetlov V, Nudler E. Basic mechanism of transcription by RNA polymerase II. Biochim Biophys Acta. 2013 Jan;1829(1):20-8.

3) Arimbasseri AG, Rijal K, Maraia RJ. Comparative overview of RNA polymerase II and III transcription cycles, with focus on RNA polymerase III termination and reinitiation. Transcription. 2013 Dec 10;4(6).

4) Liu X, Bushnell DA, Kornberg RD. RNA polymerase II transcription: structure and mechanism. Biochim Biophys Acta. 2013 Jan;1829(1):2-8

5) Corden JL. Tails of RNA polymerase II. Trends Biochem Sci. 1.990 Oct;15(10):383-7.

6) Gnatt A.Elongation by RNA polymerase II: structure-function relationship. Biochim Biophys Acta. 2002 Sep 13;1577(2):175-90.

7) Flores O, Lu H, Killeen M, Greenblatt J, Burton ZF, Reinberg D. The small subunit of transcription factor IIF recruits RNA polymerase II into the preinitiation complex. Proc Natl Acad Sci U S A. 1991 Nov 15;88(22):9999-10003.

8) Luse DS. The RNA polymerase II preinitiation complex: Through what pathway is the complex assembled? Transcription. 2013 Nov 15;5

9) Lu H, Zawel L, Fisher L, Egly JM, Reinberg D. Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. Nature. 1992 Aug 20;358(6388):641-5

10) Tae-Kyung Kim, et al. Mechanism of ATP-Dependent Promoter Melting by Transcription Factor IIH.. Science 288, 1418 (2000)

11) Buratowski S. Progression through the RNA Polymerase II CTD Cycle Stephen.2009.10.019 Molecula Cell Review

12) Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Genes Dev. 14, 2452–2460.

13) Peng, J. et al. Identification of multiple cyclin subunits of human P-TEFb. Genes Dev. 12, 755–762.(1998)

14) J.B. Kim, P.A. Sharp. Positive transcription elongation factor B phosphorylates hSPT5 and RNA polymerase II carboxyl-terminal domain independently of cyclin-dependent kinase-activating kinase. J. Biol. Chem. 276 (2001) 12317–12323.

15) Zawel, L., Kumar, K.P., and Reinberg, D.. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes & Dev.* 9: 1479–1490. 1995

16) Rudd, M.D., Izban, M.G., and Luse, D.S.. The active site of RNA polymerase II participates in transcript cleavage within arrested ternary complexes. *Proc. Natl. Acad. Sci.* 91: 8057–8061. 1994

17) Jason N. Kuehner, Erika L. Pearson and Claire Moore Unravelling the means to an end: RNA polymerase II transcription termination . NATURE REVIEWS | MOLECULAR CELL BIOLOGY VOLUME 12 | MAY 2011 | 1

18) Steinmetz, E. J. & Brow, D. A. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol. Cell. Biol.* 16, 6993–7003 (1996).

19) Konstantina Skourti-Stathaki, Nicholas J. Proudfoot, and Natalia Gromak. Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Mol Cell. 2011 Jun 24;42(6):794-805

20) Eleanor White, Kinga Kamieniarz-Gdula, Michael J. Dye and Nick J. Proudfoot. AT-rich sequence elements promote nascent transcript cleavage leading to RNA polymerase II termination Nucleic Acids Research, 2013, Vol. 41, No. 3 1797–1806

21) Weinmann R, Roeder RG. Role of DNA-dependent RNA polymerase 3 in the transcription of the tRNA and 5S RNA genes. Proc Natl Acad Sci U S A. 1974 May;71(5):1790-4.

22) Van Trung Nguyen, Federico Giannoni, Marie-Françoise Dubois, Sook-Jae Seo,Marc Vigneron, Claude Kédinger and Olivier Bensaude In vivo degradation of RNA polymerase II largest subunit triggered by a-12924–2929 Nucleic Acids Research, 1996, Vol. 24, No. 15

23)Bensaude O. Inhibiting eukaryotic transcription Which compound to choose? How to evaluate its activity? Transcription 2:3, 103-108; May/June 2011;

24)Chao SC, Fujinaga K, Marion JE, Taube R, Sausville EA, Senderowicz AM, et al. Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. J Biol Chem 2000; 275:28345-8.

25)Zhu Y, Pe'ery T, Peng J, Ramanathan Y, Marshall N, Marshall T, et al. Transcription elongation factor P-TEFb is required for HIV-1 Tat transactivation in vitro. Genes Dev 1997; 11:2622-32.

26)Kupchan, S.M., Court, W.A., Dailey, R.G. Jr., Gilmore, C.J. & Bryan, R.F. TPL and tripdiolide, novel antileukemic diterpenoid triepoxide from Tripterygium wilfordii. *J. Am. Chem. Soc.* 94, 7194–7195 (1972).

27) Immunosuppressant PG490 (TPL) inhibits T-cell interleukin-2 expression at the level of purine-box/nuclear factor of activated T-cells and NF-kappaB transcriptional activation. J Biol Chem. 1999 May 7;274(19):13443-50.

28) Yang S, Chen J, Guo Z, Xu XM, Wang L, Pei XF, et al. TPL inhibits the growth and metastasis of solid tumors. Mol Cancer Ther 2003;2:65–72.

29) Kim MJ, Lee TH, Kim SH, Choi YJ, Heo J, Kim YH.TPL inactivates Akt and induces caspase-dependent death in cervical cancer cells via the mitochondrial pathway. Int J Oncol. 2010 Nov;37(5):1177-85

30)McCallum C, Kwon S, Leavitt P, Shoop W, Michael B, Felcetto T, Zaller D, O'Neill E, Frantz-Wattley B, Thompson C, Forrest G, Carballo-Jane E, Gurnett A. In vitro versus in vivo effects of TPL: the role of transcriptional inhibition. Therapy, Vol 2, N°2, March 2005 pp. 261-273(13).

31)Leuenroth SJ, Crews CM. TPL-induced transcriptional arrest is associated with changes in nuclear substructure. Cancer Res. 2008 Jul 1;68(13):5257-66.

32)Wang Y, Lu JJ, He L, Yu Q. TPL (TPL) inhibits global transcription by inducing proteasome-dependent degradation ofRNA polymerase II (Pol II). PLoS One 2011;6:e23993.

33)Titov DV, Gilman B, He QL, Bhat S, Low WK, Dang Y, et al. XPB, a subunit of TFIIH, is a target of the natural product TPL. Nat Chem Biol 2011;7:182–8.

34)Liu, L.F., Wang, J.C., *DNA–DNA gyrase complex: the wrapping of the DNA duplex outside the enzyme.* Cell., (1978). 15 (3): p. 979–984.

35) Fogg, J.M., Catanese Jr. D. J., Randall, G. L., Swick, M. C., Zechiedrich, L. , *Differences Between Positively and Negatively Supercoiled DNA that Topoisomerases May Distinguish*, in *Mathematics of DNA Structure, Function and Interactions*. (2009), Springer New York. p. 73-121.

36) Hiasa, H., Marians, K.J., *Two Distinct Modes of Strand Unlinking during theta-type DNA Replication.* J. Biol. Chem., (1996). 271(35): p. 21529–21535.

37) Drolet, M. Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNAtopology. Mol. Microbiol. 59, 723–730. (2006)

38) Schoeffler, A.J., Berger, J.M., *DNA topoisomerases: harnessing and constraining energy to govern chromosome topology.* Quart. Rev. Biophys., (2008) 41(1): p. 41–101.

39) Bouthier de la Tour, C., Portemer, C., Huber, R., Forterre, P., Duguet, M., *Reverse gyrase in thermophilic eubacteria.* Journal of Bacteriology (1991). 173(12): p. 3921–3923

40) Champoux, J.J., *DNA topoisomerases: structure, function, and mechanism.* Annu. Rev. Biochem., (2001). 70: p. 369–413

41) Nitiss, J.L., *Investigating the biological functions of DNA topoisomerases in eukaryotic cells.* Biochim. Biophys. Acta., (1998). 1400(1-3): p. 63–81.

42) Tsao YP[1], Wu HY, Liu LF Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies Cell. 1989 Jan 13;56(1):111-8.

43) Wang, J.C., *Cellular roles of DNA topoisomerases: A molecular perspective.* Nat. Rev. Mol. Cell. Biol., (2002). 3(6): p. 430–440.

44) Gangloff, S., McDonald, J. P., Bendixen, C., Arthur, L., Rothstein, R., *The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase.* Mol. Cell. Biol., (1994). 14(12): p. 8391-8

45) Muller, M.T., *Quantitation of eukaryotic topoisomerase I reactivity with DNA. Preferential cleavage of supercoiled DNA.* Biochim. Biophys. Acta., (1985). 824(3): p. 263-7

46) Stewart, L., Redinbo, M.R., Qiu, X., Hol, W.G., Champoux, J.J., *A model for the mechanism of human topoisomerase I.* Science, (1998). 279(5356): p. 1534-41.

47) Pommier, Y., *Topoisomerase I inhibitors: camptothecins and beyond.* Nat. Rev. Cancer, (2006). 6(10): p. 789-802.

48) Khobta, A., Ferri, F., Lotito, L., Montecucco, A., Rossi, R., Capranico, G. , *Early effects of topoisomerase I inhibition on RNA polymerase II along transcribed genes in human cells.* J Mol Biol., (2006) 357(1): p. 127-38.

49) Brill, S.J., Sternglanz, R., *Transcription-dependent DNA supercoiling in yeast DNA topoisomerase mutants.* Cell., (1988) 54(3): p. 403-11

50) Merino, A., Madden, K.R., Lane, W.S., Champoux, J.J., Reinberg, D., *DNA topoisomerase I is involved in both repression and activation of transcription.* Nature (1993) 365(6443): p. 227-32.

51) Miao, Z.H., Player, A., Shankavaram, U., Wang, Y. H., Zimonjic, D. B., Lorenzi, P. L., Liao, Z. Y., Liu, H., Shimura, T., Zhang, H. L., Meng, L. H., Zhang, Y. W., Kawasaki, E. S., Popescu, N. C., Aladjem, M. I., Goldstein, D. J., Weinstein, J. N., Pommier, Y., *Nonclassic functions of human topoisomerase I: genome-wide and pharmacologic analyses.* Cancer Res., (2007). 67(18): p. 8752-61.

52) King IF, Yandava CN, Mabb AM, Hsiao JS, Huang HS, Pearson BL, Calabrese JM, Starmer J, Parker JS, Magnuson T, Chamberlain SJ, Philpot BD, Zylka MJ. Topoisomerases facilitate transcription of long genes linked to autism.Nature. 2013 Sep 5;501(7465):58-62.

53) Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J., Hol, W.G., *Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA.* Science., (1998). 279(5356): p. 1504-13.

54) Hsiang, Y.H., Lihou, M.G., Liu, L.F., *Arrest of replication forks by drugstabilized topoisomerase I-DNA cleavable complexes as a mechanism of cell killing by camptothecin.* Cancer Res., (1989). 49(18): p. 5077–82

55) Zhang, H., Wang, J.C., Liu, L.F., *Involvement of DNA topoisomerase I in transcription of human ribosomal RNAgenes.* Proc. Natl. Acad. Sci. USA, (1988). 85(4): p. 1060–64.

56) Collins, I., Weber, A., Levens,D., *Transcriptional consequences of topoisomerase inhibition.* Mol. Cell. Biol., (2001). 21(24): p. 8437–8451.

57) Sordet, O., Larochelle, S., Nicolas, E., Stevens, E. V., Zhang, C., Shokat, K. M., Fisher, R. P., Pommier, Y., *Hyperphosphorylation of RNA polymerase II in response to topoisomerase I cleavage complexes and its association with transcription- and BRCA1-dependent degradation of topoisomerase I.* J Mol Biol, (2008). 381(3): p. 540-9.

58) Squires, S., Ryan, A. J., Strutt, H. L., Johnson, R. T., *Hypersensitivity of Cockayne's syndrome cells to camptothecin is associated with the generation of abnormally high levels of double strand breaks in nascent DNA.* Cancer Res., (1993). 53(9): p. 2012-9

59) Desai, S.D., Li, T. K., Rodriguez-Bauman, A., Rubin, E. H., Liu, L. F., *Ubiquitin/26S proteasomemediated degradation of topoisomerase I as a resistance mechanism to camptothecin in tumor cells.* Cancer Res., (2001). 61(15): p. 5926–5932.

60) Desai, S.D., Zhang, H., Rodriguez-Bauman, A., Yang, J.M., Wu, X., Gounder, M.K., Rubin, E.H., Liu, L.F., *Transcription-dependent degradation of topoisomerase I– DNA covalent complexes.* Mol. Cell. Biol., (2003). 23(7): p. 2341–2350.

61) Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. Rossi F, Labourier E, Forné T, Divita G, Derancourt J, Riou JF, Antoine E, Cathala G, Brunel C, Tazi J. Nature. 1996 May 2;381(6577):80-2.

62) Genome-wide analysis of novel splice variants induced by topoisomerase I poisoning shows preferential occurrence in genes encoding splicing factors. Solier S, Barb J, Zeeberg BR, Varma S, Ryan MC, Kohn KW, Weinstein JN, Munson PJ, Pommier Y. Cancer Res. 2010 Oct 15;70(20):8055-65.

63) DNA topoisomerase I inhibition by camptothecin induces escape of RNA polymerase II from promoter-proximal pause site, antisense transcription and histone acetylation at the human HIF-1alpha gene locus. Baranello L, Bertozzi D, Fogli MV, Pommier Y, Capranico G. Nucleic Acids Res. 2010 Jan;38(1):159-71.

64) Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. Huang HS, Allen JA, Mabb AM, King IF, Miriyala J, Taylor-Blake B, Sciaky N, Dutton JW Jr, Lee HM, Chen X, Jin J, Bridges AS, Zylka MJ, Roth BL, Philpot BD. Nature. 2011 Dec 21;481(7380):185-9. doi: 10.1038/nature10726.

65) Skourti-Stathaki K, Proudfoot NJ.A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. Genes Dev. 2014 Jul 1;28(13):1384-96.

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

67) Drolet M, Bi X, Liu LF. 1994. Hypernegative supercoiling of the DNA template during transcription elongation in vitro. J Biol Chem 269: 2068–2074.

68) Drolet M, Phoenix P, Menzel R, Masse E, Liu LF, Crouch RJ.1995. Overexpression of RNase H partially complements the growth defect of an

Escherichia coli DtopA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. Proc Natl Acad Sci 92: 3526–3530.

69) El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev. 24, 1546–1558.

70) Tuduri, S., Crabbe´, L., Conti, C., Tourrie`re, H., Holtgreve-Grez, H., Jauch, A.,Pantesco, V., De Vos, J., Thomas, A., Theillet, C., et al. (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat. Cell Biol. 11, 1315–1324.

71) Sordet, O., Redon, C.E., Guirouilh-Barbat, J., Smith, S., Solier, S., Douarre, C., Conti, C., Nakamura, A.J., Das, B.B., Nicolas, E., et al. (2009). Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. EMBO Rep. 10, 887–893.

72) Powell WT, Coulson RL, Gonzales ML, Crary FK, Wong SS, Adams S, Ach RA, Tsang P, Yamada NA, Yasui DH, Chédin F, LaSalle JM. R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proc Natl Acad Sci U S A. 2013 Aug 20;110(34)

73) Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. Mol Cell. 2014 Feb 6;53(3):484-97.

74) Aguilera A. The connection between transcription and genomic instability. EMBO J 21: 195–201. 2002.

75) Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102: 553–563. 2000.

76) Gottipati P, Cassel TN, Savolainen L, Helleday T.. Transcription- associated recombination is dependent on replication in Mammalian cells. Mol Cell Biol 28: 154–164. 2008

77) Chon H, Sparks JL, Rychlik M, Nowotny M, Burgers PM, Crouch RJ, Cerritelli SM.. RNase H2 roles in genome integrity revealed by unlinking its activities. Nucleic Acids Res 41: 3130–3143.2013

78) Cerritelli SM, Frolova EG, Feng C, Grinberg A, Love PE, Crouch RJ.. Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. Mol Cell 11: 807–815. 2003

79) Cerritelli SM, Crouch RJ. Ribonuclease H: the enzymes in eukaryotes. FEBS J 276: 1494–1505. 2009.

80) Alzu A, Bermejo R, Begnis M, Lucca C, Piccini D, Carotenuto W, Saponaro M, Brambati A, Cocito A, Foiani M, Liberi G. Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. Cell. 2012 Nov 9;151(4):835-46.

81) Huertas P, Aguilera A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell 12: 711–721.

82) Paulsen, R.D., Soni, D.V., Wollman, R., Hahn, A.T., Yee, M.C., Guan, A., Hesley, J.A., Miller, S.C., Cromwell, E.F., Solow-Cordero, D.E., et al. (2009) A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. Mol. Cell 35, 228–239.

83) Li X, Manley JL.Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell. 2005 Aug 12;122(3):365-78.

84) Sollier J, Stork CT, García-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability.. Mol Cell. 2014 Dec 18;56(6):777-85.

85) Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. Nat Immunol R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells.. 2003 May;4(5):442-51.

86) PA, Lott PL, Christensen HC, Korf I, Chédin F. Mol Cell. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Ginno 2012 Mar 30;45(6):814-25.

87) Ginno PA, Lim YW, Lott PL, Korf I, Chédin F. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. Genome Res. 2013 Oct

88) R-loops induce repressive chromatin marks over mammalian gene terminators. Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. Nature. 2014 Dec 18;

89) Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, Kamieniarz-Gdula K, Dimitrov S, Pathania S, McKinney KM, Eaton ML, Kellis M, Hill SJ, Parmigiani G, Proudfoot NJ, Livingston DM BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair.. Mol Cell. 2015 Feb 19;57(4):636-47.

90) Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754-60.

91) Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. 1000 Genome Project Data Processing Subgroup. Bioinformatics. 2009

92) Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841-2.

93) Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL.TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013 Apr 25;14(4):R36.

94) Love MI, Huber W, Anders S.Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;

95) Marinello J, Chillemi G, Bueno S, Manzo SG, Capranico G.Antisense transcripts enhanced by camptothecin at divergent CpG-island promoters associated with bursts of topoisomerase I-DNA cleavage complex and R-loop formation. Nucleic Acids Res. 2013 Dec;41(22):10110-23.

96) Manzo SG, Zhou ZL, Wang YQ, Marinello J, He JX, Li YC, Ding J, Capranico G, Miao ZH.Natural product TPL mediates cancer cell death by triggering CDK7-dependent degradation of RNA polymerase II. Cancer Res. 2012 Oct 15;72(20):5363-73.

97) Brookes E, Pombo A. Modifications of RNA polymerase II are pivotal inregulating gene expression states. EMBO Rep 2009;10:1213–9.

98) Seila,A.C., Calabrese,J.M., Levine,S.S., Yeo,G.W., Rahl,P.B., Flynn,R.A., Young,R.A. and Sharp,P.A. (2008) Divergent transcription from active promoters. Science, 322, 1849–1851.

99) Flynn,R.A., Almada,A.E., Zamudio,J.R. and Sharp,P.A. (2011) Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. Proc. Natl Acad. Sci. USA, 108, 10460–10465.

100) Albert E. Almada, XuebingWu, Andrea J.Kriz, Christopher B. Burge & Phillip A. Sharp. Promoter directionality is controlled by U1 snRNP and polyadenylation signals 3 6 0 | N AT U R E | VO L 4 9 9 | 1 8 J U LY 2 0 1 3

101) Rossi F, Labourier E, Forné T, Divita G, Derancourt J, Riou JF, Antoine E, Cathala G, Brunel C, Tazi J Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I.. Nature. 1996 May 2;381(6577):80-2.

102) Soret J, Gabut M, Dupon C, Kohlhagen G, Stévenin J, Pommier Y, Tazi J.Altered serine/arginine-rich protein phosphorylation and exonic enhancer-dependent splicing in Mammalian cells lacking topoisomerase I. Cancer Res. 2003 Dec 1;63(23):8203-11.

103) Helmrich A, Ballarino M, Tora L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Mol Cell 44: 966–977.

104) Nakama M, Kawakami K, Kajitani T, Urano T, Murakami Y. 2012. DNA–RNA hybrid formation mediates RNAi-directed heterochromatin formation. Genes Cells 17: 218–233

105) Castellano-Pozo M, Santos-Pereira JM, Rondon AG, Barroso S, Andujar E, Perez-Alegre M, Garcia-Muse T, Aguilera A. 2013. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. Mol Cell 52: 583–590.

106) Thomas S. Carroll, Ziwei Liang, Rafik Salama, Rory Stark and Ines de Santiago. Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. Front Genet. 2014; 5: 75

107) Dye MJ, Proudfoot NJ. 1999. Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. Mol Cell 3: 371–378.

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

109) Teves SS, Henikoff S.Nat Struct Mol Biol Transcription-generated torsional stress destabilizes nucleosomes. Nat Struct Mol Biol 2014 Jan;21(1):88-94.

Chapter VI

PUBLICATIONS and URL:

Natural product triptolide mediates cancer cell death by triggering CDK7-dependent degradation of RNA polymerase II.
Manzo SG, Zhou ZL, Wang YQ, Marinello J, He JX, Li YC, Ding J, Capranico G, Miao ZH.
Cancer Res. 2012 Oct 15

http://cancerres.aacrjournals.org/content/72/20/5363.long

Antisense transcripts enhanced by camptothecin at divergent CpG-island promoters associated with bursts of topoisomerase I-DNA cleavage complex and R-loop formation.
Marinello J, Chillemi G, Bueno S, **Manzo SG,** Capranico G.
Nucleic Acids Res. 2013 Dec

http://nar.oxfordjournals.org/content/41/22/10110.long

The natural inhibitor of DNA topoisomerase I, camptothecin, modulates HIF-1α activity by changing miR expression patterns in human cancer cells.
Bertozzi D, Marinello J, **Manzo SG**, Fornari F, Gramantieri L, Capranico G.
Mol Cancer Ther. 2014 Jan

http://mct.aacrjournals.org/content/13/1/239.long

Novel DNA topoisomerase IIα inhibitors from combined ligand- and structure-based virtual screening.
Drwal MN, Marinello J, **Manzo SG**, Wakelin LP, Capranico G, Griffith R.
PLoS One. 2014 Dec 9

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114904