DDX47 and MECP2, two novel human functions controlling R-loop-mediated genome integrity

Esther Marchena Cruz

Tesis doctoral

Universidad de Sevilla

2021

DDX47 and MECP2, two novel human functions controlling R-loop-mediated genome integrity

Trabajo realizado en el Departamento de Genética, Facultad de Biología y en el departamento de Biología del Genoma, CABIMER, de la Universidad de Sevilla, con el fin de optar al grado de Doctora en Biología Molecular, Biomedicina e Investigación Clínica por la graduada Esther Marchena Cruz.

Sevilla, 2021

La doctoranda: **Los directores de tesis:**

TABLE OF CONTENTS

INDEX OF FIGURES

INTRODUCTION

RESULTS

DISCUSSION

MATERIALS AND METHODS

INDEX OF TABLES

APPENDIX

ABBREVIATIONS

53BP1	p53 binding protein 1
A	Alanine
AID	Activation-Induced cytidine Deaminase
AIRD	AID inducible R loop detection
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine triphosphate
A.U.	Arbitrary Units
BER	Base excision repair
bp	base pairs
BLAST	Basic local alignment search tool
BRCA	Breast Cancer
C	Cytosine
CDC25	Cell Division Cycle 25
cDNA	complementary DNA
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation-sequencing
CIN	Chromosomal instability
CSR	Class switching recombination
CNVs	Copy Number Variants
DDR	DNA Damage Response
DDX	DEAD box
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide triphosphate
DOX	Doxycycline
DRIP	RNA-DNA immunoprecipitation
dsRNA	Double-stranded RNA
E	Glutamic acid
EM	Electron Microscopy

ETS	External transcribed spacers
EU	5-ethynyluridine
G4	G-quadruplexes
GCR	Gross chromosomal rearrangements
GO	Gene ontoloty
FA	Fanconi Anemia
FACT	Facilitates Chromatin Transcription
GFP	Green Fluorescence Protein
H3S10-P	Histone H3 Phosphorylation at serine 10
HBD	Hybrid-binding domain
HDAC	Histone Deacetylase
IF	Immunofluorescence
Ig	Immunoglobulin
IGS	Intergenic spacer region
ITS	Internal transcribed spacers
K	Lysine
Kb	Kilobases
KDa	Kilodalton
lncRNA	long non-coding RNA
LOH	Loss of Heterozygosity
MCM	Mini chromosome maintenance
MDC1	Mediator DNA damage checkpoint 1
MIN	Micro- and minisatellite instability
miRNA	microRNA
MMR	Mismatch repair
MRN	Mre11-Rad50-Nbs1
mRNA	messenger RNA
mRNP	messenger Ribonucleoprotein particle
NORs	Nucleolus organizer regions

nt	nucleotides
NTS	Non-transcribed strand
ORF	Open reading frame
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase Chain Reaction
PLA	Proximal ligation assay
Pol	Polymerase
pRNA	promoter-associated RNA
qPCR	quantitative Polymerase Chain Reaction
RBP	RNA-binding protein
RBD	RNA binding domain
rDNA	ribosomal DNA
RF	Replication fork
RFB	Replication fork barrier
rRNA	ribosomal RNA
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNase	Ribonuclease
RNH1	RNase H1
RPA	Replication protein A
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SEM	Standard Error of the Mean
SHM	Somatic hypermutation
S-Pro	Space promoter
siC	Non-targeting siRNA control
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
SR	Serine/arginine-rich protein
SSB	Single-Strand Break

ssDNA	single-stranded DNA
SSUP	Small Subunit Processome
TAM	Transcription-associated mutation
TAR	Transcription-associated recombination
TAGIN	Transcription-associated genome instability
TC-NER	Transcription couple nucleotide excision repair
TCR	Transcription couple Repair
Top	Topoisomerase
TREX	Transcription-Export complex
tRNA	transfer ribonucleic acid
TS	Transcribed Strand
TSS	Transcription start site
TSA	Trichostatin A
TTS	Transcription terminal site
UBF	Upstream binding factor
UTR	Untranslated region
UV	Ultraviolet
WT	Wild Type
XPF	Xeroderma Pigmentosum Complementation Group F
XPG	Xeroderma Pigmentosum Complementation Group G

RESUMEN

La integridad del genoma es un requisito indispensable para la correcta transmisión de la información genética de una célula a su descendencia. De hecho, evolutivamente, nuestras células han desarrollado numerosos procesos que actúan de forma coordinada para evitar o resolver amenazas que puedan potencialmente comprometer la estabilidad del genoma. Tanto agentes genotóxicos externos como aquellos derivados del propio metabolismo celular procedentes de procesos básicos como la replicación, transcripción y recombinación pueden ser causantes de inestabilidad genómica. Generalmente, la inestabilidad del genoma se manifiesta en forma de mutaciones y reordenamientos cromosómicos, siendo incluso una característica asociada a predisposición a cáncer y envejecimiento.

La transcripción es un proceso esencial para la proliferación y supervivencia celular que a su vez supone un riesgo para la integridad del genoma por diversos motivos. Como consecuencia de la transcripción tienen lugar cambios topológicos en el ADN que dan lugar a ADN de cadena sencilla, el cual es más susceptible a sufrir daños que el ADN de cadena doble. Además, la maquinaria de transcripción puede ser un obstáculo para otros procesos esenciales como la replicación, causando roturas en el ADN, recombinación y, en consecuencia, reordenaciones cromosómicas. Otra de las causas más estudiadas es la formación de bucles R (en inglés *R loops*), estructuras formadas por un híbrido de ARN-ADN y la cadena sencilla de ADN desplazada. La formación de estas estructuras tiene lugar durante la transcripción cuando el transcrito de ARN naciente invade hibridando con su hebra molde de ADN, desplazando así la hebra no transcrita como ADN de cadena sencilla. A pesar de que los *R loops* pueden desempeñar funciones en procesos fisiológicos actuando como estructuras intermediarias, se ha demostrado que pueden llegar a ser una amenaza para la expresión génica y la integridad del genoma.

En esta tesis nos centramos en la búsqueda de nuevos factores que puedan estar implicados en la inestabilidad genómica dependiente de la formación de *R loops*. Para ello, hemos llevado a cabo un escrutinio de alto rendimiento con microscopía de fluorescencia en células humanas con una

colección de ARN interferentes dirigidos contra posibles genes diana de fármacos. Hemos empleado una línea celular especial que contiene un *cassette* para la expresión controlada de la enzima citidina deaminasa (AID) como herramienta para la detección indirecta de híbridos ARN-ADN combinada con la medida de daño en el ADN. Mediante esta aproximación hemos identificado 46 genes candidatos relacionados con procesos como transcripción, biogénesis del ARN, ciclo celular, degradación de proteínas y traducción, entre otros. Hemos llevado a cabo una caracterización en mayor detalle de 7 de los candidatos obtenidos para estudiar el posible efecto de su silenciamiento en la acumulación de *R loops* y daño en el ADN. Según los resultados obtenidos, decidimos centrarnos en dos factores con funciones diferentes: MeCP2, una proteína de unión a ADN metilado, y DDX47, una helicasa de ARN localizada en el nucléolo.

El silenciamiento de MECP2 da lugar a un aumento de híbridos a nivel global en el nucleoplasma y en concreto, en genes con altos niveles de híbridos previamente descritos, así como en genes cuya expresión depende de MeCP2. El fenotipo de acumulación de bucles de ARN observado podría estar asociado al papel de MeCP2 en la regulación de la transcripción y en la condensación de la cromatina.

Hemos estudiado en profundidad el papel de DDX47 en la homeostasis de los *R loops* y el efecto que tiene esta helicasa en la estabilidad del ADN. DDX47 se localiza fundamentalmente en el nucléolo y se recluta a la cromatina tanto en el locus del ADN ribosómico como en genes transcritos por la ARN polimerasa II (RNAPII). Esta unión a la cromatina es consistente con la acumulación de híbridos tanto en el ADN ribosómico como en genes transcritos por RNAPII. El silenciamiento de DDX47 produce una reducción del área total nucleolar, afecta parcialmente al reclutamiento de la ARN polimerasa I en el rDNA, y reduce drásticamente la transcripción en el nucléolo, sugiriendo un posible papel de esta helicasa nucleolar no sólo en el procesamiento del ARN sino también en la transcripción. Así mismo el silenciamiento de DDX47 está asociado a un incremento de las colisiones replicación-transcripción, lo cual podría contribuir a la inestabilidad genética observada. En colaboración con el laboratorio del Dr. Xue Xiaoyu en la Universidad de Yale, hemos descubierto que DDX47 es una helicasa de ARN-ADN capaz de resolver híbridos y estructuras *R*

loops in vitro. En paralelo hemos demostrado que la sobreexpresión de DDX47 suprime fenotipos de acumulación de híbridos en células con niveles reducidos de factores previamente implicados en la prevención/resolución de estas estructuras. Nuestros resultados confirman que DDX47 es una helicasa conservada con capacidad para resolver híbridos ARN-DNA *in vivo*.

Esta tesis nos ha permitido identificar a MECP2 y DDX47 como nuevos factores implicados en la homeostasis de los *R loops* y necesarios para prevenir la inestabilidad genómica dependiente de híbridos de ARN-DNA.

INTRODUCTION

1. GENOME INSTABILITY

The complete instructions required for life are encompassed in the genome and stored as DNA (or deoxyuribonucleic acid) molecule in the cells. To maintain genome integrity and propagate faithfully this information during cell divisions to the off-spring, cells have developed numerous processes that act in a closely coordinated manner. Genome changes may occur due to environmental genotoxic agents and endogenous metabolites action, leading to genetic variability, which although can be harmful for our cells and organism, are crucial for evolution (Aguilera & Gómez-González, 2008). Indeed, genome instability plays a key role in the mechanism of diversification of immunoglobulin genes, providing a large battery of molecules that recognize the antigens. Nevertheless, high levels of genome instability are deleterious for the cell and arise upon genotoxic stress or can be the result of certain pathologies that affect a proper DNA repair and/or replication. Importantly, genome instability is considered one of the hallmarks of cancer, aging, and other genetic diseases and disorders (Aguilera & García-Muse, 2013; Gaillard & Aguilera, 2016; Niedernhofer et al., 2018).

DNA integrity is compromised by the action of ultraviolet (UV) radiation, ionizing radiations, numerous chemicals and endogenous metabolites and as consequence of biological processes such as transcription and replication (Hoeijmakers, 2009). Depending on the mechanisms involved, DNA can undergo different type of lesions such as abasic sites, bases mismatch, DNA adducts, inter- and intra-strand crosslinks, single-stranded DNA (ssDNA) gaps and double strand breaks (DSBs). Consequently, these lesions could be translated into a wide range of genetic alteration: point mutations, micro- and minisatellite instability (MIN) expansion and contractions, chromosomal instability (CIN), gross chromosomal rearrangements (GCRs), copy number variants (CNVs) of a particular DNA fragment, loss of heterozygosity (LOH) and hyper-recombination (Aguilera & García-Muse, 2013).

To minimize detrimental consequences of DNA injury, the stability of genome is supported by multiple pathways focused on detection of DNA lesions, their signalling and subsequently DNA repair to counteract DNA damage and ensure cell proliferation or apoptosis. This set of mechanisms form a complex

signal transduction pathway known as DNA Damage Response (DDR). When such system fails increase cancer susceptibility (Gaillard & Aguilera, 2016; Jeggo et al., 2016).

1.1. DNA damage response

The DNA Damage response (DDR) is a complex signal transduction pathway in which a signal (DNA damage) is detected by a protein sensor that triggers the activation of a transduction system that amplifies and diversifies the signal to protein effectors for an efficient DNA repair (Ciccia & Elledge, 2010).

Depending on the specific type of DNA damage, DDR signalling pathway can differ in a wide variety of DNA repair mechanisms. DDR signalling pathways are initiated by the recognition of DNA lesions by the complex MRN (Mre 11 Rad50 Nsb1), that recognizes DSBs and RPA (replication protein A). Then, the upstream kinases, Mec1/ATR (the ataxia telangiectasia and RAD3-related) and Tel1/ATM (ataxia telangiectase mutated), are activated and phosphorylate different substrates. As a consequence of ATR/ATM phosphorylation, a quality-control mechanism that blocks cell cycle to guarantee genome integrity, known as checkpoint, is activated preventing progression from G1 to S phase and from G2 to M phase (Sulli et al., 2012). DDR signalling is spread around the DSB by the phosphorylation of the histone variant γH2AX on Ser139. This mark is required to recruit MDC1, (mediator DNA damage checkpoint 1), that maintains DDR signalling by enforcing accumulation of the MRN complex and ATM activation at DSBs sites (van Attikum & Gasser, 2009). Other factors as BRAC1 and 53BP1 (p53-binding protein 1) are also recruited to DSB sites. Downstream kinases CHK2 and CHK1, mainly phosphorylated by ATM and ATR respectively, are activated and the signalling pathway converges on downstream effectors to coordinate different outcomes: transient cell cycle arrest to repair DNA damage and cell death by apoptosis or cellular senescence (Jackson & Bartek, 2009; Sulli et al., 2012). Otherwise, a long-term state of global genome instability is a characteristic of almost all human cancers (Niedernhofer et al., 2018).

1.2. Replication

Propagation of the genetic information to the off-spring requires timely and accurate duplication of DNA prior to cell division. The replication fork (RF) must travel along the DNA to accomplish its function. During replication the RF has to

pass through the chromatinized DNA template and must overcome frequent obstacles such as DNA damage, protein barriers, torsional stress, heterochromatin, non-B DNA structures or transcription machinery itself, that will hamper replication fork progression (Gómez-González & Aguilera, 2019). Indeed, these obstacles are an important source of replication stress and genome instability since encountering an obstacle can cause RF stalling, which turns into an irreversible collapse of the synthesis, causing replisome disassembly and generating hazardous DSBs and ssDNA (Aguilera & García-Muse, 2013; Aguilera & Gómez-González, 2008). In general, any condition that compromises replication it is referred to as replication stress. The surveillance mechanisms that prevent genome instability upon replication stress are mitotic and S phase checkpoint pathways to guarantee replication completion and to prevent replication fork breakage (Gaillard et al., 2015).

1.3. Transcription-associated genome instability

Transcription is an essential cellular process in all living cells for proliferation and survival, in which the genetic information encompassed in form of DNA is copied into new RNA molecules that can either play their function directly or be translated into proteins. During transcription, the two complementary DNA strands are separated to allow the RNA polymerase (RNAP) to copy one of the strands as a template into RNA. Specifically, in the transcription bubble, the transcribed strand (template) generate a complementary RNA chain, forming an RNA-DNA hybrid, whereas the non-transcribed strand remains unpaired as ssDNA (Gaillard et al., 2013).

During RNAP journey across a gene, the organization and integrity of the genome are transitory affected by the positive and negative DNA supercoiling and chromatin remodelling changes, to allow the progression of the RNAP through the DNA template (Selth et al., 2010). Due to the fact that DNA molecule is the substrate of transcription, this process creates the optimal conditions for high levels of mutations (transcription associated mutation, TAM), recombination (transcription associated recombination, TAR) as well as DSBs, chromosome rearrangements and chromosome loss by making chromatin more susceptible to DNA insulting agents. Both TAM and TAR are conserved processes from prokaryotes to higher eukaryotes (Aguilera, 2002).

In the 70s by two different laboratories, transcription-associated mutagenesis (TAM) was first reported in *Escherichia coli*. Both studies showed that transcription activation in presence of mutagenic agents increases reversion rates of point mutations in transcribed genes (Brock, 1971; Herman & Dworkin, 1971). Later works suggested that high levels of transcription impact differently mutagenesis of the non-transcribed strand (NTS) versus the transcribed strand (TS). The analysis of mutation of both DNA strands (NTS and TS) showed that mutations along transcribed regions occur preferentially in NTS (Beletskii & Bhagwat, 1996, 1998). TAM has been evidenced in a wide variety of assays and systems to sense mutations from prokaryotes to eukaryotes (Jinks-Robertson & Bhagwat, 2014). One of the most frequent mutations are C-to-T transitions, caused by spontaneous cytosine deamination to uracil conversion. If uracil is not removed, it is paired with adenine, resulting into C-to-T mutations after replication (Jinks-Robertson & Bhagwat, 2014). In this thesis we have used the overexpression of AID an human cytosine deaminase to increase genome instability as we will explain bellow.

In the case of TAR, numerous works based on systems to sense ectopic recombination have shown the transcription can lead to hyper-recombination and chromosomal rearrangements (Gaillard et al., 2013). Studies in bacteria, yeast and mammals indicate that as a consequence of transcription, TAR is enhanced by making DNA more susceptible to undergo recombinogenic DNA damages and more accessible to the enzymatic activities required for recombination directed by site-specific recombinases. Moreover, TAR is necessarily linked to replication. Several studies showed that when RF (replication fork) progression was opposed to transcription, TAR occurred at high levels (Gaillard & Aguilera, 2016; Hamperl & Cimprich, 2016).

Genome instability associated with transcription can be due to several factors. During transcription, the chromatin is subjected to topological changes to allow the movement of the transcription machinery, which favoured negative and positive supercoiling accumulated behind and ahead of the elongating RNAP, respectively. Such changes can facilitate the transient formation of ssDNA that is more vulnerable to damage (Figure I1A). According that, mutants of topoisomerases accumulate torsional stress at the transcribed genes and favour

transient underwound DNA that might lead to stretches of ssDNA behind RNAP resulting in transcription-associated genome instability in different ways. These stretches could be damaged due to their susceptibility to spontaneous base modifications, genotoxic agents or enzymatic activities that could cause ssDNA breaks and base damage.

This particular scenario, when DNA unwinding occurs during replication, transcriptional or recombination, also encourages non-B DNA structures formation in different forms as hairpins, G-quadruplexes (G4) DNA or RNA-DNA hybrids (known as R loops when formed outside the transcription bubble) (Figure I1B). It is worth to notice that genome instability associated with these structures is due to their relationship with replication, by blocking replisome (Aguilera & García-Muse, 2013), or impeding transcription by the accumulation of stalled RNAPs (Belotserkovskii et al., 2013) which in turn can serve as a signal for the machinery of repair called transcription-coupled excision repair (TC-NER) in the absence of DNA lesions (Hanawalt & Spivak, 2008), undergoing genome instability at risk in all cases (Gaillard & Aguilera, 2016).

1.4. Transcription-replication conflicts

Within the numerous obstacles to replication fork progression, transcription is likely the main source of replicative impairments since the complexity of transcription machinery, together multiple mRNA processing steps and mRNP assembly factors are a challenge to advancing replication fork. Although there is a temporal or spatial separation between transcription and replication (for most of the genes) both processes use the same DNA as template, thus unavoidable encounters on the same DNA region at the same time can be occur, causing transcription-replication conflicts (Figure I1C) (Gómez-González & Aguilera, 2019; Hamperl & Cimprich, 2016; Oestergaard & Lisby, 2017).

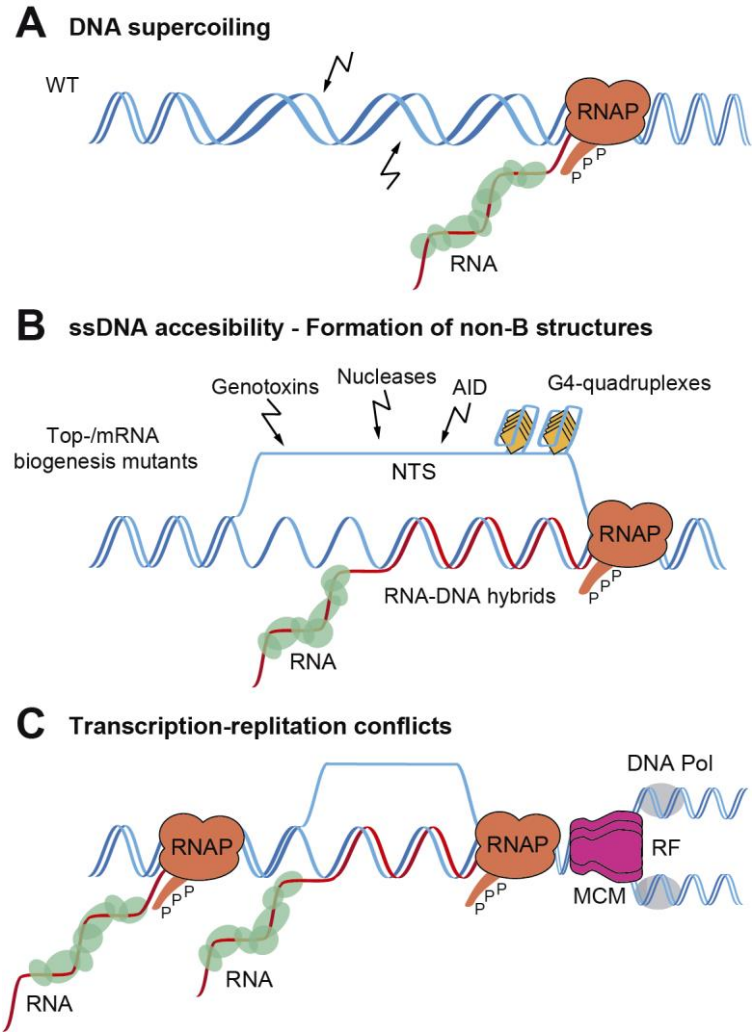


Figure I1. Transcription-associated genome instability. (A) The progression of transcription promotes the formation of local negative supercoiling behind the RNAP and positive supercoiling ahead the RNAP that make the DNA more vulnerable to damage. **(B)** In the absence of certain factors, non-B structures such as G4-quadruplexes, RNA-DNA hybrids or R loops are produced. The nascent RNA transcript hybridizes with its DNA template strand and displaces NTS remaining as ssDNA, that is more susceptible to the action of genotoxic agents, including reactive oxygen species (ROS), nucleases and other modifying enzymes such as AID. **(C)** Encounters between replication and transcription machineries cause transcription-replication collisions. Depending on the orientation to the replication origin can occur in a head-on orientation (if both machineries progress in the opposite direction) which are more deleterious than codirectional orientation (when both machineries advance in the same direction). MCM, minichromosome maintenance complex; DNA pol, DNA polymerase; RF, Replication Fork; RNAP, RNA polymerase; NTS, non-transcribed strand. Figure adapted from (Gaillard et al., 2013).

DNA replication and transcription exhibit different rates favouring these conflicts to occur (Bermejo et al., 2012; Helmrich et al., 2013). Depending on the orientation to the replication origin of a particular relative gene, collision between transcription and replication machineries can occur in a head-on orientation (genes transcribed from the lagging strand template) or a codirectional orientation (genes transcribed from the leading strand template). Experimental data suggest that head-on collisions seem to be more deleterious and constitute a stronger

hindrance for replisome progression than codirectional both *in vitro* and *in vivo* (Liu & Alberts, 1995; Prado & Aguilera, 2005). These encounters can compromise genome integrity not necessarily by physical collisions between both machineries, but also enhance the occurrence of structures that hamper fork progression by modifying topology, the structure of the DNA template, as well as in the surrounding chromatin organization (Gómez-González & Aguilera, 2019).

The detrimental consequences of collisions are manifested in forms of TAM or TAR, or TAGIN (transcription-associated genome instability) as a general phenomenon, that may be enhanced by the accumulation of torsional stress, compromising the integrity of RF (Herrera-Moyano et al., 2014; Tuduri et al., 2009). However, cells use different strategies to avoid and resolve potentially deleterious transcription-replication conflicts. Indeed, most of the genes and virtually all rDNA loci of bacterial genomes are found in codirectional orientation with replication, preventing head-on collisions. The resolution of these such controversial structures demand of the action of specific factors to allow RF progression and maintenance of genome integrity (Hamperl & Cimprich, 2016). A mechanism of restraining replication at highly transcribed regions is another way to avoid collisions that can be achieved by the presence of an RF barrier (RFB), being so important for the stability of the locus (Ide et al., 2010; Kobayashi et al., 1998). The sophisticated organization of the rDNA region is the major challenge to replication since high frequency of collisions can be occurred. Thus, in yeast, the rDNA intergenic spacer contains an origin of replication (rARS) and a DNA element acting as an RFB downstream of the 35S genes, as will be discussed below (Takeuchi et al., 2003). Delaying origin firing and slowing down of RF progression in conditions requiring extensive transcription is another strategy to avoid collision (Duch et al., 2013). Opposite strategies consist on transcription repression prior to the passage of the replication machinery (V. C. Nguyen et al., 2010). Indeed, one strategy to avoid them is to restrain the amount of arrested RNAPs, for this transcription elongation factors are required (Tehranchi et al., 2010; Trautinger et al., 2005).

2. R LOOPS AS A SOURCE OF GENOME INSTABILITY

R loop is a non-B three-strand nucleic acid structure formed by an RNA-DNA hybrid plus a displaced DNA strand (ssDNA) (see Figure I2). The most accepted mechanism for R loop formation is known as “thread-back” model, where the nascent RNA transcript invades hybridizing with its DNA template strand and leading the complementary strand as ssDNA (Aguilera & García-Muse, 2012).

RNA-DNA hybrids occur naturally since they are intermediary structures formed during some physiological cellular processes such as replication and transcription. These hybrids are normally found at active site of the RNA polymerase (RNAP) at the transcription bubble or during lagging-strand DNA synthesis (Aguilera & García-Muse, 2012).

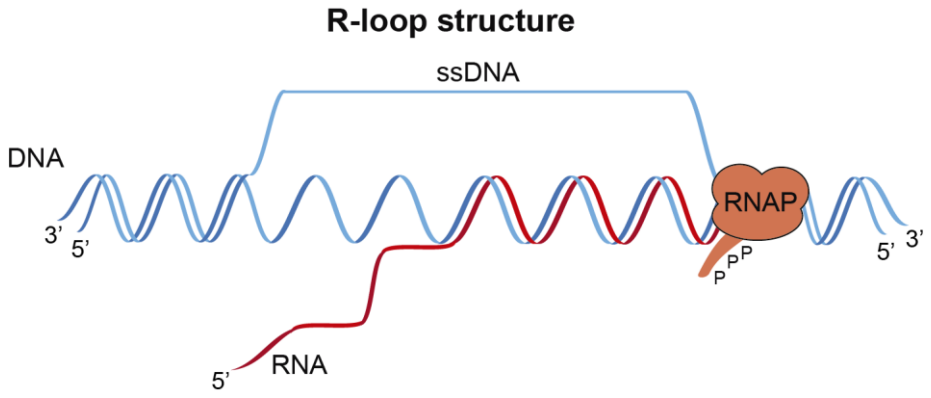


Figure I2. Schematic representation of an R loop structure. R loops is a non-B three-strand nucleic acid structure formed by an RNA-DNA hybrid plus a displaced DNA strand (ssDNA).

Genome-wide distribution data of RNA-DNA hybrid prone loci analysis indicate that these structures accumulate preferentially at highly transcribed genes, including the rDNA and tRNA loci, together with also transposable elements, centromeres and telomeres, antisense-RNAs or ncRNA regions (Chan et al., 2014; Liang Chen et al., 2017; El Hage et al., 2014; Ginno et al., 2012; Wahba et al., 2016). In *S. cerevisiae*, R loops are also detected at a subset of open reading frames with a high GC content and generally highly transcribed (Chan et al., 2014; El Hage et al., 2014; Wahba et al., 2016). In mammalian cells, active genes at the promoter and terminator are R loop-prone regions (Ginno et al., 2012; Sanz et al., 2016). Others R loop-prone regions identified are GC-rich sequences, CpG islands, transcription termination sites with high GC skew, or G-quadruplex (G4)-containing sequences (Liang Chen et al., 2017; Ginno et al., 2012; Sanz et al., 2016; Stork et al., 2016).

2.1. Physiological roles of R loops

Several reports have described transient programmed R loops as key intermediates in different specific physiological processes such as in *E. coli* plasmid replication (Kogoma, 1997), mitochondrial DNA replication (Xu & Clayton, 1996), immunoglobulin (Ig) class switching recombination (CSR) in B-cells (Yu et al., 2003) or CRISPR-Cas9 gene editing (Figure I3) (Jiang et al., 2016).

Cumulative evidences indicate that R loops play a role in gene expression, consistently with the enrichment of these structures at promoter and termination regions (Figure I3). Genome-wide analysis have revealed that R loops are preferentially accumulated at the G-rich 5'-UTR regions immediately downstream of the CpG-non-methylated promoters in humans (Liang Chen et al., 2017; Dumelie & Jaffrey, 2017; Ginno et al., 2012; Nadel et al., 2015; Sanz et al., 2016). On one hand, R loops may facilitate transcription by protecting DNA from methylation and influencing the binding and activity of chromatin remodelers at promoters (Ginno et al., 2012; Grunseich et al., 2018). On the other hand, R loops can affect the binding of transcription factors, and in some cases this regulation is mediated by noncoding RNAs as shown in yeast and human cells (Arab et al., 2019; Boque-Sastre et al., 2015; Cloutier et al., 2016). Thus, R loops may regulate the activation/inactivation of genes depending on different mechanisms (Figure I3) (Crossley et al., 2019).

Different studies suggest that R loops are enriched at the 3' end of some RNAPII-driven genes and it has been proposed that they could pause RNAPII and mediate efficient transcription termination. Some reports support that helicases as SETX and DHX9 could resolve these R loops and favour the action of the nuclease XRN2 leading to transcription termination (Cristini et al., 2018; Morales et al., 2016; Skourti-Stathaki et al., 2011). R loops at some terminators regions can induce antisense transcription that could also contribute to transcription termination via the recruitment of silencing machinery (Figure I3) (Skourti-Stathaki et al., 2014; Castel et al 2014). It is also worth mentioning that R loops could hamper transcription as an initial pause signal to slow-down RNAPII that would facilitate the co-transcriptional splicing process or the correct

termination and polyadenylation of the nascent mRNA (Proudfoot, 2016; Wahba et al., 2016).

Although R loops may have different physiological roles, a number of reports suggest a problematic scenario where R loops are formed at higher frequency as aberrant by-products of transcription interfering with DNA replication, repair and transcription and as a consequence becoming a threat to gene expression and genome integrity (García-Muse & Aguilera, 2019).

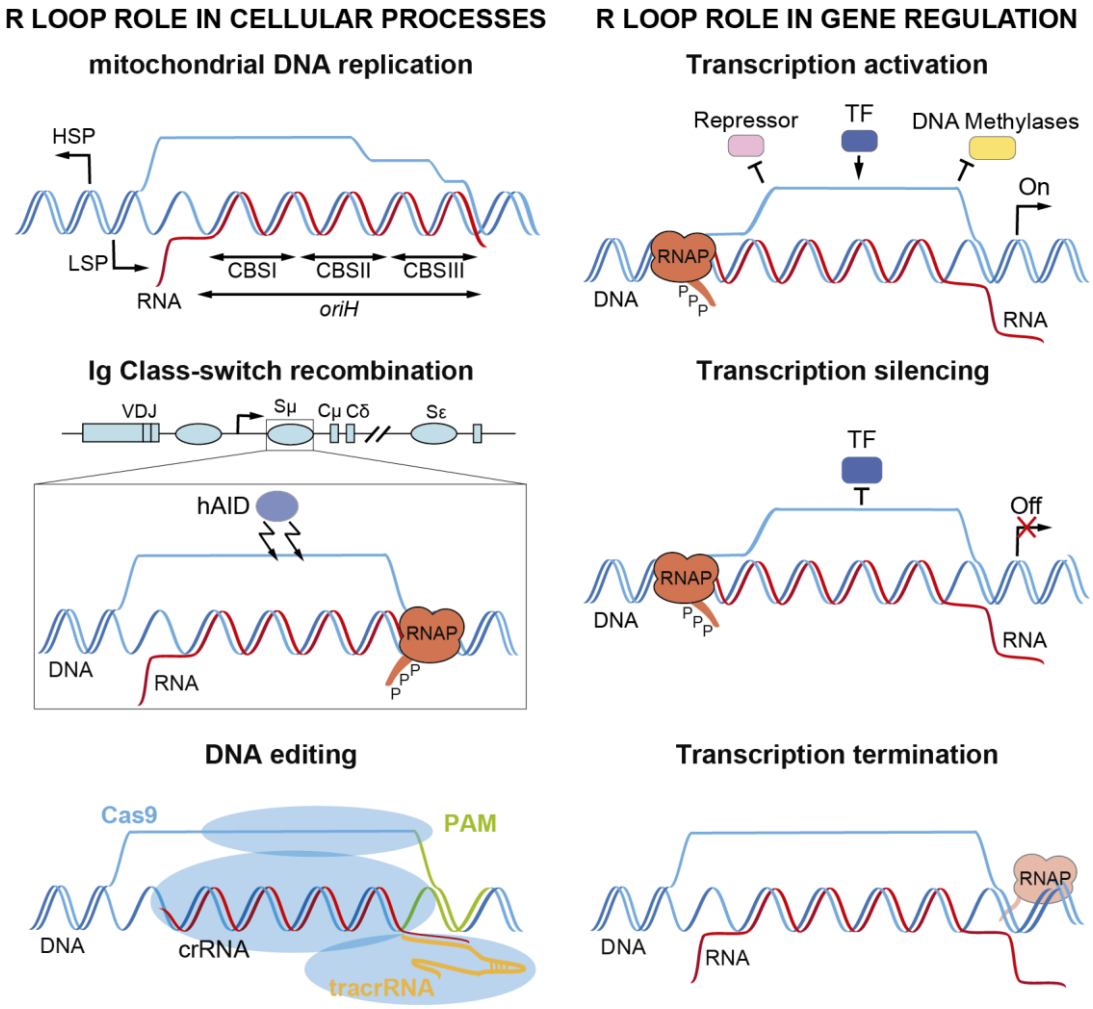


Figure I3. Physiological role of R loops. R loops role in cellular processes, including mitochondrial light DNA strand replication, Ig class-switch recombination or identification of a target DNA site by a guide RNA in the CRISP-Cas9 system. R-loop role in gene regulation. At promoters, R loops can activate transcription by promoting transcription-factor or by blocking binding of transcriptional repressors or DNA methylases or silence transcription by impeding transcription-factor binding. At terminators, R loops can facilitate transcription termination. HSP, H-strand promoter; LSP, L-strand promoter; TF, transcription factor; RNAP, RNA polymerase; AID, activation-induced cytidine deaminase; PAM, protospacer adjacent motif. Figure adapted from (García-Muse & Aguilera, 2019).

2.2. R loops as genomic threats

2.2.1. R loop-mediated genome instability

R loops are potentially linked to different forms of genome instability including mutations, recombination, and chromosome rearrangements as well as chromosome loss (Aguilera & García-Muse, 2012). Given their structure, the displaced ssDNA within the R loop makes it more vulnerable to DNA genotoxic agents and the mutagenic action of specific DNA-modifying enzymes such as activation-inducted cytidine deaminase (AID) that acts on ssDNA (see below section 4.1) and nucleases, as a consequence ssDNA nicks could be formed, that if remaining unrepaired might lead to RF collapse and DSBs formation (Figure I4A). It is noteworthy that the recognized role of R loop in triggering DSBs has recently emerged in a new layer of complexity about whether R loops are detrimental by-products that must be removed to preserve genome instability or whether these structures actively participate in DSB repair (Gómez-González & Aguilera, 2020; Marnef & Legube, 2021). Other observations evidence that hypermutation could be mediated by the RNA strand within the R loops since it can act as a highly mutagenic primer of unscheduled error-prone DNA synthesis, as has been described in *E. coli* (Kogoma, 1997) or in the rDNA region of yeast deprived of Top1 and ribonuclease H (RNase H) enzymes (Stuckey et al., 2015).

Transcription can be also altered by R loops. Mutants affected in different steps of transcription, concurrent with excessive stalling, pausing, arrest or backtracking of RNAP, and mRNA processing show a transcription-associated hyperrecombination phenotype and elevated chromosome and plasmid loss (Aguilera & Gómez-González, 2008; Huertas & Aguilera, 2003; Saponaro et al., 2014). R loops structure itself, when the nascent RNA hybridizes with the DNA template, could impair transcription elongation destabilizing the transcription complex and leading to transcription blockages. Indeed, transcription blockages caused by RNAP collision with preformed R-loops have been shown *in vitro* (Tous & Aguilera, 2007) but the effect of this structures *in vivo* needs to be further explored (Kouzine et al., 2017). It remains to clarify different scenarios, whether problems with transcription derive from a RNAP stalled with the R loop or whether the R loop by itself act as a barrier to upstream polymerases.

2.2.2. R loops and replication stress

R-loop-mediated genome instability is also related to the increase of recombination, chromosome rearrangements and chromosome instability as a result of DNA breaks. However, the way by which an R loop can lead to a recombinogenic DNA break may be diverse and are not fully understood yet. The most accepted mechanism is based on replication due to the potential R loops capacity to impede RF progression and lead to fork stalling, potential collapse and as a final consequence breakage. Indeed, the most prevalent source of R loop-dependent DNA damage seems to take place during S phase (Gan et al., 2011; Madireddy et al., 2016; Wellinger et al., 2006). Many examples of RF progression impairment in R-loop-accumulating cells from bacteria to human have been described supporting this idea by different techniques such as 2D gel electrophoresis, DNA combing or the recruitment of Rrm3 DNA helicase (Gómez-González et al., 2011; Herrera-Moyano et al., 2014; Salas-Armenteros et al., 2017; Tuduri et al., 2009; Wellinger et al., 2006). In addition, in these cases, replication impairments, DNA breaks and recombination are entirely or partially supressed by overexpression of ribonuclease H1 (RNase H1), which specially cleave the RNA moiety of RNA-DNA hybrids, suggesting that part of this phenotype could be mediated by potentially harmful RNA-DNA hybrids (Aguilera & García-Muse, 2012).

There are several possible explanations about how DNA breaks can occur as a consequence of R loops. As mentioned before, the single-strand nature of the displaced DNA strand within the R loop increase the susceptibility to be damaged, leading to single-strand breaks (SSBs). If these DNA lesions remain unrepaired, RF progression can be stalled which in turn can lead to DSBs formation (Figure I4A). Furthermore, the ssDNA also favours the formation of non-B DNA structures such as hairpin, G-quadruplexes, which may become an obstacle to replication. On the other hand, the RNA-DNA hybrid structure can act as a barrier for replication by stalling or blocking RF progression directly. Indirectly, transcription-replication conflicts appear when the presence of a stable R loop force the RNAP to remain attached at the transcription site (Figure I4B) (Aguilera & García-Muse, 2012; Gaillard & Aguilera, 2016). It is noteworthy that recent studies report that the replisome acts as an orientation-dependent

regulator of R loops levels, whereas in the co-directional orientation reduce R-loops levels, in head-on orientation promote their formation (Hamperl et al., 2017; Lang et al., 2017). However, it is an unsolved question whether an R loop is sufficient to provoke RF stalling or if a more complex chromatin structure is involved (Castellano-Pozo et al., 2013; García-Pichardo et al., 2017; Rondón & Aguilera, 2019).

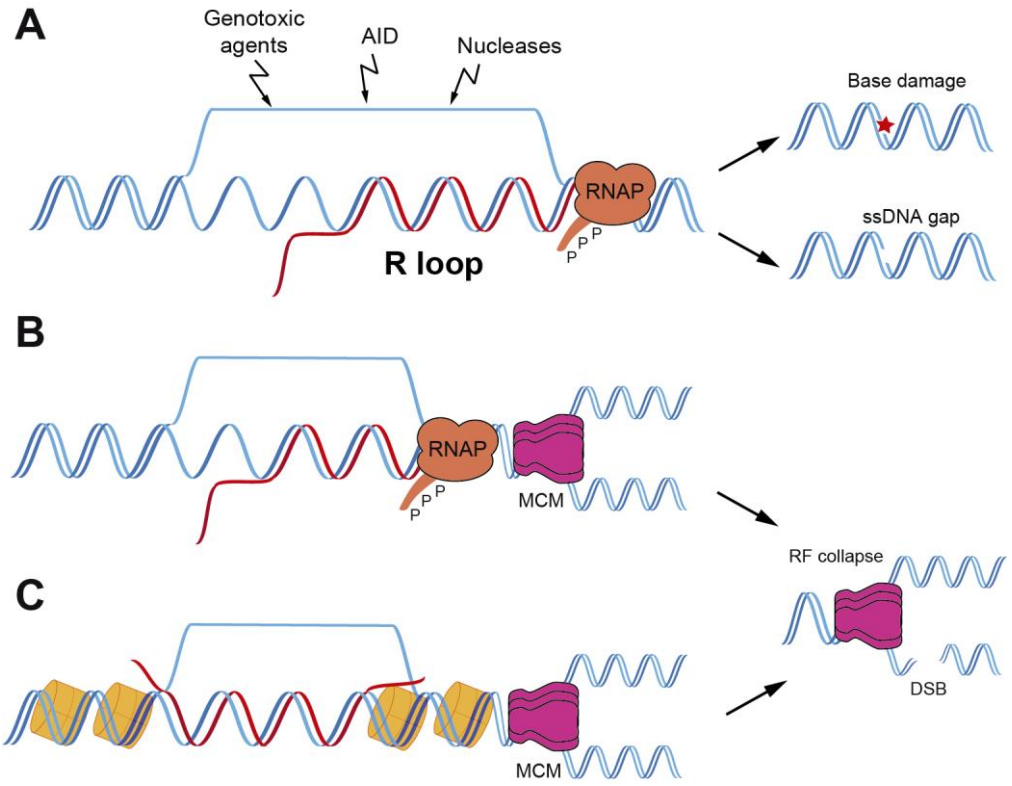


Figure I4. R loop-mediated genome instability. (A) As a consequence of R loop formation, the non-transcribed strand that remains unpaired as displaced ssDNA molecule can be more susceptible to the attack of different genotoxic agents or enzymatic activities such as AID. This could lead to DNA lesions such as base damage (red star) or ssDNA gaps, which can block the progression of replication forks, generating genome instability that can ultimately lead to cell death or cancer. **(B)** Directly, R loops by themselves could act as an obstacle of replication forks or indirectly when a stable R loop force the RNAP to remain attached at the transcription site promoting transcription-replication conflicts. **(C)** Chromatin compaction as a consequence of R loop formation could also interfere with replication. In all this cases, R loop could hamper RF progression, leading to RF stalling, potential collapse and breakage generating genome instability. Adapted from (Gaillard & Aguilera, 2016).

2.2.3. R loops and chromatin

A nascent connection between chromatin changes and R loops homeostasis have recently emerged as possible cause of R-loop associated genome instability (Chédin, 2016). Histone hyper-acetylation has been proposed to facilitate R loop accumulation due to the fact that the high chromatin accessibility facilitates the probability of R loop generation. In fact, it has been shown that depletion of the

histone deacetylase mSin3a complex or the acetyltranferase MOF or de-acetylation inhibition by chemical compounds (TSA and SAGA) causes R loop accumulation and genome instability (Salas-Armenteros et al., 2017; Singh et al., 2018; Wahba et al., 2011).

In contrast, aberrant R loop accumulation have been associated with a mark of chromatin condensation (H3S10-P), in yeast, *C. elegans* and human cells (Castellano-Pozo et al., 2013; García-Pichardo et al., 2017). In particular, depletion of THO complex or the helicase SETX/Sen1 leads to an accumulation of H3S10-P in yeast and human cells. These observations open the possibility that local chromatin compaction triggered by R loops contributes to replication fork impairment or to delayed or less-efficient replication initiation as possible causes of R-loop dependent genome instability (Figure I4C) (Castellano-Pozo et al., 2013). Moreover, mutants of the chromatin-reorganizing complex FACT (facilitates chromatin transcription) show high levels of R loops and R-loop mediated transcription-replication conflicts in yeast and human cells. This phenotype is likely linked with a specific chromatin organization since this complex is involved in nucleosome disassembly around the RNA polymerase during transcription elongation (Herrera-Moyano et al., 2014). To understand the biological meaning of R-loop chromatin signatures, recent screening studies in yeast have been performed where it has been identified specific histone mutants that accumulate R loops but do not cause genome instability. Thus, R loops do not cause DNA damage by themselves, but requiring a two-step mechanism in which, first, an altered chromatin prone to R loops, and second, a modified chromatin that including H3S10-P for compromising genome integrity (García-Pichardo et al., 2017). Histone H3 acetylation, Histone 3 lysine 4 (H3K4) mono- and di-methylation and tri-methylation of H3K36 are also significantly enriched over R-loops prone sites, linking R-loop-forming regions with open chromatin and high RNA polymerase occupancy (P. B. Chen et al., 2015; Sanz et al., 2016).

Another piece of evidence between R loops and chromatin is a recent work that demonstrates a key role of the conserved chromatin remodeler complex SWI/SNF on the maintenance of genome instability by helping resolve R-loop-mediated T-R conflicts (Bayona-Feliu et al., 2021).

3. MECHANISMS AND FACTORS INVOLVED IN PREVENTING R LOOP ACCUMULATION

Cells have developed mechanisms to control R loop homeostasis and mitigate their harmful effect if deregulated (Chédin, 2016; Santos-Pereira & Aguilera, 2015; Skourti-Stathaki & Proudfoot, 2014). In general, two types of mechanisms or factors have been reported: those that help prevent R loop formation (mRNP biogenesis and topoisomerases) and those which remove R loops (ribonucleases, helicases and others).

3.1. Topoisomerases

Transcription induces local topological changes, whereas positive supercoiling (overwinding) ahead, negative supercoiling (underwinding) is located behind the advancing RNAP. The advance of replication machinery also leads to transient negative supercoiling accumulation. This more open DNA can favour the hybridization of the nascent RNA with the DNA template and the formation of R loop structures. Supporting this idea topoisomerases, enzymes required to resolve or relax negative and positive supercoils respectively, also prevent R loop accumulation (Gaillard & Aguilera, 2016; Santos-Pereira & Aguilera, 2015). Indeed, yeast topoisomerases mutants show R loop accumulation at rDNA locus (El Hage et al., 2010) and human TOP1-deficient cells accumulate stalled replication forks and chromosome breaks in an R-loop dependent manner (Tuduri et al., 2009). Recently, genome-wide analysis indicate that toposiomerase I prevents transcription-replication conflicts at R-loop-enriched transcription termination sites (TTS) (Promonet et al., 2020). Therefore, topoisomerases are considered as crucial factors involved in preventing R-loop formation during transcription (Figure I5A).

3.2. mRNP biogenesis

During eukaryotic gene expression, nascent mRNA need to be processed and correctly packaged to export a properly assembled and integrated messenger ribonucleoparticle (mRNP) to the cytoplasm. The different steps of mRNP biogenesis are linked from transcription, mRNA processing to mRNP export and can influence each other (Bentley, 2014; Proudfoot et al., 2002). mRNP biogenesis is not only important for gene expression, but also for the maintenance of genome integrity. Thus, a wide variety of RNA-binding proteins (RBPs) are

associated with the nascent mRNA, promoting completely its package and protection and also reducing the capacity to hybridize back with the DNA template (Figure I5A) (Rondón et al., 2010). The connection between mRNA biogenesis and R loops was first shown in yeast mutant of THO/TREX, a conserved complex involved in the coupling of transcription with mRNP biogenesis and export (Huertas & Aguilera, 2003). Cells lacking a functional THO complex accumulate RNA-DNA hybrids and show hyper-recombination and DNA breaks that are R loop dependent (Castellano-Pozo et al., 2012; Chávez & Aguilera, 1997; Domínguez-Sánchez et al., 2011; Gómez-González et al., 2011; Huertas & Aguilera, 2003). Importantly, genome instability of THO mutants is associated with an increase in transcription-replication conflicts (Gómez-González et al., 2011; Salas-Armenteros et al., 2017; Wellinger et al., 2006).

Furthermore, the connection between transcripcion/mRNP biogenesis/export and genome instability is extended to other factors involved in splicing, SRF1 (serine/arginine splicing factor 1) (X. Li & Manley, 2005); in mRNA 3´ end processing and degradation (Figure I5A) (Gavaldá et al., 2013; Luna et al., 2005; Pefanis et al., 2015; Stirling et al., 2012); mRNP biogenesis and export such as THSC/TREX2 among others (González-Aguilera et al., 2008). Nowadays, several global and specific studies have evidenced more factors involved in RNA metabolism that have an important association with R loops accumulation (Chan et al., 2014; Paulsen et al., 2009; Stirling et al., 2012; Wahba et al., 2011), thus highlighting that the structure and fate of the nascent mRNA are crucial for R-loop prevention. Nevertheless, there are too few studies about molecular mechanisms underlying R loop homeostasis, and not all factors protect the genome in an R-loop-dependent manner. Thus for example, certain splicing factors could contribute to genome integrity indirectly by regulating gene expression (Salas-Armenteros et al., 2019).

3.3. Ribonucleases

Apart from the active prevention of R loops, cells also possess diverse mechanisms actively involved in removing these structures once formed. First of all, the RNase H enzymes specifically degrade the RNA moiety of RNA-DNA hybrids (Figure I5B). There are two types of RNase H enzymes: RNase H1, which is monomeric, removes RNA-DNA hybrids like primers in Okazaki fragments and

RNase H2, which is monomeric in bacteria but is composed of three subunits in eukaryotes, is involved in ribonucleotide excision repair (Cerritelli & Crouch, 2019). Despite, their different structure and substrate specificity, both type are able to resolve RNA-DNA hybrids (Cerritelli & Crouch, 2009; Skourti-Stathaki & Proudfoot, 2014). RNase H1 is the main player in removing co-transcriptional R loops (Chon et al., 2013). Indeed, RNase H1 overexpression is widely used to remove R loops and to supress R-loop dependent genome instability phenotypes.

3.4. Helicases

R loops can be also removed by RNA-DNA helicases through unwinding these structures (Figure I5B). Helicases are essential enzymes that use ATP-driven motor force to unwind DNA or RNA duplex substrates and are involved in different aspects of nucleic acid metabolism: DNA replication, repair, transcription and recombination (Singleton et al., 2007; Valentini & Linder, 2021). The yeast protein Sen1 and its human homologue senataxin (SETX) are extensively studied examples of proteins implicated in R-loop homeostasis (Figure I5B). In *S. cerevisiae,* Sen1 helicase-inactive *sen1-1* mutants show R loops accumulation and hyper-recombination phenotype (Mischo et al., 2011). In human cell, SETX depletion causes R-loop accumulation at transcription termination pause sites, suggesting that this protein is required to resolve R loops, in particular during transcription termination process (Skourti-Stathaki et al., 2011). Other well-described examples of helicases related to R loops homeostasis are DHX9 (Cristini et al., 2018), DDX1 (L. Li et al., 2016), DDX23 (Sreerama Chaitanya Sridhara et al., 2017), DDX21 (Song et al., 2017), DDX19 (Hodroj et al., 2017), UAP56 (Pérez-Calero et al., 2020), DDX5 (Sessa et al., 2021) among others. However, in most of the cases, it is unclear whether these RNA helicases have a direct role in unwinding RNA-DNA hybrids or whether they have the ability to reduce R-loop accumulation based on their activity as RNA chaperones or RNPs disassemblers.

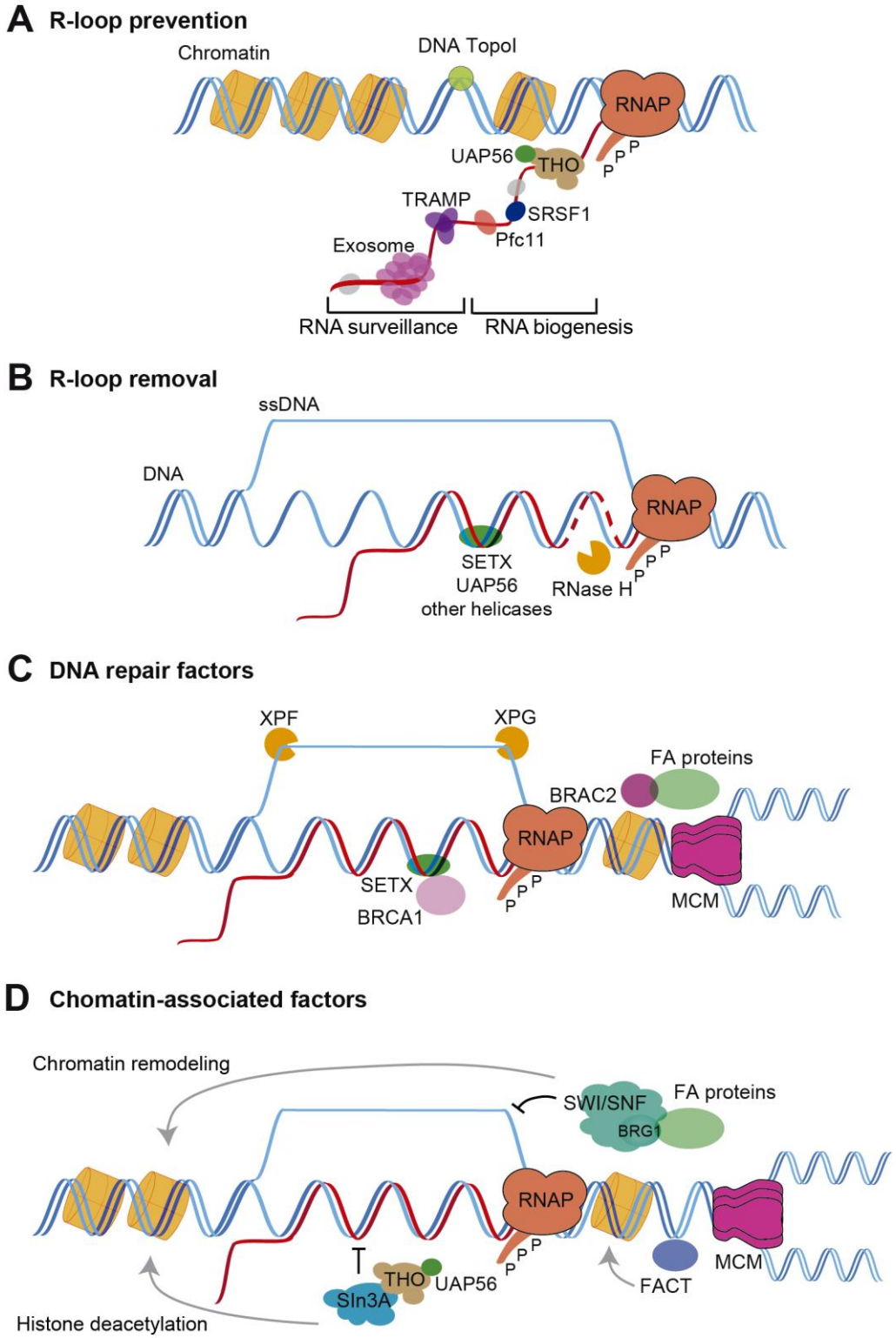


Figure I5. Mechanisms and factors to prevent and resolve R loop accumulation. (A) Topoisomerase 1 (TOP1) avoid the local negative supercoiling accumulation behind the elongating RNAP. Specific RNA binding proteins involved in different steps of RNA metabolism from RNA biogenesis (THO complex, UAP56, SRSF1 and Pcf11) to RNA surveillance (exosome, TRAMP) prevent the formation of R loops. **(B)** RNase H enzymes can remove R loops, degrading RNA moiety. Moreover, helicases as SETX and UAP56 could unwind the RNA-DNA hybrids within the R loop. **(C)** DNA repair factors as BRAC1, BRAC2, TC-NER factors (XPF, XPG) and potentially other Fanconi Anemia (FA) proteins can help counteract this situation. **(D)** Chromatin-associated factors such as the FACT, the histone deacetylase mSin3a and SWI/SNF complexes have a role in prevention of R loop-mediated genome instability.

3.5. Other factors

In addition to direct players in R loop homeostasis, there are many factors that resolve other forms of stresses that may indirectly contribute to resolve R loops. Importantly, most of them have a role in DNA repair and/or in mediation of transcription-replication conflicts. For instance, this is the case of the transcription-coupled nucleotide excision repair (TC-NER) factors that may contribute to mitigate R-loop-induced DNA damage and instability. Specifically, it is proposed that TC-NER nucleases XPG and XPF could excise R loops, being actively processed into DSBs (Figure I5C) (Sollier et al., 2014; Yasuhara et al., 2018). The DSB repair factors and tumor suppressors BRAC1 and BRAC2 have been shown to play a role regulating transcription elongation by RNA Polymerase II to prevent R-loop accumulation and help to suppress R-loop-mediated transcriptional stress (Figure I5C) (Bhatia et al., 2014; Shivji et al., 2018; X. Zhang et al., 2017). Furthermore, it is proposed that BRAC1 recruits SETX to remove R loops at termination sites (Hatchi et al., 2015). Cells lacking Fanconi Anemia factors show an R loop accumulation and genome instability phenotype (M. L. García-Rubio et al., 2015; Hatchi et al., 2015; Schwab et al., 2015).

Chromatin related factors are also relevant in R loop homeostasis. Thus the histone deacetylase Sin3A histone deacetylase complex interacts with THO/UAP56 and suppresses co-transcriptional R-loops (Luna et al., 2019; Pérez-Calero et al., 2020; Salas-Armenteros et al., 2017). Different works have shown that other chromatin-associated factors as FACT, and SWI/SNF complexes play a role in the resolution of R-loop-mediated transcription-replication (Figure I5D) (Bayona-Feliu et al., 2021; Herrera-Moyano et al., 2014).

4. TOOLS AND TECHNIQUES TO DETECT R-LOOP ACCUMULATION

A main issue of R loop biology is the development of strategies to detect these structures *in vivo* (Figure I6A). Electron microscopy (EM) and physical analysis of nucleic acid resistant to RNase A and sensitive to RNase H are direct methods used for R loop detection (Drolet et al., 1995; Duquette et al., 2004; Huertas & Aguilera, 2003). Moreover, the most widely direct tool used is the S9.6 monoclonal antibody that detects RNA-DNA hybrids, which originally was isolated by Carrico and colleagues in 1985. Specifically, Mouse (BALB/c) B-cell was fused with Mouse (BALB/c) Sp2/0-Ag14 myeloma to produce monoclonal antibodies against RNA-DNA hybrids. Animals were immunized with RNA-DNA hybrids formed by using single stranded PhiX174 DNA as template for *E. coli* DNA dependent RNA polymerase (Boguslawski et al., 1986). RNA-DNA hybrids immunoprecipitation from digested DNA (DRIP) and immunofluorescence (IF) using the monoclonal antibody are methods commonly used to capture RNA-DNA hybrids (M. García-Rubio et al., 2018). However, recent reports have revealed that S9.6 antibody has ability to bind also to dsRNA (Hartono et al., 2018; Silva et al., 2018). As a consequence, indirect strategies have been developed to overcome S9.6 limitations and detect the existence of R loop accumulation including the use of bisulfite which converts specifically cytosines into uracils of the ssDNA displaced by the RNA-DNA hybrid (Ginno et al., 2012; Yu et al., 2003), the use of the hybrid-binding domain of the RNase H enzyme fused either to GFP (HBD-GFP) (Bhatia et al., 2014) or to MNase (Yan et al., 2019) and the use of an catalytically inactive RNase H in chromatin immunoprecipitation (ChIP) which is bound to the RNA-DNA hybrids but without resolving them (Liang Chen et al., 2017). It worth to notice that all these techniques are applied for genome-wide detection of R loops (Liang Chen et al., 2017; Dumelie & Jaffrey, 2017; Ginno et al., 2012; Sanz et al., 2016; Sanz & Chédin, 2019).

Complementarily, the suppression of R loop-dependent phenotypes by *in vivo* RNase H overexpression or *in vitro* RNase H and RNase III treatment are essential tools to decipher if the signal measured comes from RNA-DNA hybrids or not.

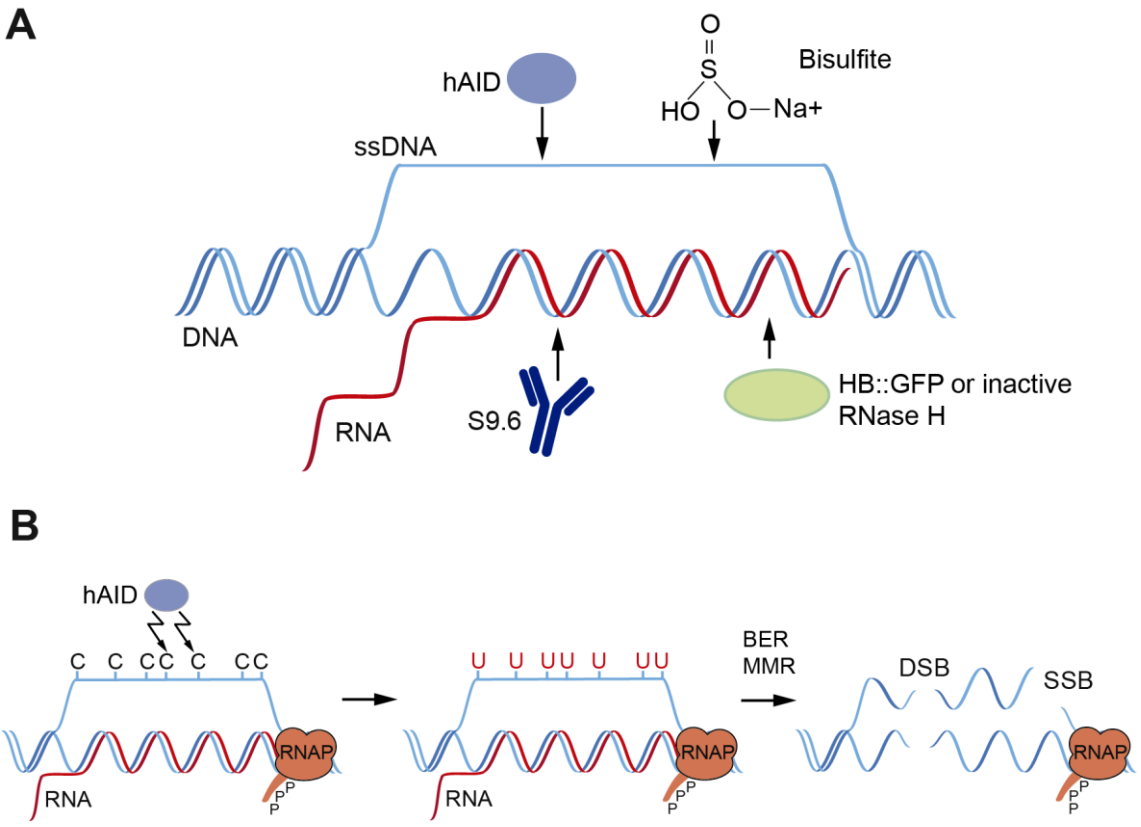


Figure I6. Direct and indirect tools for R-loop recognition. (A) RNA-DNA hybrids can be detected with the S9.6 antibody, GFP fused to an inactive RNaseH or the RNA binding domains (HB) of the RNase H. Induced-mutagenesis with bisulfite *in vitro* or with activation-induced cytidine deaminase (AID) *in vivo* are used to detect the displaced ssDNA within the R loop. **(B)** Schematic illustration of the AID-based principle for detecting R-loop dependent DNA damage. AID deaminate cytosines of ssDNA converting them into uracil, whose repair processing by base excision repair (BER) and mismatch repair (MMR) leads to DSBs. Figure adapted from (García-Muse & Aguilera, 2019).

4.1. AID as an indirect tool for R loop-detection

There is another tool to indirectly detect R loops based on the expression of the human cytosine deaminase AID (activation-induced cytidine deaminase) (Figure I6B). This enzyme acts on ssDNA and deaminate cytosines of ssDNA turning them into uracils. AID is required to initiate the processes of somatic hypermutation (SHM) and class switch recombination (CSR) in the immunoglobulin genes required for antibody diversification (Chaudhuri & Alt, 2004). The R loop formation behind RNP provides the substrate to the AID enzyme (Yu et al., 2005). Nevertheless, beyond its physiological activity, AID can erroneously act on a subset of off-target genes in a transcription-dependent manner, particularly in R loop prone accumulating genes, promoting chromosome translocations and mutations prone to activate oncogenes (Hendriks et al., 2010; Ruiz et al., 2011).

Heterologous expression of human AID was first used in an study in yeast with mutants of the THO complex (Gómez-González & Aguilera, 2007). A strong and transcription-dependent hypermutation overall in the non-transcribed strand and hyperrecombination was induced by AID in these mutants validating the accumulation of R loop structures in the absence of these mRNP factors. Later on, the ectopic expression of AID has been reported in yeast and human cells as a tool to increase R loop-dependent genome instability (García-Benítez et al., 2017; García-Pichardo et al., 2017; Gómez-González & Aguilera, 2009; Mischo et al., 2011). These approaches are based on the ability to AID to act on the ssDNA of an R-loop structure deaminating Cs turning them into uracils that finally lead to DSB by the action of other enzymes and as a consequence, causing AID-mediated genomic instability (Figure I6B) (Aguilera & Gómez-González, 2008; Basu et al., 2011; Gómez-González & Aguilera, 2007; Ruiz et al., 2011).

Taken advantage of the previous experience in our laboratory of the use of AID to exacerbate R-loop genome instability associated phenotypes, in this thesis, we have used a system based on AID expression to identify new factors involved in R loop-homeostasis in human cells.

5. rDNA LOCUS

Genome-wide sequencing of hybrid-harboring loci have revealed a higher R loop distribution in retrotransposons, telomeres and highly expressed genes, such as the ribosomal RNA and tRNA loci. Specifically, in *Saccharomyces cerevisiae* it has been described that 50% of all mapped R loop are in the rDNA locus in wild-type cells (Wahba et al., 2016). Similar to yeast, R-loops accumulate at repetitive sequences such as transposable elements, ribosomal DNA, centromeres and telomeres in human cells (Ginno et al., 2012).

In this section we will summarize some characteristics of the transcription, processing and transcription associated-genome instability of the human rDNA locus.

rDNA genes are arranged in clusters of multiple tandem repeats (150-200 in yeast and up to 350 in humans) and symbolize the most highly transcribed genome locus. The chromosomal landmarks that contain rRNA repeats are

termed nucleolus organizer regions (NORs). In humans, each repeat unit is ~43 kb in length, with a ~13 kb rRNA coding region and a ~31 kb IGS (intergenic sequence). Human rDNA tandem arrays are distributed on the short arms of the five acrocentric chromosomes in human (chromosomes 13, 14, 15, 21, and 22) (Figure I7A). The RNA polymerase I (RNAPI) drive rRNA synthesis, generating a single large polycistronic pre-ribosomal RNA (pre-rRNA) transcript (47S in mammalian species). This transcript undergoes an intricate sequence of processing steps that lead to the mature 18S rRNA component of the 40S subunit and to the 5.8S and 25S/28S rRNA components of the 60 subunit (Fatica & Tollervey, 2002). The major part of this process occurs in the nucleolus, but additional maturation events take place in the nucleoplasm. During rRNA processing the external transcribed spacers (5'ETS and 3'ETS) and internal transcribed spacers (ITS) are removed (Figure I7A). Whereas 18S synthesis pathway requires four successive endonuclease cleavages, 5.8S and 25S/28S synthesis is more complex and requires endonuclease cleavage and multistep exonuclease digestions. In addition to the nucleases, numerous ribosomal and non-ribosomal proteins and ribonucleoproteins (RNPs) containing large varieties of small nucleolar RNAs (snoRNAs) are involved in rRNA processing (Kiss, 2001). rRNA modifications are carried out by pseudouridine synthases and methyltransferases. RNA helicases and RNA chaperons mediate RNA folding/remodelling as DDX47, which is studied in more detail in this thesis. GTPases and AAA ATPases make possible protein associations and dissociations during rRNA processing (Raška et al., 2004; Tschochner & Hurt, 2003). Finally, downstream pathways of pre-rRNA processing and maturation occur in the nucleoplasm and cytoplasm (Gamalinda & Woolford, 2015).

Maintaining a balanced level of rDNA transcription is essential for cellular homeostasis, for this, in mammalian cells, heterochromatin-specific histone modifications and CpG methylation of the 45S rDNA promoter (Pro) are marked in silent rDNA repeats (Figure I7B) (Santoro et al., 2002). Moreover, Transcribed by RNAPI, S-Pro gives rise to the promoter-associated RNA (pRNA) as lncRNAs that mediates rDNA silencing linked to the 45S rRNA gene promoter (Pro) (Figure I7B) (Mayer et al., 2008). As in yeast rDNA, the human IGS of a rDNA unit also contains an rDNA origin of replication (rARS) (Figure I7B) (Vydzhak et al., 2020).

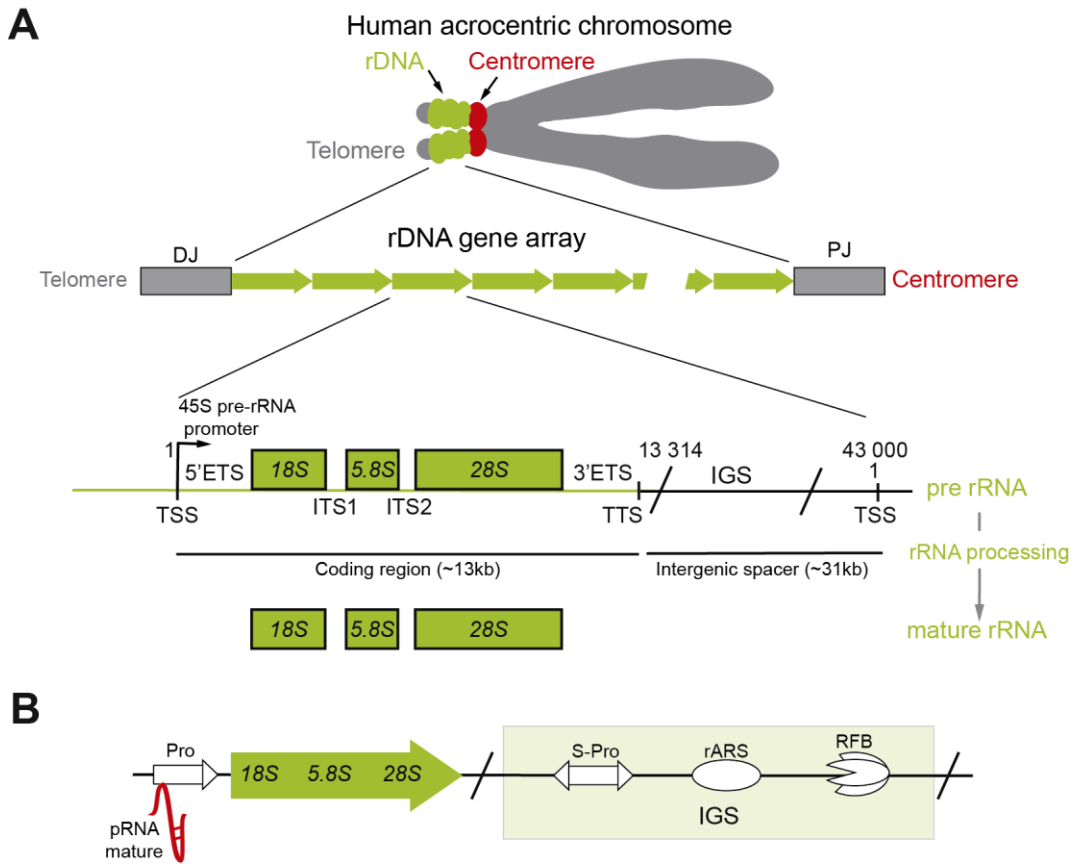


Figure I7. Schematic representation of motifs at the human rDNA loci. (A) In the human cells, the 45S cluster locates on the short arms of acrocentric chromosomes (13,14, 15, 21, 22) between telomeres and centromeres, surrounded by proximal and distal junctions (PJ and DJ). Each 45S rDNA repeat unit is transcribed and processed into 18S, 5.8S and 28S. The coding parts of 45S sequence are separated by internal transcribed spacers (ITS1 and ITS2) and external transcribed spacers (5'ETS and 3'ETS) are part of boundaries of the coding region. The coding regions are separated by the long intergenic spacer region (IGS). **(B)** Transcription termination elements are located both upstream and downstream of the transcription unit, including the 45 pre-rRNA promoter (Pro), the space promoter (S-pro), an origin of replication (rARS) and the replication fork barrier (RFB). pRNA, promoter-associated RNAs. Figure adapted from (Potapova & Gerton, 2019; Vydzhak et al., 2020).

To meet the high cellular requirements, ribosomal DNA (rDNA) locus becomes a paradigm of genome instability-prone region because of two main aspects very remarkable: it contains hundreds of tandem repeats of the DNA coding for ribosomal RNA (rRNA) and also the rDNA region is the most highly transcribed genomic locus (Gómez-González & Aguilera, 2019; Salim & Gerton, 2019). Indeed, the yeast rDNA loci are likely the most unstable and studied regions for T-R conflicts in the genome (Kobayashi, 2014). In *S. cerevisiae,* the RFB, when bound by Fob1 protein, avoids head-on collisions between replisome and RNAPI at 35S gene by stalling replication forks that progress in a direction opposite to RNAPI transcription (Kobayashi, 2003). This mechanism seems to be

conserved in humans, since RNAPI transcription terminator complex can act as a RFB at the rDNA accommodating the progress of replication with rRNA transcription activity (Figure I7B) (Akamatsu & Kobayashi, 2015).

OBJECTIVES

The main goal of this thesis is to further explore the wide variety of factors that could contribute to induce R loop-dependent genome instability in human cells. For this purpose, we addressed the following specific objectives:

1. To carry out a high-throughput screening to identify new factors involved in R loop-dependent genome instability in human cells.
2. To validate and characterize specific hits: MECP2 and DDX47 as novel R-loop related factors.
3. To analyze the role of DDX47 in the maintenance of genome integrity, with the aim of getting new insights into the molecular mechanism of this helicase to prevent genome instability.

RESULTS

1. SCREENING FOR FACTORS INVOLVED IN R LOOP HOMEOSTASIS

With the aim at screening for factors involved in R loop-homeostasis we have developed in our laboratory a system based on the action of the human cytidine deaminase AID on single strand DNA (ssDNA), particularly on the displaced DNA strand of an R loop, deaminating Cs that finally leads to DNA double strand breaks (DSBs) by the action of other enzymes (Figure R1A) (Aguilera & Gómez-González, 2008; Basu et al., 2011). The system, that we referred from now on, AIRD (AID inducible R loop detection cell line), consists of a stable U2OS cell line carrying a tetracycline-regulated human cytidine deaminase AID gene cassette (Figure R1B). We expected that cells accumulating high-levels of R loops would generate high levels of DNA breaks that could be detected by immunofluorescence as γH2AX foci.

In order to get further insight into the connection between RNA-DNA hybrids and genome integrity, in this thesis we have used the AIRD system as a tool to perform an siRNA high-throughput microscopy screening by γH2AX immunofluorescence and look for new factors related to R-loop-mediated genome instability.

1.1. A high-throughput screening for factors involved in R loop-mediated genome instability

To validate AIRD system, we measured DNA damage by γH2AX immunofluorescence in cells depleted of FANCD2, a factor involved in R loop metabolism, with and without AID expression (M. L. García-Rubio et al., 2015; Schwab et al., 2015). As previously reported, transient depletion of FANCD2 in HeLa cells leads to a higher number of γH2AX foci per cell in comparison with siC control cells. Importantly, a significant increase of DNA damage could be observed in siFANCD2 after AID induction (+DOX), and to less extent in siC, compared to their respective non-induced conditions (-DOX) (Figure R1C). Furthermore, in collaboration with Jose Calderón in Aguilera's lab, the AIRD system was validated for FANCD2 and other factors such as THOC1, an mRNP protein that prevents R loop formation (Domínguez-Sánchez et al., 2011), in 96-well plates and automated microscope in order to set up the conditions for a high-throughput screening (Appendix 1).

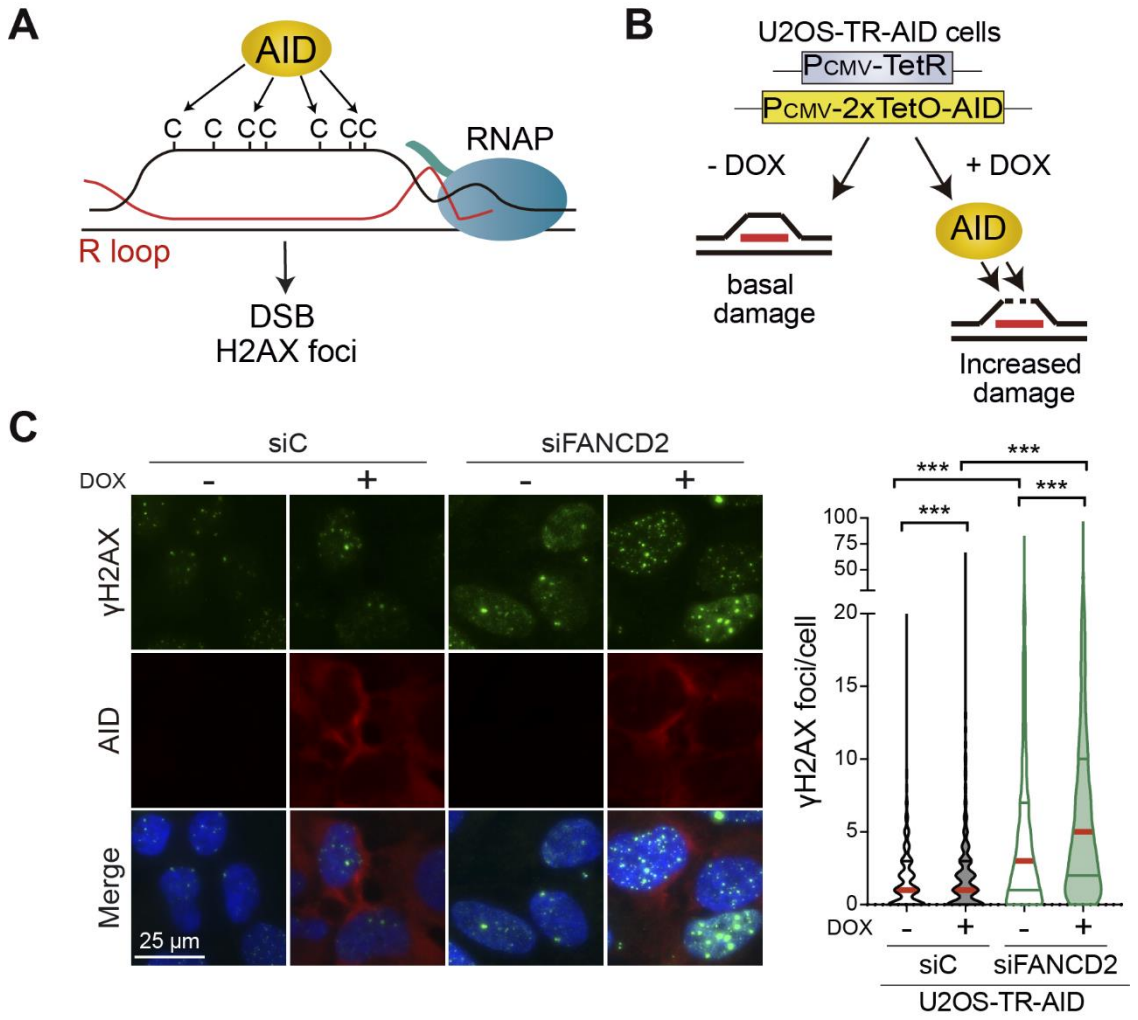


Figure R1. Validation of the U2OS-TR-AID cell line for inducible AID-based R-loop detection system (AIRD). (A) Schematic illustration of the AID-based principle for detecting factors involved in R-loop metabolism. **(B)** U2OS-TR-AID stable cell line carrying a tetracycline-regulated AID gene cassette is used as a tool to induce an increase of DNA double-strand breaks (DSBs) in an R-loop dependent manner. **(C)** Immunofluorescence of γH2AX in siC and siFACND2- transfected U2OS-TR-AID cells (clone #8) expressing (+DOX) or not expressing AID (-DOX) as indicated. Nuclei were stained with DAPI. Scale bars, 25 μm. Quantification of the number of γH2AX foci per cell is shown. The red line indicates the median (n=3). The statistical significance of the difference was calculated with Mann-Whitney U-test; ***P<0.001.

In collaboration with Sonia Silva, Lola P. Camino and Jose Javier Marqueta-Gracia, we performed an siRNA high-throughput screening with the AIRD system and analyzed a collection of siRNAs targeting 3205 human genes involved in different processes, as apoptosis, nucleic acid binding, autophagy and others, that are considered potential targets for therapeutic drugs (3205 out of 4796 siRNAs of Dharmacon-ON TARGET Plus-Druggable genome siRNA library) (Appendix 2). We prepared 96-well plates containing a duplicate of each four-siRNA pool for every targeted gene, and siRNA against FANCD2 and a non-targeting siRNA (siC) were included as positive and negative control respectively.

For the screening, U2OS-TR-AID cells were transfected, treated or not with doxycycline (DOX) to induce AID expression and γH2AX detection was acquired by automated microscopy of cells with IMAGE-Express (137239) equipment available at Genomic Unit at CABIMER (Materials and Methods 6; Figure M1; Figure R2A). We reasoned that comparison of γH2AX signal before and after expression of AID in siRNA transfected cells would allow for the identification of factors involved in R-loop metabolism.

The screening was performed in duplicate, and siRNAs whose depletion lead to an increase (≥1.2) in the percentage of cells with γH2AX foci upon AID induction (siRNA +DOX), versus non-induced conditions (siRNA -DOX), that was higher (≥1.2) to that observed in control cells (siC +DOX) were selected as candidates. With these criteria, 156 out of 3205 siRNAs tested were chosen for further γH2AX immunofluorescence analysis (Figure R2B; Appendix 3 and 4). These candidates were analyzed in a second round and 46 hits were confirmed, taking into account the increase in the percentage of cells with γH2AX foci and the reproducibility between duplicates of the two different experiments (Figure R2C; Table R1; Appendix 5).

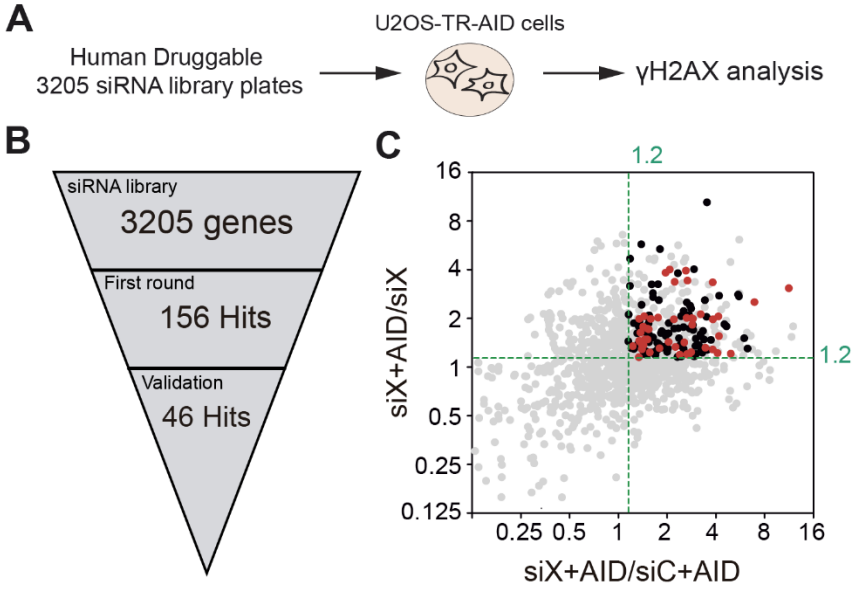


Figure R2. Candidates of high-throughput screening validation. (A) Schematic workflow for testing the human druggable siRNA library (3205 siRNAs). **(B)** Scheme of hits obtained in the different steps of the screening. **(C)** Results of the second round of γH2AX immunofluorescence screening analysis. The percentages of cells with ≥5 or ≥10 γH2AX foci was used to calculate the indicated ratios, represented in both axis: Y, siRNA with AID versus itself without AID expression; X, siRNA with AID versus the median of siC with AID expression. The dotted lines show the cutoff (1.2) used to designate a siRNA as positive. Two ratios for each siRNA are represented by a dot: black dots: candidates with ≥10 γH2AX foci per cell; red dots: candidates with ≥5 γH2AX foci per cell; grey dots: others.

	Replicate 1								Replicate 2								
	Duplicate 1				Duplicate 2				Duplicate 1				Duplicate 2				
	≥ 5 foci/cell		≥ 10 foci/cell		≥ 5 foci/cell		≥ 10 foci/cell		≥ 5 foci/cell		≥ 10 foci/cell		≥ 5 foci/cell		≥ 10 foci/cell		
Gene Symbol	Ratio 1	Ratio 2	Ratio 1	Ratio 2	Ratio 1	Ratio 2	Ratio 1	Ratio 2	Ratio 1	Ratio 2	Ratio 1	Ratio 2	Ratio 1	Ratio 2	Ratio 1	Ratio 2	Reproducibility score
MECP2	2.0	1.4	2.0	3.8	1.5	1.5	1.5	1.7	1.2	1.3	1.6	2.0	1.4	1.4	2.2	3.4	64
ARHGAP1	1.4	1.4	4.1	1.2	1.4	1.4	2.4	1.2	1.3	1.6	2.6	2.0	1.5	1.1	2.8	2.0	42
ATP11A	2.2	2.0	3.8	3.3	1.3	2.0	1.8	2.0	2.8	1.1	3.8	1.3	2.4	1.4	2.8	1.2	42
PSMC5	1.3	1.5	4.1	2.1	3.2	2.1	6.9	2.5	0.9	1.6	2.9	2.0	2.1	1.8	2.9	1.8	42
PSMD7	3.8	2.0	11.3	3.1	2.6	1.4	4.9	1.2	1.3	1.1	3.4	1.3	2.4	1.9	2.3	1.3	42
SPINK5	1.4	1.1	1.3	1.2	1.4	2.1	2.1	4.0	1.4	1.6	1.5	1.7	1.4	1.3	1.8	1.3	42
ARNTL2	2.7	3.4	2.6	4.0	2.7	1.2	2.2	0.6	1.5	1.5	1.9	2.8	1.4	1.8	1.3	0.9	24
GNAI1	1.6	1.2	4.2	1.6	2.7	1.2	4.5	1.0	1.4	1.1	1.5	1.3	1.2	1.0	1.2	1.3	20
PAM	2.2	1.7	3.0	1.2	1.7	1.4	3.2	1.6	1.1	0.9	2.0	0.9	1.4	1.5	1.6	1.1	20
PHOX2A	2.3	1.1	3.1	1.7	1.3	0.8	2.8	0.8	2.8	1.7	4.7	1.8	1.9	1.7	2.0	1.4	20
CCS	1.5	1.4	2.0	1.6	2.0	1.6	2.9	1.8	1.1	1.5	1.0	2.2	0.1	0.3	0.2	0.5	16
EGLN1	3.5	1.3	6.3	1.3	2.5	1.3	3.5	2.0	1.1	0.8	0.2	0.2	0.5	0.4	0.1	0.2	16
GFRA3	1.5	1.9	2.7	2.1	1.4	1.2	3.0	1.4	0.6	0.6	0.3	0.4	0.7	1.2	0.6	0.9	16
MYOG	2.0	1.6	4.2	2.8	1.6	1.2	2.4	1.2	0.9	2.9	1.0	5.8	0.5	1.3	0.2	1.0	16
RHOC	1.1	1.5	0.9	1.8	1.1	1.8	0.8	1.3	1.5	1.6	2.1	1.9	2.4	1.4	2.4	2.1	16
SLC25A12	2.6	1.2	2.9	1.2	1.5	2.1	1.4	5.7	1.1	0.8	0.5	0.3	0.8	0.8	0.1	1.1	16
APOB	2.2	1.8	3.6	3.2	1.3	0.8	2.1	0.7	2.9	1.7	4.6	2.4	2.3	1.2	2.5	0.9	10
RPUSD1	0.9	1.4	1.6	2.7	1.2	2.6	1.8	3.3	1.0	0.7	0.9	1.3	1.9	2.2	2.7	2.9	10
SLC24A1	0.8	1.7	1.4	2.4	1.3	1.9	1.2	4.7	1.2	1.0	1.0	0.9	1.2	1.5	1.3	1.7	10
TNPO1	1.3	2.0	2.0	1.3	1.1	1.3	1.2	2.1	1.6	1.5	3.9	1.8	1.5	0.9	1.6	0.7	10
TUBB6	0.9	0.9	1.3	1.6	1.5	1.3	3.4	1.5	1.2	1.6	1.3	2.2	0.6	1.0	0.1	0.9	10
CHRAC1	1.1	1.0	3.0	1.5	2.0	1.7	3.3	1.7	1.2	1.6	0.8	1.1	0.4	0.7	0.8	2.0	8
CHST8	1.3	1.0	3.4	2.6	1.5	1.2	1.8	1.6	1.0	1.5	2.1	1.4	1.1	0.9	0.9	0.7	8
COX8C	2.5	1.8	5.6	2.9	1.3	1.4	1.0	0.7	1.1	1.3	2.8	1.9	0.9	0.5	0.7	0.4	8
CTPS2	1.0	0.5	1.8	0.8	1.6	1.6	1.9	1.1	1.1	1.2	2.8	2.4	2.7	1.6	5.6	2.7	8
EEF1D	2.1	4.8	3.5	10.5	1.1	2.2	1.6	3.2	1.4	1.3	0.3	0.3	0.7	0.6	0.0	0.0	8
KIR3DL3	1.3	21.8	0.9	-	0.8	2.0	0.3	-	1.1	1.2	2.5	2.2	1.6	1.9	2.3	3.8	8
SLC30A2	1.8	1.4	1.9	1.0	1.5	2.0	2.5	3.3	1.0	2.5	0.4	-	0.6	1.9	1.2	2.0	8
TOPBP1	1.2	1.2	0.8	0.6	0.6	0.9	0.1	0.2	0.8	0.5	1.8	1.4	4.1	1.3	6.0	1.5	8
ACADL	2.6	0.9	6.2	0.8	1.7	0.8	3.7	0.8	3.3	1.2	7.9	1.1	3.2	1.3	7.3	1.2	6
DDX42	2.1	1.2	3.1	1.4	1.0	1.2	1.8	5.4	0.9	0.9	0.9	1.1	1.1	0.7	0.9	0.8	6
DDX47	0.8	0.7	1.2	1.4	1.4	1.3	1.3	1.3	1.0	1.5	1.0	1.8	0.6	0.9	0.5	0.8	6
DHX9	1.0	2.2	0.4	-	0.4	1.3	0.1	-	0.4	0.7	1.4	1.9	1.4	1.6	1.5	1.6	6
EIF3A	1.0	1.1	2.0	1.8	0.7	1.4	0.6	1.4	2.1	1.0	3.7	1.5	2.5	0.9	4.5	1.8	6
FGF10	0.9	1.0	1.0	0.4	0.9	0.6	0.9	0.4	1.4	1.2	1.2	1.3	1.1	1.0	1.2	1.9	6
HDAC8	1.8	1.3	2.2	1.2	1.7	1.2	1.8	1.1	0.4	2.1	0.3	2.1	0.4	1.5	0.6	-	6
HSDL1	1.9	1.3	3.4	1.0	2.4	1.5	3.1	2.1	0.7	0.9	0.9	0.5	1.3	0.8	1.8	1.0	6
IL4I1	1.5	1.2	1.2	1.1	1.7	1.3	2.4	1.2	0.7	1.4	0.3	2.0	0.4	1.1	0.4	2.1	6
MCCC1	1.0	0.6	1.3	0.3	1.0	1.2	1.6	2.0	1.1	1.4	2.6	2.5	2.0	1.1	3.3	1.3	6
NR2E3	0.9	1.1	0.6	0.9	0.9	0.8	0.3	0.3	1.6	1.1	2.0	1.2	2.9	1.4	5.5	2.8	6
PIGZ	1.4	1.1	1.6	1.5	1.7	1.4	2.8	2.3	0.9	1.0	0.6	1.2	0.2	0.5	0.0	0.0	6
SLC15A2	2.0	1.8	1.8	2.6	1.4	0.9	1.6	2.9	0.8	2.6	1.1	6.6	0.3	0.6	0.2	0.2	6
SLC41A2	2.0	1.0	2.9	4.0	1.5	1.3	1.3	1.8	0.8	0.9	0.6	0.8	0.6	1.6	0.2	0.6	6
ALOX5	1.0	2.2	1.7	1.3	1.9	1.1	2.3	1.2	0.6	1.3	0.3	0.6	0.6	1.1	0.5	3.2	4
EIF3C	4.0	1.1	7.4	1.0	2.7	0.9	7.4	0.9	2.1	1.1	3.2	1.6	0.9	1.2	1.2	3.2	4
STAMBP	0.8	1.3	0.4	0.7	0.8	1.4	0.3	0.6	1.4	1.7	2.4	0.8	2.3	1.7	1.8	1.0	4

***Ratio 1**: siX+AID +/siC+ AID
***Ratio 2**: siX+AID/siX

Table R1. γH2AX foci/cell analysis of validated candidate genes. Data show the ratios between percentages of cells with ≥5 or ≥10 γH2AX foci with and without AID expression upon siRNA depletion of the 46 candidate genes. The two ratios for each siRNA consist on: one versus itself with and without AID expression (siX+DOX/siX), and the other one versus the median of siC with AID expression in its plate (siX+DOX/siC+DOX). The selected hits (46 genes) defined as those with both ratios ≥ 1.2, at least in two duplicates of one replicate.

HITS	Gene Name	Molecular Function
ACADL	Long-chain specific acyl-CoA dehydrogenase	fatty-acyl-CoA binding; acyl-CoA dehydrogenase activity
ALOX5	Arachidonate 5-lipoxygenase	arachidonate 5-lipoxygenase activity; iron ion binding
APOB	Apolipoprotein B-100	lipid transporter activity
ARHGAP1	Rho GTPase-activating protein 1	SH3/SH2 adaptor activity; GTPase activator activity
ARNTL2	Aryl hydrocarbon receptor nuclear translocator-like protein 2	DNA-binding transcription factor activity, RNA polymerase II-specific
ATP11A	Probable phospholipid-transporting ATPase IH	ATPase-coupled intramembrane lipid transporter activity
CCS	Copper chaperone for superoxide dismutase	copper ion binding; oxidoreductase activity
CHRAC1	Chromatin accessibility complex protein 1	RNA polymerase II proximal promoter sequence-specific DNA binding
CHST8	Carbohydrate sulfotransferase 8	sulfotransferase activity
COX8C	Cytochrome c oxidase subunit 8C	cytochrome-c oxidase activity
CTPS2	CTP synthase 2	identical protein binding; ligase activity
DDX42	ATP-dependent RNA helicase DDX42	nucleic acid binding
DDX47	Probable ATP-dependent RNA helicase DDX47	nucleic acid binding
DHX9	ATP-dependent RNA helicase A	RNA binding; 3'-5' DNA helicase activity
EEF1D	Elongation factor 1-delta	guanyl-nucleotide exchange factor activity
EGLN1	Egl nine homolog 1	iron ion binding
EIF3A	Eukaryotic translation initiation factor 3 subunit A	translation initiation factor activity; mRNA binding
EIF3C	Eukaryotic translation initiation factor 3 subunit C	translation initiation factor activity
FGF10	Fibroblast growth factor 10	fibroblast growth factor receptor binding
GFRA3	GDNF family receptor alpha-3	cytokine binding
GNAI1	Guanine nucleotide-binding protein G(i) subunit alpha-1	G protein-coupled receptor binding
HDAC8	Histone deacetylase 8	histone deacetylase activity
HSDL1	Inactive hydroxysteroid dehydrogenase-like protein 1	protein binding
IL4I1	L-amino-acid oxidase	oxidoreductase activity
KIR3DL3	Killer cell immunoglobulin-like receptor 3DL3	transmembrane glycoprotein
MCCC1	Methylcrotonoyl-CoA carboxylase subunit alpha	biotin carboxylase activity
MECP2	Methyl-CpG-binding protein 2	DNA binding
MYOG	Myogenin	RNA polymerase II proximal promoter sequence-specific DNA binding
NR2E3	Photoreceptor-specific nuclear receptor	DNA-binding transcription factor activity
PAM	Peptidyl-glycine alpha-amidating monooxygenase	peptidylglycine monooxygenase activity
PHOX2A	Paired mesoderm homeobox protein 2A	DNA-binding transcription factor activity, RNA polymerase II-specific
PIGZ	GPI mannosyltransferase 4	mannosyltransferase activity
PSMC5	26S proteasome regulatory subunit 8	TBP-class protein binding
PSMD7	26S proteasome non-ATPase regulatory subunit 7	protein binding
RHOC	Rho-related GTP-binding protein RhoC	GTP binding
RPUSD1	RNA pseudouridylate synthase domain-containing protein 1	intramolecular transferase activity
SLC15A2	Solute carrier family 15 member 2	dipeptide transmembrane transporter activity
SLC24A1	Sodium/potassium/calcium exchanger 1	potassium ion transmembrane transporter activity
SLC25A12	Calcium-binding mitochondrial carrier protein Aralar1	L-glutamate transmembrane transporter activity
SLC30A2	Zinc transporter 2	metal ion transmembrane transporter activity
SLC41A2	Solute carrier family 41 member 2	cation transmembrane transporter activity
SPINK5	Serine protease inhibitor Kazal-type 5	endopeptidase inhibitor activity
STAMBP	STAM-binding protein	thiol-dependent ubiquitin-specific protease activity
TNPO1	Transportin-1	signal sequence binding; transporter activity
TOPBP1	DNA topoisomerase 2-binding protein 1	DNA binding; protein binding
TUBB6	Tubulin beta-6 chain	GTP binding

Table R2. List of validated hits by γH2AX foci screening. Table contains the list and description of genes identified in this study. Data information about molecular function was obtained from The Human Gene Database (https://www.genecards.org/).

1.2. Network analysis of hits from high-throughput screening

To study the spectrum of biological functions within the selected hits from the γH2AX foci analysis, we used PANTHER (protein Analysis through Evolutionary Relationships) and tested the 46 hits against the 3205 siRNA library as background. The enriched categories of genes we found included those with roles in regulation of RNA stability, ribonucleoprotein complex biogenesis, translational initiation, nucleobase-containing compound transport, cell cycle and RNA processing (Figure R3). Thus, we found a significant enrichment in genes related with transcription and RNA biogenesis, cell cycle as well, something expected of genes that could be related to R loop and transcription-associated genome instability, providing confidence in our results. Indeed, among the candidates we found DHX9, an helicase that has been identified in an RNA-DNA hybrid interactome and that plays a role in the prevention of R-loop associated DNA damage (Cristini et al., 2018). A detailed description of the functions of the 46 candidates is shown in Table R2.

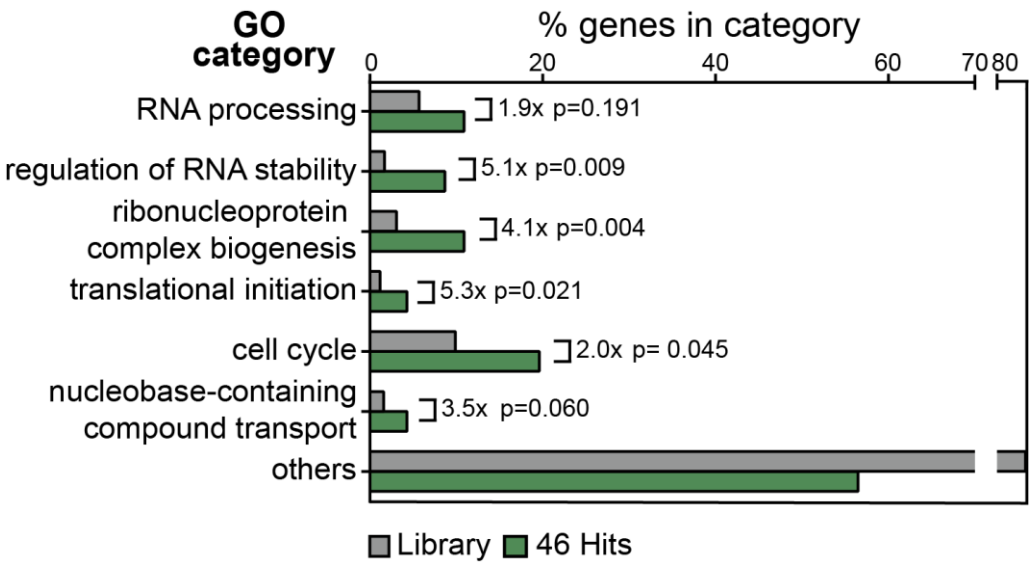


Figure R3. Gene Ontology analysis for putative R-loop forming factors from high-throughput screening. Fold enrichment and significance levels for Gene Ontology (GO) Biological Process enrichment analysis for putative R-loop forming factors. The percentage of genes contained in the most representative GO categories is plotted for primary library (3205 siRNAs) and validated hits. Enrichment was analyzed using PANTHER (P-value calculated using Fischer's exact test).

We completed the analysis with a network interaction map of the hits obtained (using the GeneMANIA plugin for Cytoscape v3.8.0 to build a network of known interactions, both physical and genetic, shared protein domains and co-localization) that highlighted a close relationship for most of the selected genes (Figure R4A). Interestingly, strong interaction modules were those encompassing pathways related to protein degradation, translation, some RNA helicases and signal transduction by Rho-GTPases.

In order to have a representative sample, specific hits were chosen based on their reproducibility (Figure R4B), their described functional categories (Table R2) and our network analysis (Figure R4A). With these criteria, in this thesis, we focused on 7 selected genes to be validated, including proteins related with chromatin modification (MECP2 a Methyl-CpG Binding Protein and HDAC8 an histone deacetylase); RNA helicases (DDX42 and DDX47 DEAD box helicases); transcription factors (MYOG); and factors involved in protein degradation (PSMC5 and PMSD7 proteasome subunits). Thus, we preferred in this thesis to focus on factors closed related to DNA and RNA metabolism, and we also included the proteasome given its relationship with transcription (Ferdous et al., 2001; Geng et al., 2012). It is worth to notice that the increase in DNA damage associated to AID induction was also observed upon depletion of genes involved in compound transport, metabolism and cell signaling such small GTPases and other proteins. Further analysis would be needed to study the putative role of these factors on R loop metabolism.

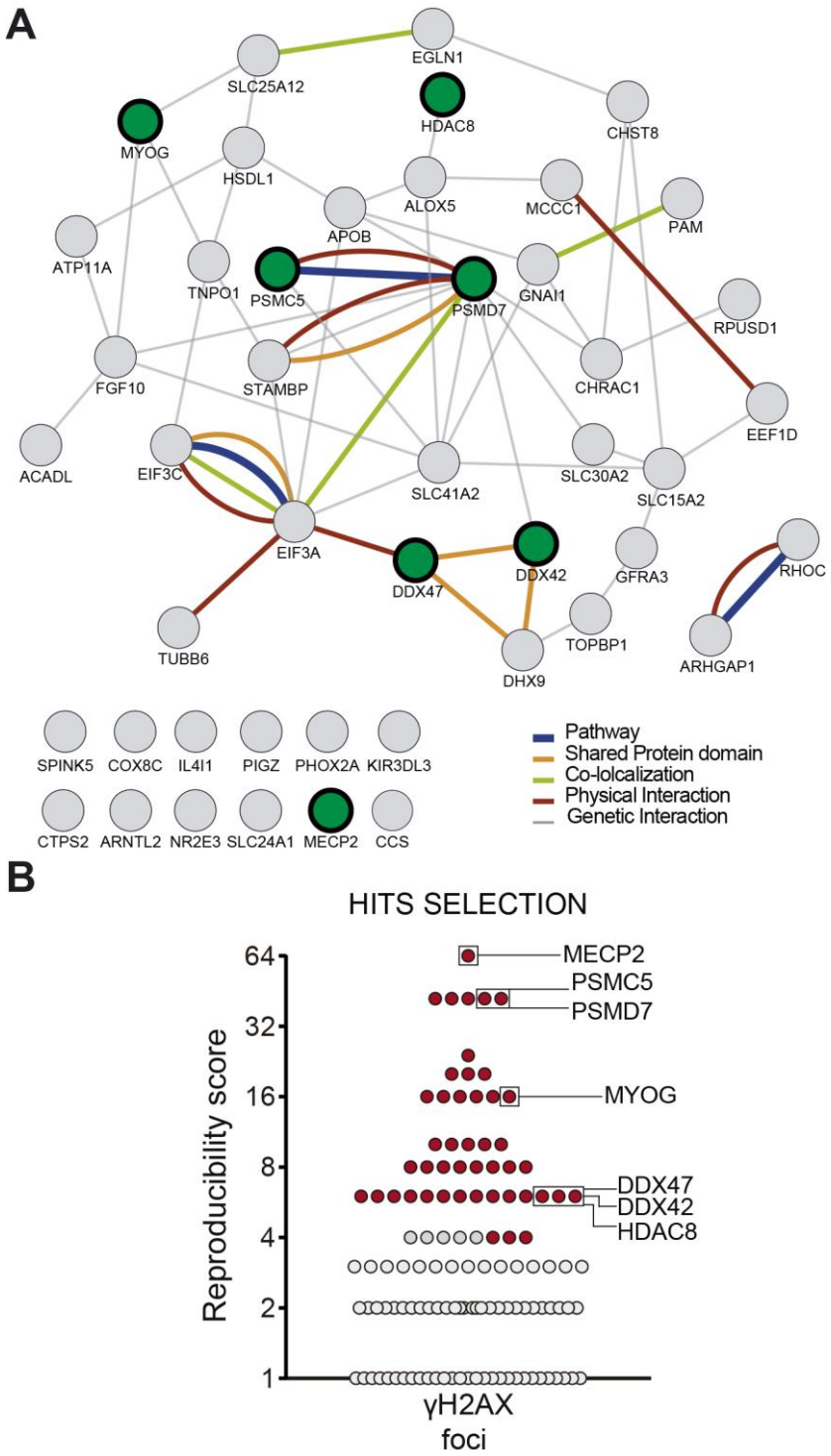


Figure R4. Functional network interaction analysis and selection of hits from high-throughput screening. (A) Networks of interacting proteins identified in the screening using (GeneMANIA software). γH2AX foci hits (grey circles), selected hits (dark green and highlighted circles). **(B)** Graphic showing the genes selected from the γH2AX foci screening (red circles). Reproducibility scores are plotted on the Y-axis. The name of selected hits for further studies is shown.

1.3. Analysis of spontaneous DNA damage in the selected candidates

Since our screening was focused on DNA damage induced by AID, to complete the first step of characterization, we were also interested in determine the levels of spontaneous DNA damage by γH2AX foci in U2OS cells depleted of the selected hits. For this purpose, we performed the experiments in cover slips and with a pool of two siRNAs per each gene, validated by RT-qPCR (Figure R5A). Immunofluorescence analysis indicate that the number of γH2AX foci per cell was similar to those of siC control cells in siRNA-depleted cells, except for HDAC8 siRNA transfected cells. A slight increase in spontaneous DNA damage was also observed in DDX42 siRNA depleted cells (Figure R5B).

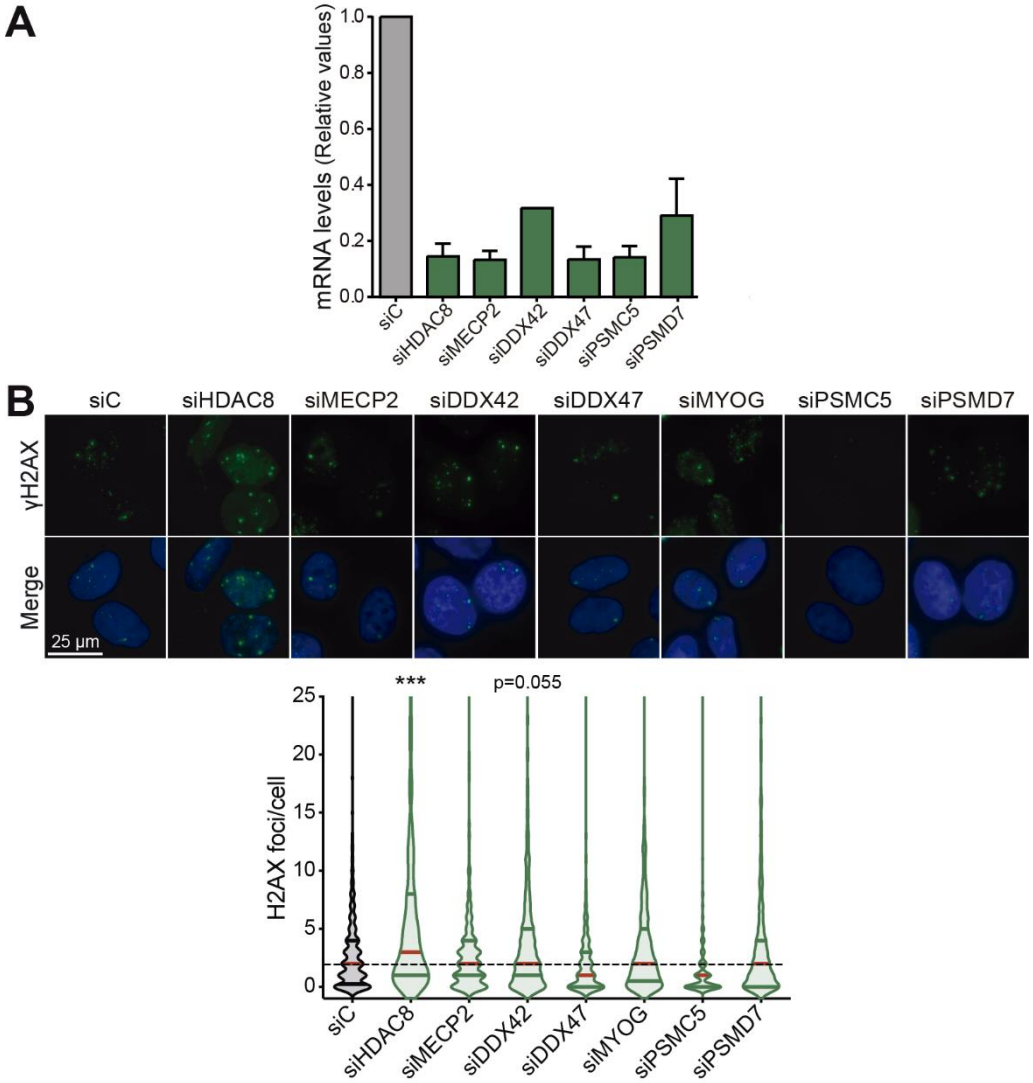


Figure R5. Spontaneous damage upon siRNA silencing of selected candidate genes. (A) Relative mRNA levels as measured by RT-qPCR in siRNA transfected U2OS cells after 72 hours of transfection. MYOG depletion is not determined (nd) because technical problems. HPRT housekeeping gene was used to normalize mRNA expression values. The mean ± SEM is shown. **(B)** Representative images of immunostaining with γH2AX (green) antibody in U2OS cells upon siRNA depletion. Quantification of the number of γH2AX foci per cell is shown (n=3). The red line in violin plot indicates the median. ***P < 0.001 (Mann-Whitney U-test, two-tailed).

1.4. Analysis of R loop accumulation in the selected hits

Our screening was based on the comparison of DNA damage before and after DNA damage in siRNA-depleted cells with the aim to identify factors involved in R-loop metabolism. We explored whether this was due to an increased accumulation of R loops. We first assayed R loops by immunofluorescence (IF) using the anti-DNA–RNA hybrid S9.6 monoclonal antibody in siRNA-depleted cells. As shown in Figure R6, all these candidates showed a significant increase in S9.6 nuclear signal respect to siC control cells.

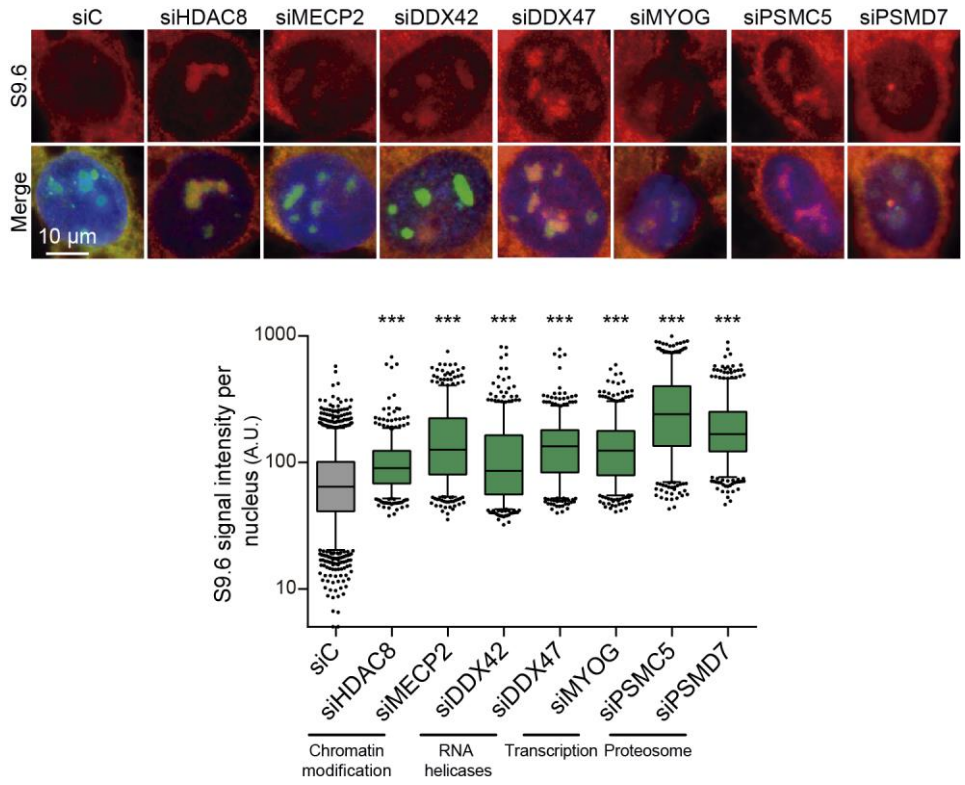


Figure R6. RNA-DNA hybrids analysis in candidate-siRNA depleted cells by S9.6 immunofluorescence microscopy. Representative images of U2OS cells immunostained with S9.6 (red) and nucleolin (green) antibodies after transfection with at least a pool of two siRNAs of each original pool. The median of S9.6 signal intensity per nucleus after nucleolar signal removal in U2OS cells transfected with the indicated siRNAs is shown (n=3). Boxes and whiskers indicate 5-95 percentiles. ***P < 0.001 (Mann-Whitney U-test, two-tailed). A.U., Arbitrary units.

To confirm further the accumulation of RNA-DNA hybrids, we performed RNA-DNA immunoprecipitation (DRIP) in two human genes (APOE and RPL13A) that have been previously validated for R-loop detection (Ginno et al., 2013; Herrera-Moyano et al., 2014). *In vitro* treatment with RNAse H was used to confirm the specificity of immunoprecipitation of DNA-RNA hybrids. Depletion of most of the candidates (6 out 7) increased the DRIP signal respect to the siC levels, in at least one of the studied genes (Figure R7). Then, we considered as the top-hit candidates those genes in which an accumulation of RNA-DNA hybrids was detected by both, S9.6 IF and DRIP analysis (relative values above 1.5 of siC levels) (MECP2, DDX42, DDX47, MYOG) (Figure R6 and R7).

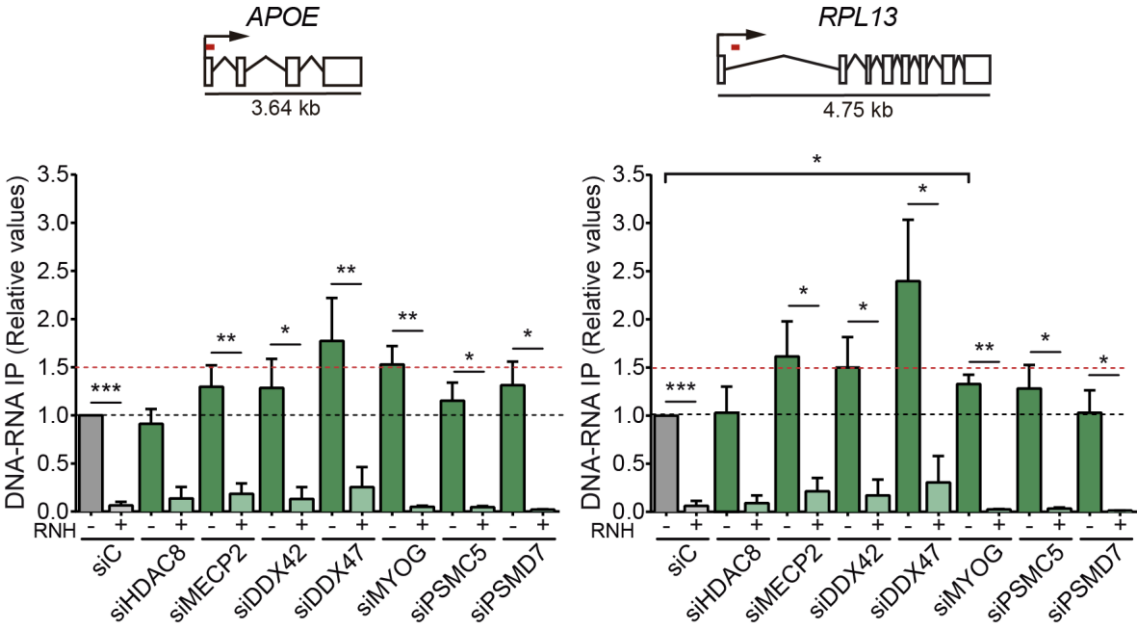


Figure R7. RNA-DNA hybrids analysis in candidate-siRNA depleted cells by DRIP assays. Relative DRIP-qPCR signal values at APOE and RPL13A genes in U2OS cells transfected with the indicated siRNAs and samples were treated in vitro with RNase H prior immunoprecipitation where indicated. Red lines indicate the regions where the primers used for PCR amplification were located. Data are plotted as mean ± SEM (n = 3). *P < 0.05; **P < 0.01; ***P < 0.001 (one-tailed paired t-test).

We decided to focus our study in two out of the top-hits R-loop accumulating candidates validated by S9.6 IF and DRIP that showed different activities or properties on the nucleic acids metabolism: DDX47, a nucleolar helicase, that showed the highest values in the DRIP analysis (Figure R7) and a DNA binding protein, MECP2, the methyl-CpG binding protein 2, being the candidate with the highest reproducibility (Figure R4B).

2. VALIDATION OF DDX47 AND MECP2 AS FACTORS RELATED TO R LOOP METABOLISM AND GENOME INSTABILITY

2.1. R loop accumulation in DDX47- and MECP2- depleted cells

To validate the selected candidates, DDX47 and MECP2, we determined the levels of R loops after siRNA depletion in another cell line different to the U2OS cells used in the screening. HeLa cell were transfected with specific siRNAs and first a clear reduction in DDX47 and MECP2 mRNA levels and proteins was detected by RT-qPCR and western, respectively after 72 hours of transfection (Figure R8A and R8B).

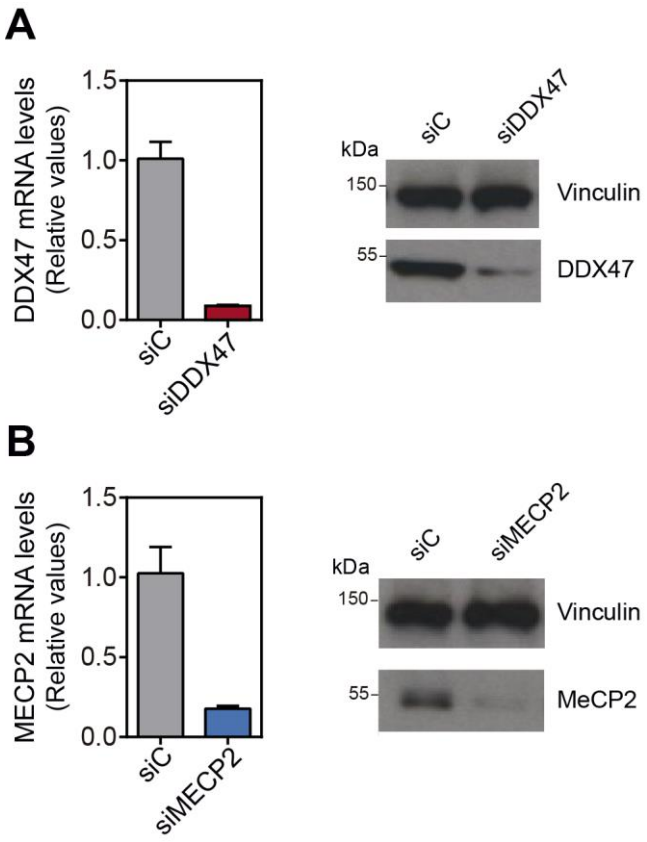


Figure R8. siRNA silencing of DDX47 and MECP2. (A) Relative DDX47 mRNA levels as measured by RT-qPCR in siRNA transfected HeLa cells after 72 hours of transfection (left panel). HPRT housekeeping gene was used to normalize mRNA expression values. Error bars represent relative target quantity (RQ) minimum and maximum from three technical replicates. Western blot analysis of siRNA-treated HeLa cells with siC or siDDX47 (right panel). Vinculin protein is used as a loading control. **(B)** Analysis of MECP2-siRNA silencing in HeLa cells. Other details as in figure A.

Then, S9.6 IF assays were performed in siRNA depleted cells including an *in vitro* treatment with RNase H, which degrades the RNA from the RNA-DNA hybrid. A significant enrichment of the antibody S9.6 nuclear signal was observed for both siDDX47 and siMECP2 cells. Importantly, this increase was suppressed by RNaseH1 treatment, confirming that they correspond to RNA-DNA hybrids (Figure R9).

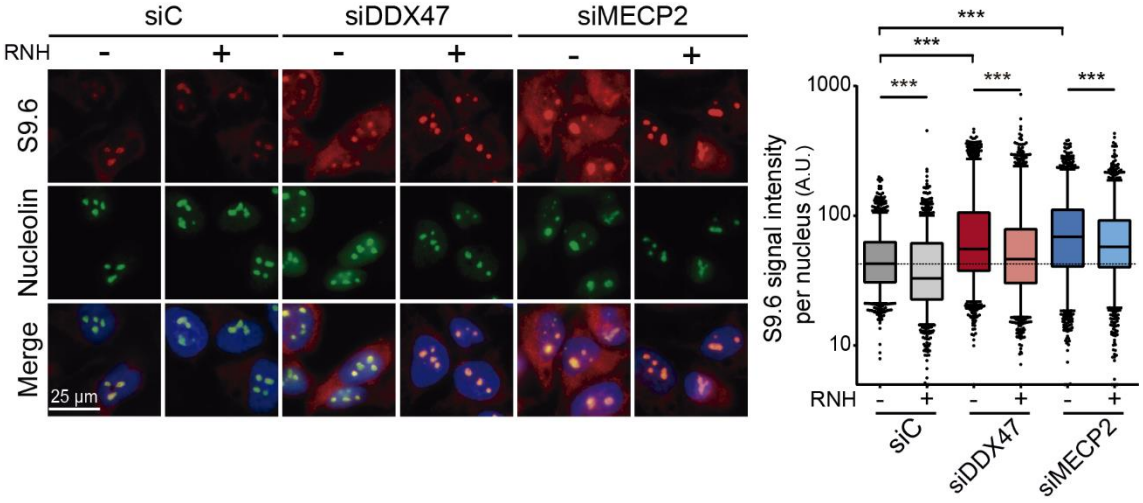


Figure R9. RNA-DNA hybrids accumulation in siDDX47 and siMECP2 HeLa cells by S9.6 immunofluorescence microscopy. Representative images of immunostaining with S9.6 (red) and anti-nucleolin (green) antibodies in HeLa cells upon DDX47 and MECP2 depletion. The median of S9.6 signal intensity per nucleus after nucleolar signal removal in siC, siDDX47 and siMECP2 HeLa cells and treated in vitro with RNase H where indicated. Data from more than 250 total cells from at least three independent experiments is shown. Boxes and whiskers indicate 5-95 percentiles. ***P < 0.001 (Mann-Whitney U test, two-tailed). A.U., Arbitrary units.

R loop accumulation at molecular level was also validated by DRIP-qPCR analysis in the standard human genes (APOE and RPL13A) and in 18S and 28S rDNA genes, given the nucleolar localization of the helicase DDX47 (Sekiguchi et al., 2006) (Figure R10A). *In vitro* RNase H treatment was used in order to confirm the RNA-DNA hybrids specific immunoprecipitation. Results clearly show a higher accumulation of RNA–DNA hybrids in siDDX47 cells in rDNA regions, but also in RNAPII transcribed (Figure R10B, left panel). In the case of siMECP2, we extended the analysis including *PHLDA2*, a gene that is regulated by MECP2 (Meng et al., 2014), and *YIF1A* as another example of MECP2 ChIP-enriched region (determined by ChIP seq in HCT116 cell type- GSE47677). A significant RNA-DNA hybrid accumulation was observed in *APOE*, *RPL13*, and *PHLDA2* genes, and high levels but no significant were detected at *YIF1A* and 28S regions (Figure R10B, right panel).

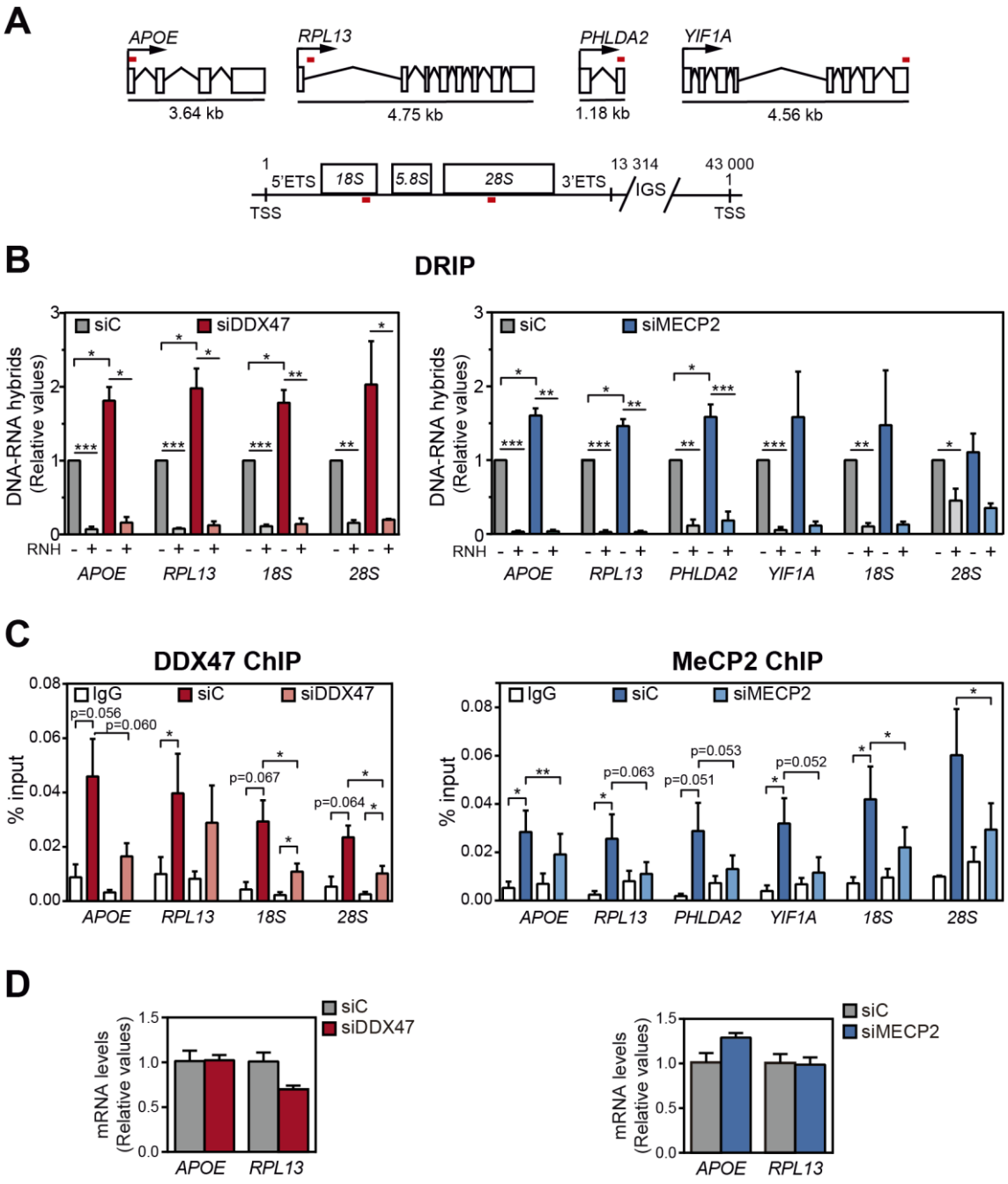


Figure R10. R loop accumulation in siDDX47 and siMECP2 Hela cells. (A) Schematic diagrams of APOE, RPL13, PHLDA2, YIF1A, 18S and 28S genes. Exons are depicted as open boxes, the arrows indicate the start of transcription and red lines indicate the regions where the primer pairs used for amplification were located. **(B)** Relative DRIP-qPCR signal values in siDDX47 (red) and siMECP2 (blue) HeLa cells at the indicated regions. Samples were treated in vitro with RNase H prior immunoprecipitation where indicated. Signal values normalized with respect to the siC control are plotted (n=3) as mean and SEM. *P < 0.05; **P < 0.01; ***P < 0.001 (one-tailed paired t-test). **(C)** ChIP analysis of DDX47 and MECP2 proteins at APOE, RPL13, 18S, 28S, PHLDA2 and YIF1A genes where indicated. Values represent the percentage of the precipitated DNA (IP) to input DNA (INPUT). Immunoprecipitation in siDDX47 and siMECP2 cells were included as control of specificity. IgG was used as negative control. Data are plotted as mean ± SEM (n = 3). *P < 0.05 (one-tailed paired t-test). **(D)** mRNA levels of APOE and RPL13 in siDDX47 and siMECP2 cells as determined by RT-qPCR. Error bars represent relative target quantity (RQ) minimum and maximum from three technical replicates.

In parallel, we performed ChIP analysis to study the recruitment of DDX47 and MECP2 to the selected genes analysed in DRIP assays, and found that both proteins were bound to these chromatin regions consistently with a possible role of these factors in the prevention of R loop (Figure R10C). Importantly, these high levels of RNA-DNA hybrids were not due to an increase in transcription at least in APOE and RPL13 genes, since no significant differences were observed in mRNA levels, as detected by RT-qPCR (Figure R10D).

Altogether data indicate that DDX47 and MECP2 were necessary to prevent R loop accumulation.

2.2. R loop-dependent genome instability phenotype in DDX47 and MECP2 depleted cells

Once we validated R loop accumulation and AID dependent genome instability upon DDX47 and MECP2 depletion, we asked whether these factors play a role in genome instability that could be mediated by RNA-DNA hybrids. First, we analyzed the spontaneous DNA damage in HeLa siRNA transfected cells by γH2AX immunofluorescence. A significant increase in γH2AX foci number per cell was detected in siMECP2 and a mild increase was also found in siDDX47 cells (Figure R11A). Single-cell electrophoresis analysis revealed that depletions of both proteins increased DNA breaks, that were partially supressed by RNH1 overexpression (Figure R11B) in agreement with the AID-dependent genome instability phenotype of these genes in the siRNA screening.

Altogether data confirm and validate that DDX47 and MECP2 are genes involved in R loop homeostasis and that are necessary to prevent R-loop dependent genome instability. In spite of the interesting role of MECP2 as a critical regulator of the maintenance of chromatin architecture (Ragione et al., 2016) and its connection with genome instability (Enikanolaiye et al., 2020), we decided to focus on a protein closed related to RNA metabolism, the helicase DDX47. In this thesis, previous experience and available tools were used to explore further the role of this helicase in transcription and RNA-DNA hybrid metabolism.

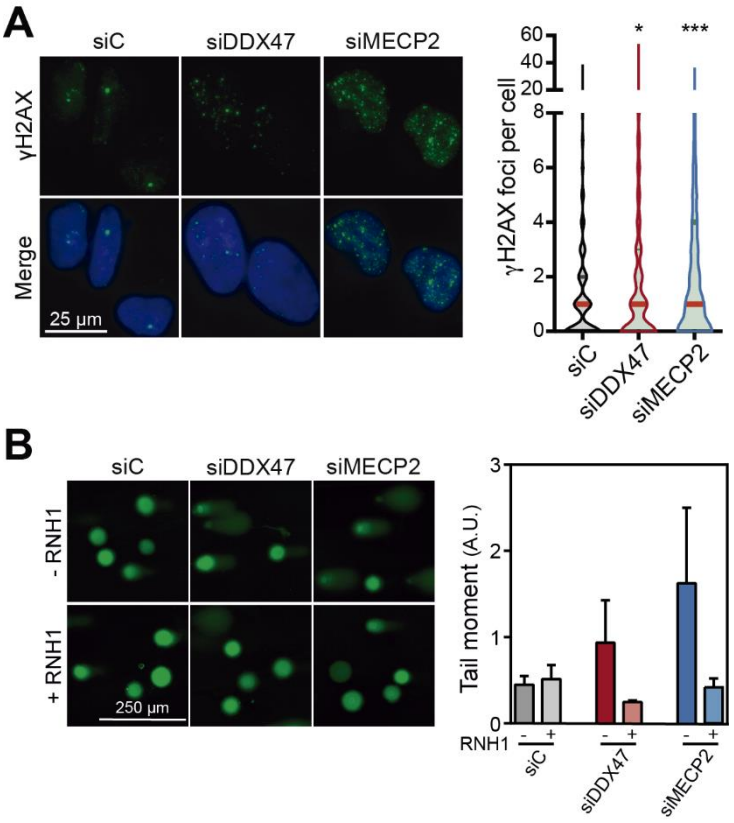


Figure R11. DNA damage analysis in siDDX47 and siMECP2 HeLa cells. (A) Representative images of immunostaining with γH2AX (green) antibody in HeLa cells upon DDX47 and MECP2 depletion. Quantification of the number of γH2AX foci per cell in HeLa cells transfected with the indicated siRNAs is represented (n=3). The red line in the violin plot indicates the median. ***P < 0.001; *P < 0.05 (Mann-Whitney U-test, two-tailed). **(B)** Representative images of single-cell alkaline gel electrophoresis (comet assay) of HeLa cells upon DDX47 and MECP2 depletion with and without RNaseH1 overexpression. Medians of comet-tail moments are plotted as mean ± SEM (n=4) (one-tailed paired t-test).

3. ROLE OF DDX47 IN R-LOOP METABOLISM AND GENOME INSTABILITY

3.1. DDX47 depletion impairs transcription

DDX47 is recruited to the nucleolus via the nucleolar protein NOP132 (Sekiguchi et al., 2006). DDX47 colocalizes with the proteins Nucleolin and Fibrillarin, commonly used as markers of granular component and fibrillar centers respectively (Figure R12A). DDX47 has been proposed to play a role in rDNA processing (Sekiguchi et al., 2006). As shown in Figure R12B, DDX47 ChIP analysis using primers spanning the highly transcribed human rDNA repeat (5'-external transcribed spacer (ETS), 18S, 28S rRNA coding regions and the intergenic spacer (IGS) region) revealed that this helicase is recruited at chromatin at ribosomal DNA locus.

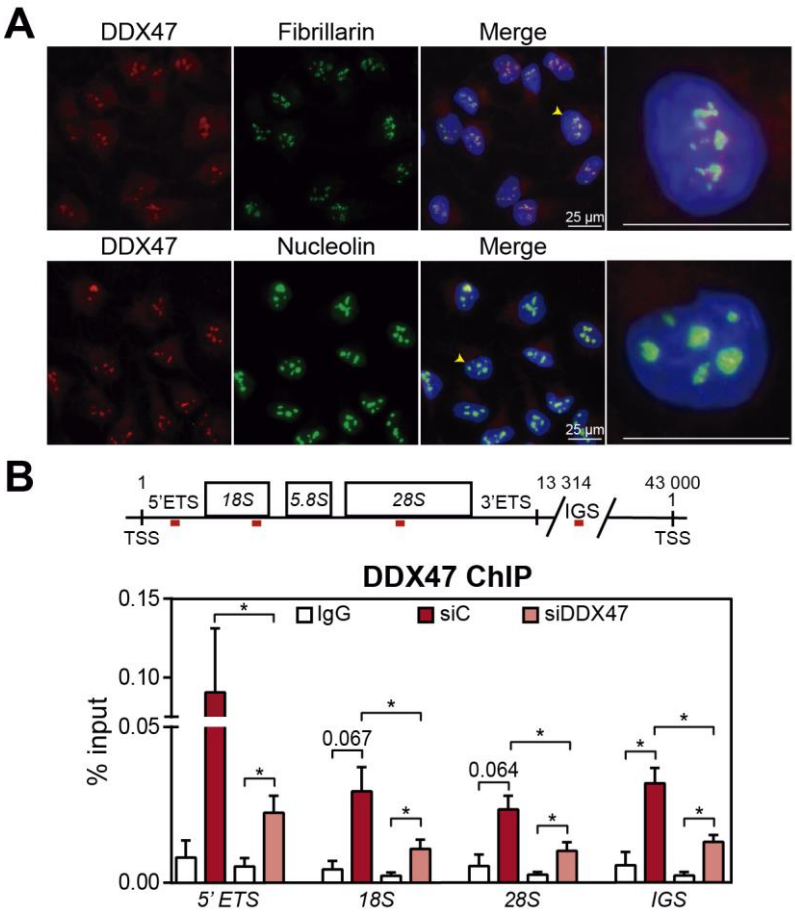


Figure R12. DDX47 colocalizes with nucleolin and fibrillarin and is recruited at rDNA. (A) Colocalization of endogenous DDX47 (red) with markers of granular component and fibrillar centers visualized by anti-fibrillarin (green, upper panel) or anti-nucleolin (green, lower panel) respectively in HeLa cells. Right panels are a magnification of the region indicated in the previous panel. Nuclei were stained with DAPI. **(B)** Schematic diagrams of 5' ETS, 18S, 28S and IGS genes. Exons are depicted as open boxes and red lines indicate the regions where the primer pairs used for amplification were located. ChIP analysis of DDX47 protein at the indicated regions. Values represent the percentage of the precipitated DNA (IP) to input DNA (INPUT). Immunoprecipitation in siDDX47 cells were included as control of specificity. IgG was used as negative control. Data are plotted as mean ± SEM (n = 3). *P < 0.05 (one-tailed paired t-test).

Interestingly, we observed a general decrease in total nucleolar area defined by nucleolin staining in immunofluorescence of siDDX47 cells (Figure R13A), suggesting defects in the homeostasis of the nucleolus in the absence of this protein. Since the primary function of the nucleoli consists in ribosomal DNA transcription we asked the possibility that DDX47 silencing could affect RNA polymerase I (RNAPI) function. Immunofluorescence analysis with antibodies against RP194, the largest subunit of RNAPI, showed a higher signal in siDDX47 compared to siC cells (Figure R13B). In addition, western blot analysis of RNAPI and UBF, a factor necessary for rDNA transcription (Panov et al., 2006), also revealed an increase in the amount of these proteins upon depletion of DDX47 (Figure R13C). To explore further the state of the RNAPI we studied its occupancy at rDNA in siRNA depleted cells. ChIP analysis using primers spanning the human rDNA repeat (5'-external transcribed spacer (ETS), 18S, 28S rRNA coding regions and the intergenic spacer (IGS) region), revealed a reduction of RNAPI from 5´ to 3´ in siDDX47 versus siC cells suggesting that DDX47 downregulation could affect the recruitment of RNAPI (Figure R13D). Interestingly, DDX47 ChIP profile at human rDNA repeat was similar to those of RNAPI except at IGS, supporting a possible role of this protein in RNAPII transcribed regions and rDNA transcription (Figure R12B and R13D).

In order to test this possibility, we performed a timecourse 5´ ethynyluridine (EU) incorporation assay followed by immunofluorescence labelling. 5-ethynyluridine is a uridine analog that is incorporated into the newly transcribed RNA *in vivo* (Jao & Salic, 2008), which allows the visualization of rDNA transcription sites by tracking the nucleolar incorporation. Results strongly indicated that DDX47 depletion significantly delayed EU incorporation into the nucleolus, especially at short times of incubation (20 min), and this phenotype was significantly maintained throughout the assayed time (Figure R14, upper panel). Altogether these results indicated that rDNA transcription was impaired upon DDX47 silencing. A reduction of EU labelling in the nucleoplasm was also detected, but mainly at long incubation times, suggesting an indirect and broad effect on global transcription in DDX47 siRNA depleted cells (Figure R14, lower panel).

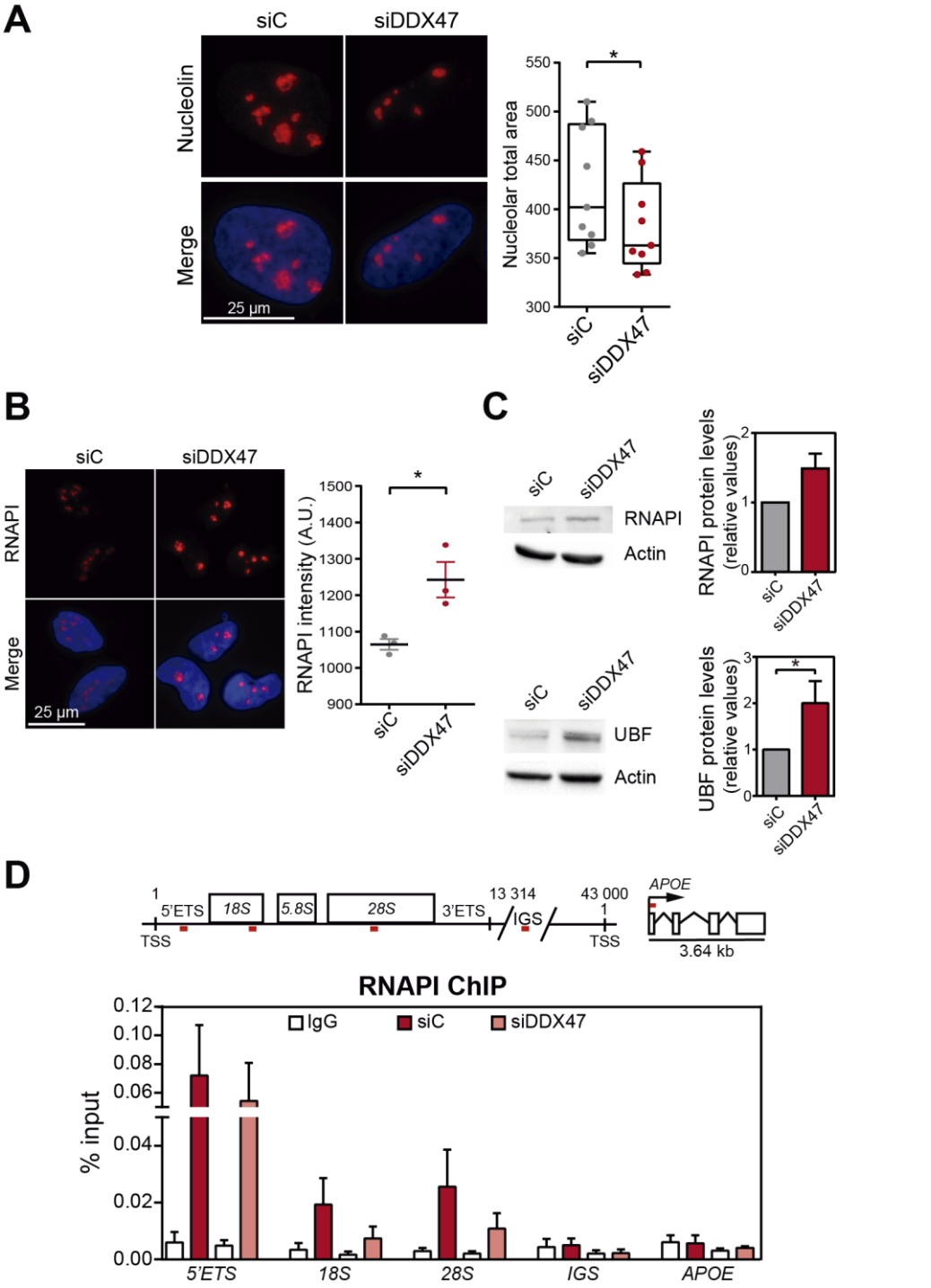


Figure R13. Analysis of nucleolus area and RNAPI in DDX47-depleted cells. (A) Representative images of immunostaining with anti-nucleolin (red) in HeLa cells upon DDX47 depletion. Quantification shows the median of nucleolar total area in siRNA transfected cells from at least three biological repeats. Boxes and whiskers indicate minimum and maximum. *P < 0.05 (two-tailed paired t-test). **(B)** Representative images of immunostaining with anti-RNAPI (red) in HeLa cells upon DDX47 depletion. Quantification of RNAPI intensity in cells transfected with siC and siDDX47 cells. Data are plotted as mean ± SEM (n = 3). *P < 0.05 (two-tailed paired t-test). **(C)** Relative RNAPI and UBF protein levels as measured by western blot in siRNA transfected HeLa cells with siC and siDDX47 after 72 hours. Actin protein is used as a loading control. **(D)** Schematic diagrams of rDNA region and APOE gene. Exons are depicted as open boxes and red lines indicate the regions for the primer pairs used for amplification. ChIP analysis of RNAPI protein at the indicated regions. Values represent the percentage of the precipitated DNA (IP) to input DNA (INPUT). IgG was used as negative control. Data are plotted as mean ± SEM (n = 3) (one-tailed paired t-test).

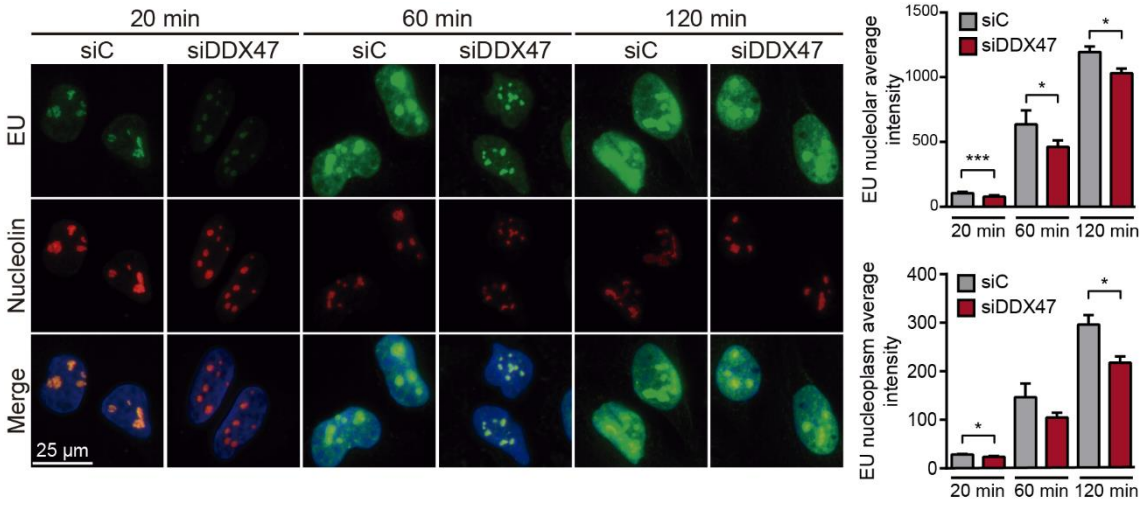


Figure R14. DDX47 depletion delays 5-ethynyluridine (EU) incorporation in the nucleolus. Representative immunofluorescence images of EU incorporation in HeLa cells transfected with siC and siDDX47 after 72 hours. Samples were collected at the indicated times of EU incubation (min). Anti-nucleolin was used as a marker of nucleolus. Quantification of EU nucleolar (upper panel) and EU nucleoplasm (lower panel) incorporation signal in siC and siDDX47 depleted cells. Data are plotted as mean ± SEM (n = 3). *P <0.05, ***P < 0.001 (one-tailed paired t-test).

3.2. Transcription-replication conflicts in DDX47-depleted cells

Given the transcription impairment and R-loop dependent genome instability in siDDX47 cells (Figure R11B), we wonder whether transcription defects associated to DDX47 depletion could compromise genome integrity in a replication-mediated manner (Gómez-González & Aguilera, 2019). Thus, we explored a possible increase in rDNA transcription-replication collision as a result of the encounters between RNAPI transcription and replication machineries in DDX47-deprived cells. For this purpose, we performed proximity ligation assay (PLA), a technique that detects protein-protein association *in situ* and used antibodies against RNAPI and antibodies against PCNA (Proliferating Cell Nuclear Antigen), as an essential protein marker of DNA replication. PLA signals of both proteins have been previously reported as an approach to measure transcription-replication collisions (Bayona-Feliu et al., 2021). As shown in Figure R15, a significant increase of PLA signal was observed in DDX47 depleted cells as compared with siC control. Thus, our results suggest that rDNA transcription-replication collision events could be a key factor in genome instability observed in DDX47-deprived cells.

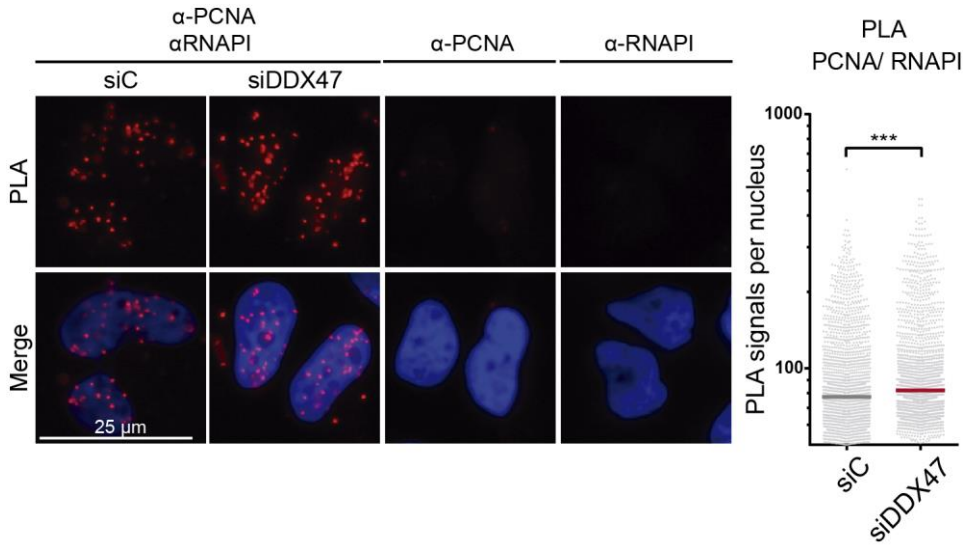


Figure R15. Analysis of conflicts between RNAPI and DNA polymerase in DDX47-depleted cells. Proximity ligation assay (PLA) showing interactions of RNAPI and PCNA endogenous proteins. Red spots are indicative of a positive PLA signal. Negative control with only one of the antibodies are also shown. DNA was stained with DAPI. The median of PLA signal intensity per nucleus in siC and siDDX47 is plotted. More than 250 total cells from at least three independent experiments is shown. ***P < 0.001 (Mann-Whitney U test, two-tailed).

3.3. DDX47 physically associates to R loops

Since R loops have been proposed to be an obstacle for transcription, and considering that these structures are increased upon DDX47 silencing we asked whether DDX47 could associate to RNA-DNA hybrids *in vivo*. For this purpose, we performed proximity ligation assay (PLA). A strong PLA signal was detected with DDX47 and S9.6 antibodies, supporting a close association between this factor and RNA-DNA hybrids in human cells (Figure R16). As expected, such association is observed mainly in the nucleolus, because DDX47 is a nucleolar protein.

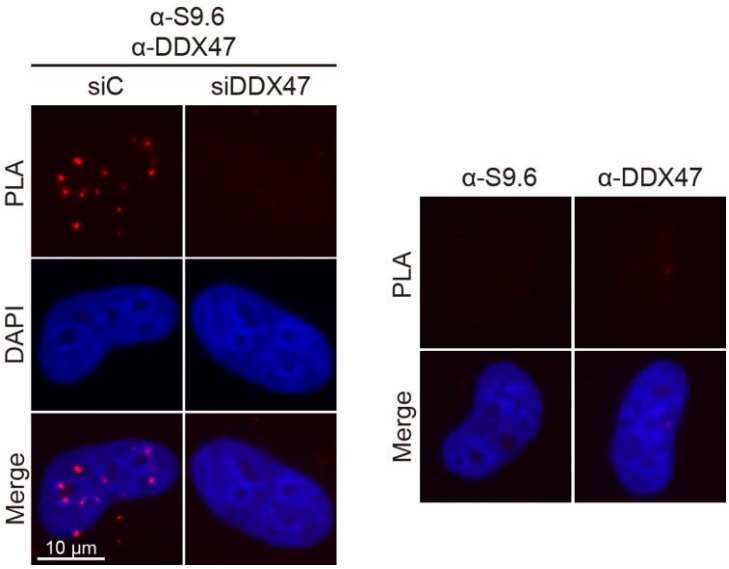


Figure R16. In situ proximity ligation assay between endogenous DDX47 and RNA-DNA hybrids. Representative images of proximity ligation assay (PLA) showing interactions of DDX47 and RNA-DNA hybrids in HeLa cells treated *in vitro* with RNase III. Red spots are indicative of a positive PLA signal. siDDX47 HeLa cells were included as control of specificity. Negative control with only one of the antibodies are also shown (right panel). DNA was stained with DAPI.

3.4. DEAD/box RNA helicase DDX47 is an RNA-DNA helicase

The DEAD/H box helicases belong to helicase superfamily 2 (SF2), which consist of two groups: 44 members in the DEAD-box group and other 15 in the DEAH-box family. Mostly, these enzymes use ATP to bind or remodel RNA and RNA-protein complexes (RNP complexes), with roles in processes including ribosome biogenesis, RNA processing and folding, and different steps of ribonucleoprotein remodelling and mRNA biogenesis. DEAD/H box proteins are built around a highly conserved helicase core of two virtually identical RecA-like domains (Helicase domain I and Helicase domain II) which contribute to the binding site for RNA substrates and ATP hydrolysis. Specifically, at least 12 characteristic sequence motifs are located at conserved position in the helicase core. The highly conserved Motif II inspire the name of the family: D-E-A-D (asp-glu-ala-asp) or D-E-A-H (asp-glu-ala-his) (Linder & Jankowsky, 2011; Schütz et al., 2010). Specifically, whereas the conserved motifs I, II, VI and the Q-motif are crucial for ATP binding and hydrolysis, the motif Ia, Ib, IV, IVa and V are important to RNA binding. Conserved motifs III and Va are essential for communication between ATP-binding and RNA-binding sites (Cordin et al., 2006; Linder & Jankowsky, 2011; Schütz et al., 2010).

We compared the amino acid sequence of DDX47 protein (Human) with its ortholog in *Saccharomyces cerevisiae* Rrp3, the putative ortholog in *Drosophila Melanogaster* DmeI and *Oriza sativa* TOGR1 using the Clustal Omega web service (McWilliam et al., 2013) and the BLAST analysis (Basic Local Alignment Search Tool). These proteins have a percent identity with DDX47 about 56.72% (*S. cerevisiae*), 72.75% (*D. melanogaster*) and 56.94% (*O. sativa*) and they share 12 conserved domains in the helicase core as determined and highlighted with black boxes in Figure R17A. This data support that DDX47 is a conserved helicase from yeast to humans.

Since siDDX47 cells show an accumulation of RNA-DNA hybrids (Figure R9 and R10), we were interested in to know whether this conserved DEAD-box RNA helicase DDX47 possesses RNA-DNA unwinding activity.

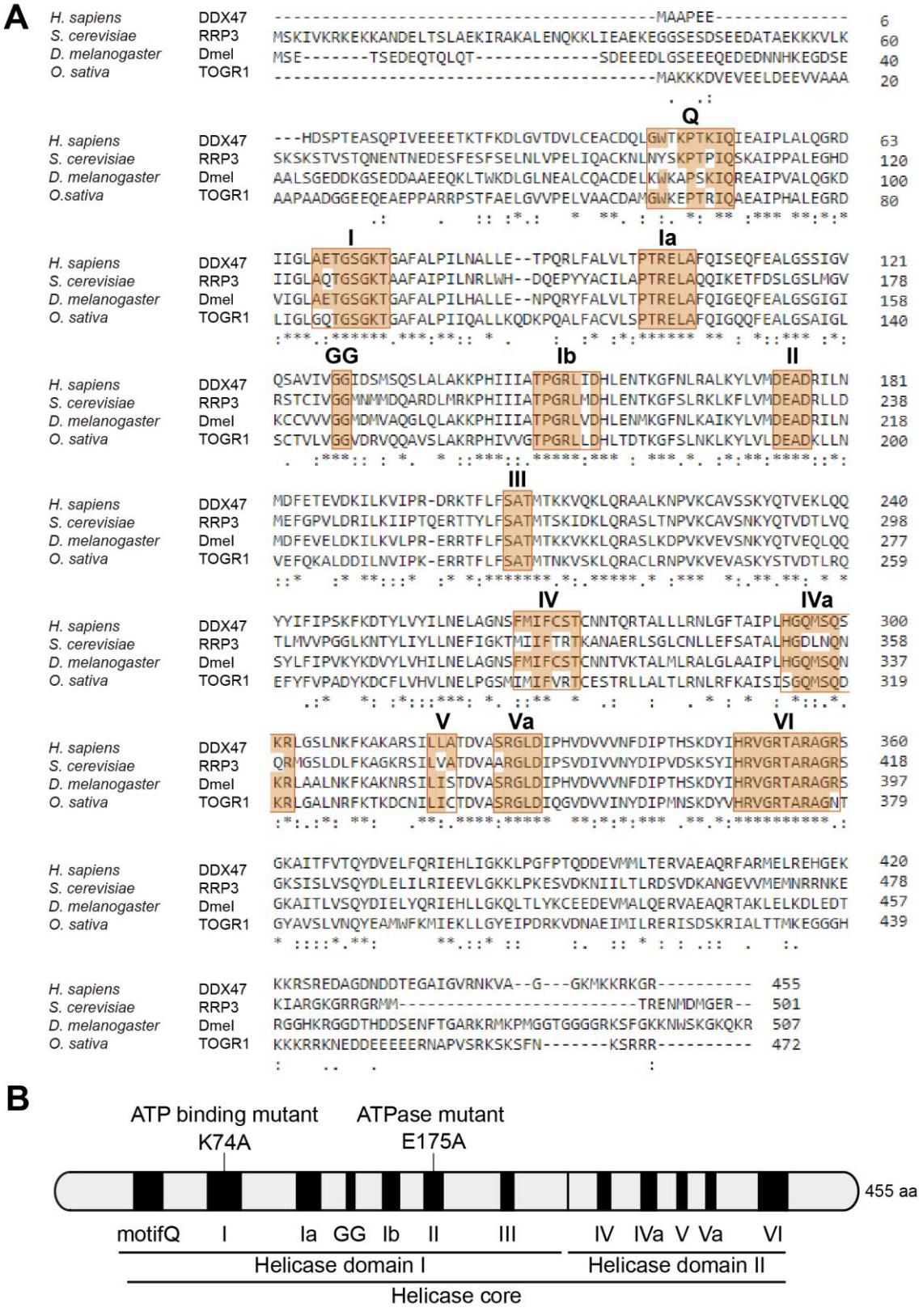


Figure R17. DDX47 conserved domains and mutations in the helicase core analyzed in this study. (A) Multiple sequence alignment of *H. sapiens* DDX47 (NP_057439.2), *S. cerevisiae* RRP3 (NP_011932.2), *D. melanogaster* Dmel (NP_610090.1) and O. sativa TOGR1 (A2XKG2.1) was conducted using Clustal Omega web service (McWilliam et., 2013). Conserved domains are highlighted with orange boxes. (*) conserved residues (:) conservation of aas with strongly similar properties (.) or aas with weakly similar properties. **(B)** Schematic representation of mutations in the helicase core of DDX47 analyzed in this study.

3.4.1. Generation of DDX47 mutants for functional *in vitro* analysis

We carried out site-direct mutagenesis to generate mutants in the conserved motif I (DDX47-K74A) and in the motif II (DDX47-E175A) involved in ATP binding and ATP hydrolysis respectively in other helicases (Cordin et al., 2006; Linder & Jankowsky, 2011). For a schematic representation of mutations in the helicase core of DDX47, see Figure R17B. The desired mutations were introduced into a plasmid containing the full DDX47-WT open reading frame sequence (see Materials and Methods 7). Then, DDX47-WT and DDX47 mutant cDNAs were cloned into the vector pET300/NT-DEST that allows high-level prokaryotic expression controlled by the strong bacteriophage T7 promoter and N-terminal 6xHis tag for protein purification (Figure R18, upper scheme; Table M4; Materials and Methods 8.1). Once we checked the induction of the protein at different conditions as determined by coomassie staining and western blot (Figure R18), a scientific collaboration was established with the Professor Xue Xiaoyu in Yale University to perform the biochemical analysis.

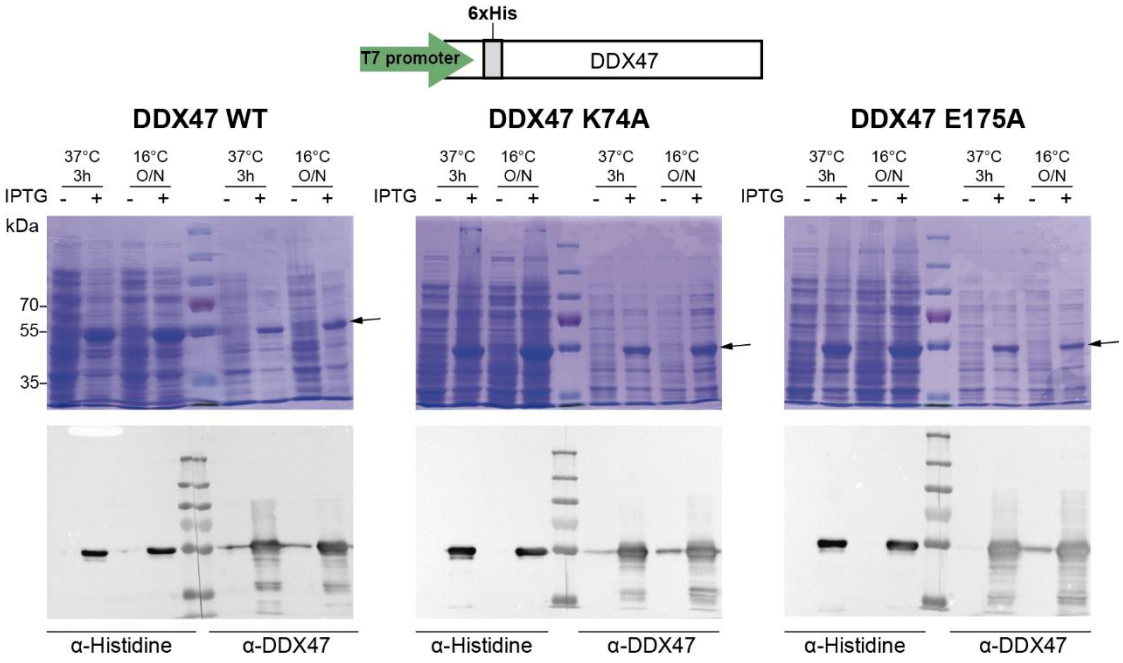


Figure R18. Development of an *in vitro* tool for DDX47-WT and DDX47-mutant protein purification. Schematic representation of vector pET300/NT-DEST containing the full DDX47 cDNA with a N-terminal 6xHis tag controlled by T7 promoter (Table M4). His6-DDX47 expression was induced by adding IPTG to the cultured at the indicated conditions. For protein staining, acrylamide gels were incubated with Coomassie Brilliant Blue. The arrows indicate the band at the DDX47 expected molecular weight (50 kDa). Western blot analysis with anti-Histidine and anti-DDX47 antibodies was performed to validate the in vitro tool (Material and Methods 15).

3.4.2. *In vitro* unwinding analysis with DDX47 proteins

DDX47-WT, and the two mutants DDX47-K74A and DDX47-E175A were purified to homogeneity (Figure R19A). RNA-RNA unwinding activity was tested using a blunt-ended RNA duplex (dsRNA) of 13 base pairs and different amounts of DDX47 proteins. The results indicated that DDX47-WT unwinds this substrate in an ATP and protein concentration-dependent manner (Figure R19B) as well as other dsRNA substrates with either a 5' or 3' overhangs (Figure R19C). In parallel, blunt-ended and either a 5' or 3' overhangs dsRNA unwinding activity of DDX47 mutants were tested. As can be seen in Figures R19B and R19C, DDX47 mutants unwind dsRNA substrates with a lower efficiency to those of DDX47WT.

Next we tested unwinding of RNA-DNA hybrid substrates. DDX47 was able to unwind RNA-DNA hybrids in a protein concentration and ATP-dependent manner. Indeed, our data suggested that DDX47 could have a preference for RNA-DNA hybrids, since the percentage of unwound hybrid product was up to fourfold of that obtained for dsRNA (Figure R20A). Therefore, DDX47 is more adept at unwinding RNA-DNA hybrids than dsRNA. To study the molecular identity of RNA-DNA unwinding activity, DDX47-K74A and DDX47-E175A mutants were tested indicating that an intact DDX47 is crucial for maintaining full RNA-DNA unwinding activity (Figure R20B). These mutants were not able to unwind blunt RNA-DNA hybrids (Figure R20B), but a basal unwinding activity was detected when 5' overhang RNA-DNA hybrids were assayed (Figure R20C).

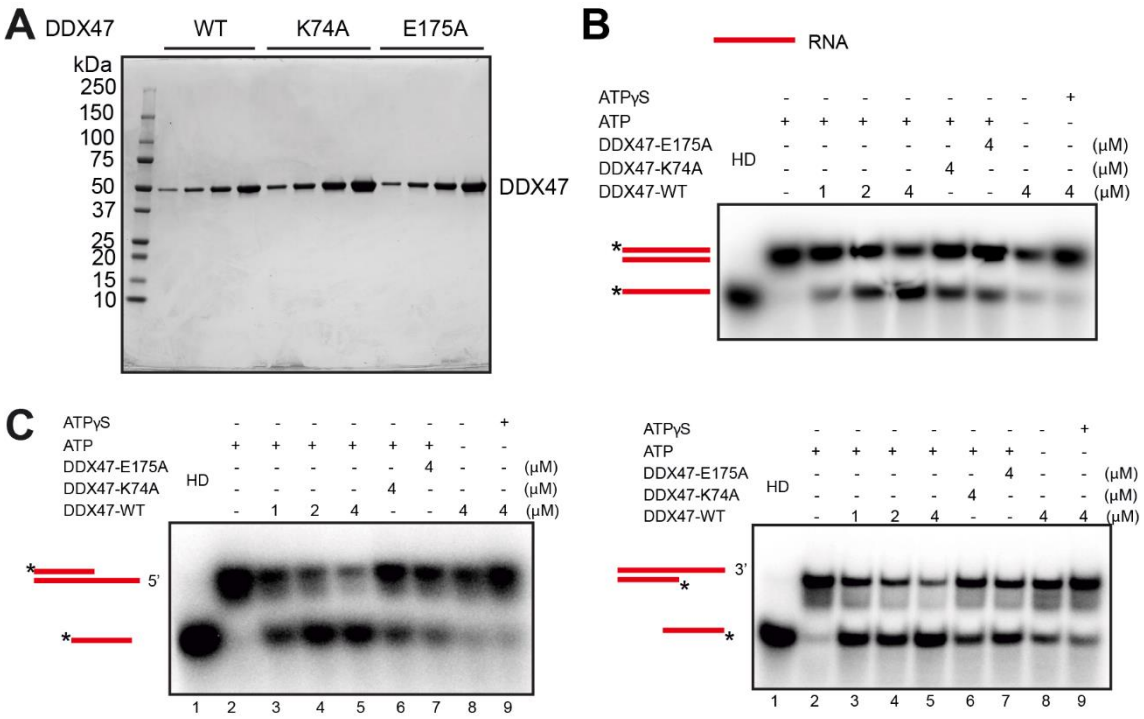


Figure R19. DDX47 RNA helicase in vitro analysis. (A) DDX47-WT and DDX47 mutants protein purification. Purified wild type and DDX47 mutants were analyzed on SDS-polyacrilamide gel and stained with Coomassie Blue. First four lines representing DDX47-WT, the next four lines DDX47-K74A mutant and last line DDX47-E175A mutant. **(B)** DDX47 RNA-RNA unwinding activity. RNA-unwinding assay was performed using a blunt-ended RNA duplex (dsRNA) as substrate with different amounts of DDX47-WT, DDX47-K74A and DDX47-E175A. **(C)** DDX47 RNA-RNA unwinding activity. RNA-unwinding assay was performed using a dsRNA with a 5' or 3' overhang as substrates with different amounts of DDX47-WT, DDX47 K74A and DDX47-E175A. The positions of duplex substrate and unwound products are indicated at the left, where the stars show the position of the radiolabel. Gels were dried and subject to phosphorimaging analysis. Performed in collaboration with Xue Xiaoyu's group.

Finally, DDX47 activity was tested with a 5' RNA-DNA flap structure that resembles a branch migratable R-loop (Schwab et al., 2015). As shown in Figure R20D, DDX47-WT protein could dissociate the flap structure to yield a dsDNA product. DDX47 mutants showed some residual unwinding activity of these structures. Therefore, our results indicate that DDX47 resolves RNA-DNA hybrids as well as R-loop mimicking structures, and that the mutants DDX47-K74A and DDX47-E175A showed a residual activity, so it is possible that other residues could contribute to this activity.

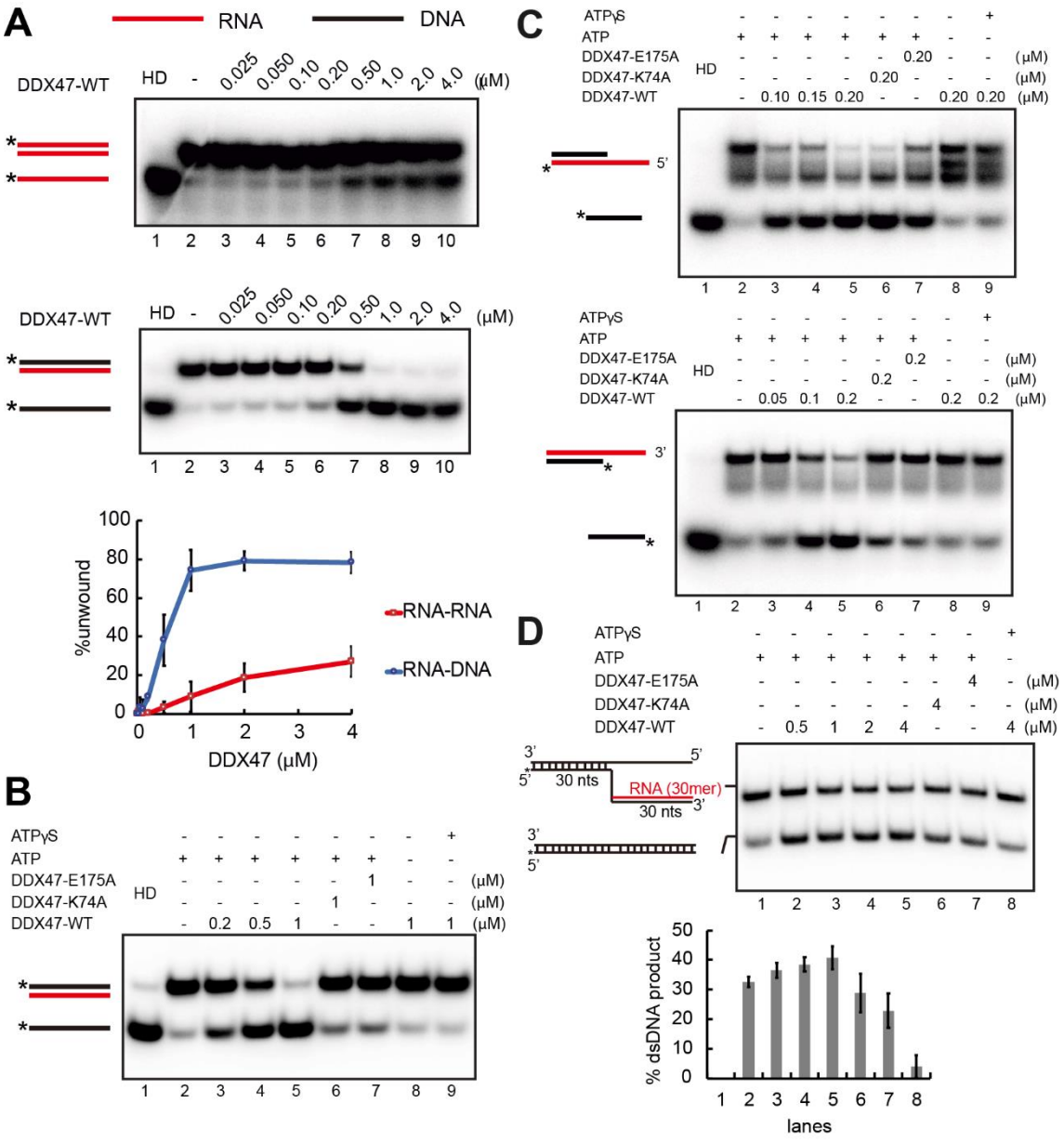


Figure R20. DDX47 RNA-DNA helicase in vitro analysis. (A) Comparison between DDX47 RNA helicase and RNA-DNA helicase activity using same amount of dsRNA or RNA-DNA duplex with a serial dilution of DDX47-WT protein. Graph shows the percentage of unwound product respect to the DDX47 concentration-dependent manner. **(B)** DDX47 RNA-DNA unwinding activity. RNA-DNA unwinding assay was performed using a blunt-ended RNA-DNA duplex as substrate with different amounts of DDX47-WT, DDX47-K74A and DDX47-E175A. **(C)** DDX47 RNA-DNA unwinding activity. RNA-DNA unwinding assay was performed using a RNA-DNA with a 5' or 3' overhang as substrates with different amounts of DDX47-WT, DDX47 K74A and DDX47-E175A. The positions of duplex substrate and unwound products are indicated at the left, where the stars show the position of the radiolabel. Gels were dried and subject to phosphorimaging analysis. **(D)** DDX47 unwinds RNA-DNA flap structures mimicking R-loops. RNA-DNA unwinding assay with DDX47-WT, DDX47-K74A and DDX47-E175A using RNA-DNA flap structures mimicking R-loops as substrates. Other details as in (C). Graph shows the percentage of dsDNA product recovered after the reaction. Performed in collaboration with Xue Xiaoyu's group.

3.4.3. DDX47 *in vivo* activity

Concerning the mechanism in which DEAD-box proteins unwind duplexes (Yang et al., 2007) could be expected that any RNA helicase could have RNA-DNA unwinding activity *in vitro*, so we wondered whether DDX47 shows RNA-DNA hybrids unwinding activity *in vivo*.

We expressed DDX47-WT and DDX47-K74A and DDX47-E175A under the control of the strong promoter CMV (Figure R21A). To confirm DDX47 overexpression we performed immunofluorescence assays with anti- DDX47 and anti-Flag antibodies to detect endogenous and tagged-DDX47 proteins respectively (Figure R21A). A strong nuclear signal was detected in cells transfected with DDX47-Flag proteins, whereas a weaker and nucleolar signal was detected in cells transfect with the empty plasmid (endogenous DDX47). Overexpression was also confirmed by western blot analyses (Figure R21B).

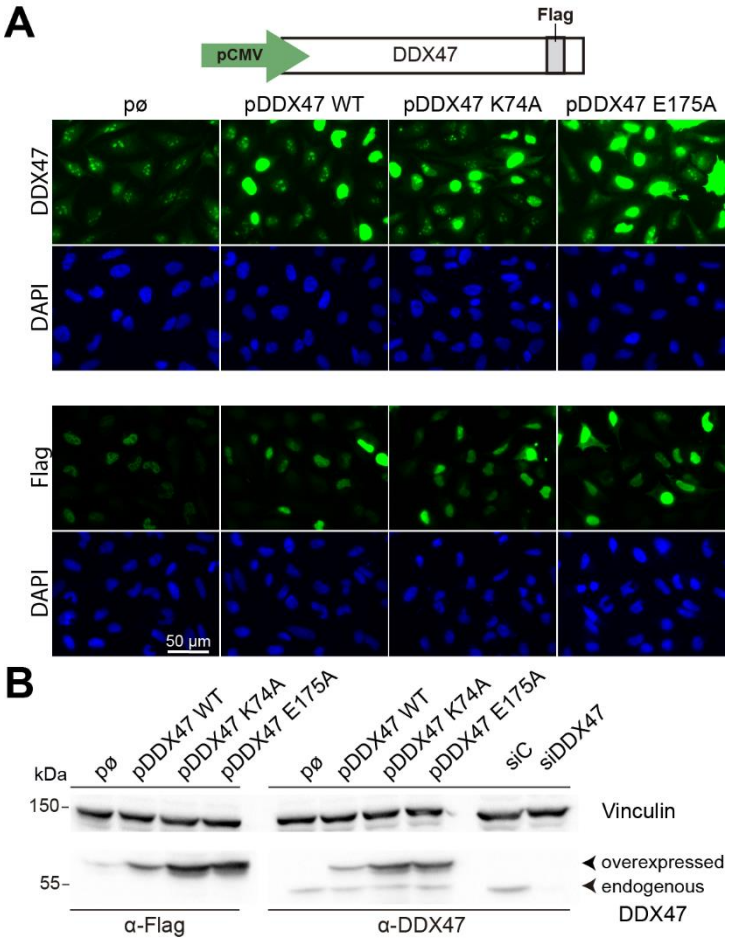


Figure R21. Tools for in vivo DDX47 overexpression. (A) Schematic representation of expression vector containing full-length human DDX47-WT or DDX47 mutants with C-terminal Myc-DDK-tag under the control of CMV promoter. Representative images of immunostaining with anti-FLAG (green) in Hela cells transfected with the empty plasmid, pDDX47-WT, pDDX47 K74A and pDDX47 E175A after 24 h of overexpression. Nuclei was stained with DAPI. **(B)** Western blot analysis with anti-DDX47 and anti-Flag of HeLa cells transfected with the indicated plasmids. siC and siDDX47 cells were included as control of endogenous DDX47.Vinculin protein is used as a loading control.

Then, we assayed whether overexpression of WT and helicase-dead mutants of DDX47 could supress or not the R-loop accumulation phenotype caused by DDX47 depletion itself. For overexpression experiments, we used a different pool of two siRNAs against the 3' UTR, just to ensure DDX47 genomic silencing but not those of DDX47 expressed from the plasmid. A similar silencing was achieved with the 3' UTR siRNAs, as determined by western blot (Figure R22A) and importantly these siRNAs conferred a RNA-DNA hybrid accumulation similar to those of siRNA pool (Figure R22B) without affecting transcription levels (Figure R22C).

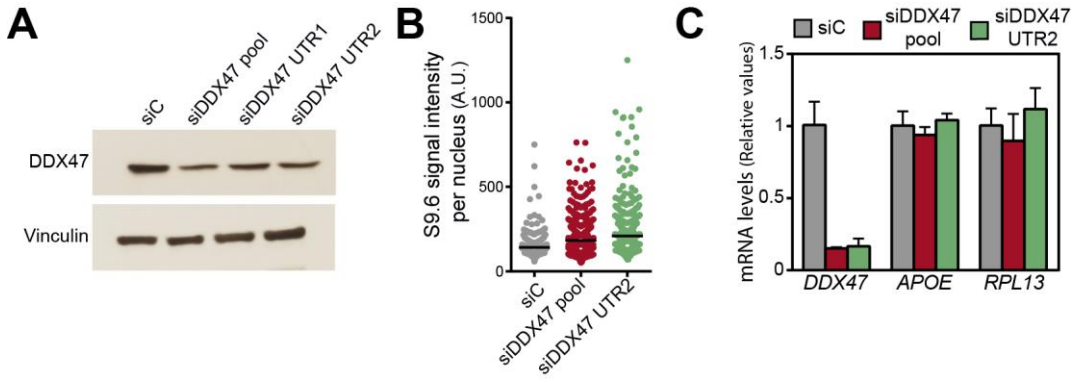


Figure R22. Tools for specific siRNA depletion and overexpression of DDX47. (A) Western blot analysis with anti-DDX47 of siRNA-treated HeLa cells with siC, siDDX47 or different pools of two siRNAs against the 3'UTR after 72 h. **(B)** The median of S9.6 signal intensity per nucleus after nucleolar signal removal in the indicated siRNA-treated HeLa cells is plotted. At least 200 cells were scored. A.U., Arbitrary units. **(C)** mRNA levels of DDX47, APOE and RPL13 in the indicated siRNA-treated cells as determined by RT-qPCR. Error bars represent relative target quantity (RQ) minimum and maximum from three technical replicates.

We measured R-loop accumulation upon DDX47 overexpression by S9.6 IF assays. To measure exclusively R-loop accumulation in nucleoplasm and nucleolus, S9.6 IF experiments were carried out including an *in vitro* treatment with RNase III, which degrades dsRNA structures (Silva et al., 2018) and nucleolin staining was used to determine nucleolar area. First, a significant enrichment of the antibody S9.6 nuclear signal was observed in siDDX47 cells in both analysis, confirming the RNA-DNA hybrids accumulation phenotype in the nucleoplasm as well as in the nucleolus upon depletion of the helicase DDX47. Concerning the antibody S9.6 nucleoplasm signal, DDX47-WT overexpression rescued the RNA-DNA hybrid accumulation associated to DDX47 depletion (Figure R23). Furthermore, in the nucleoplasm (left panel) the two helicases mutant forms reduced the R-loop accumulation in siDDX47 depleted cells,

although to a lesser extent than DDX47-WT protein. These results could be explained by the residual unwinding activity of mutants on hybrids structures (Figure R20C and R20D). Nevertheless, we cannot discard that these mutant proteins could stick to the hybrids reducing the accessibility of S9.6 antibody to these structures and the immunofluorescence signal as can be observed not only in siDDX47 cells, but also in siC cells.

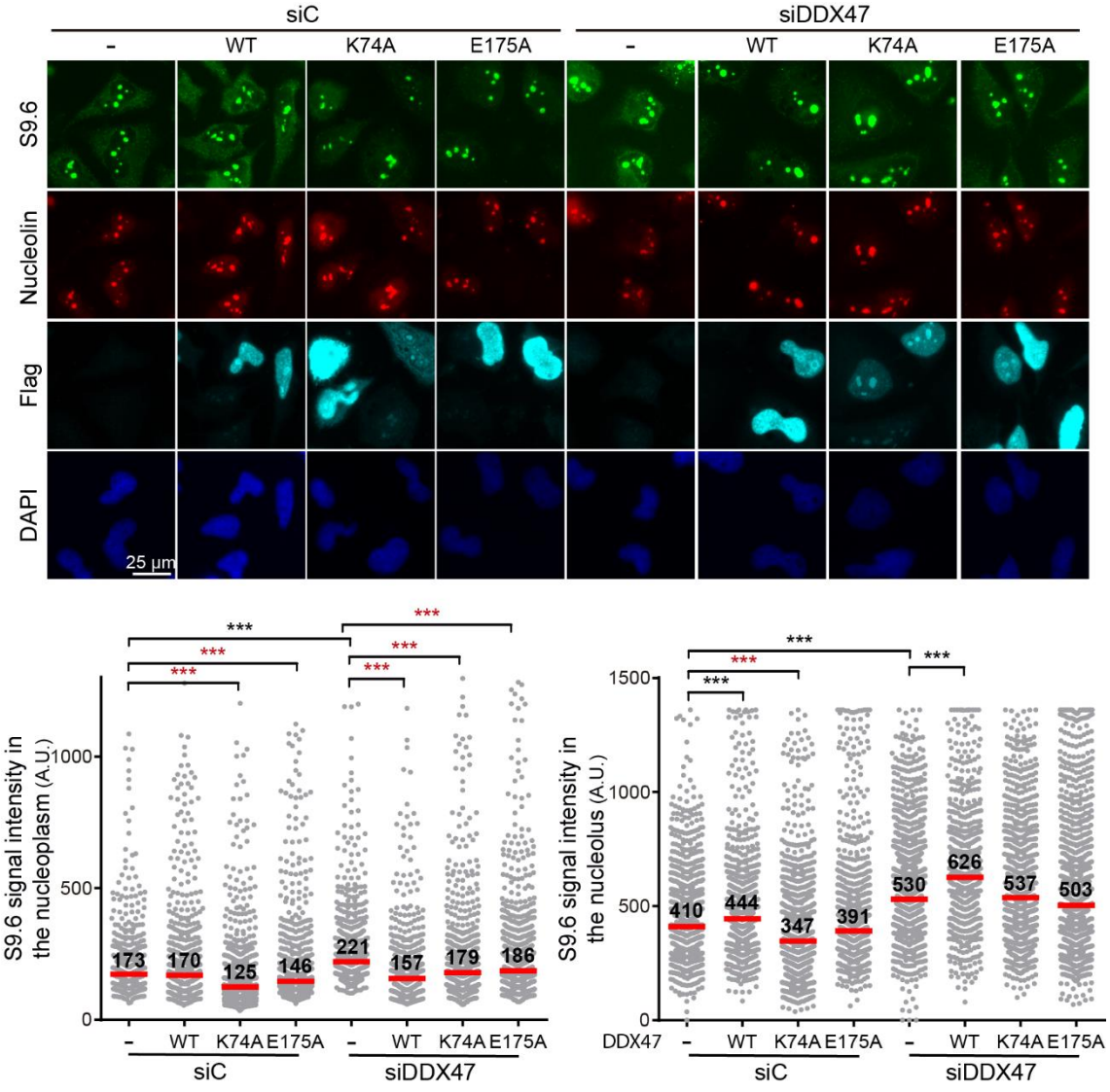


Figure R23. DDX47 overexpression effects in RNA-DNA hybrids. Representative images and quantification of S9.6 immunofluorescence signal in HeLa transfected cells with siDDX47 and with pFLAG (-), pFLAG-DDX47 (+DDX47) for DDX47-WT overexpression, and pFLAG-DDX47-K74A (+DDX47-K94A) or pFLAG-DDX47-E175A (+DDX47-E175A) for helicase-dead mutant DDX47 overexpression (blue). All samples were treated with RNase III. The graph shows the median of S9.6 signal intensity in the nucleoplasm after nucleolar signal removal (left panel) or the S9.6 signal in the nucleolus (right panel) (n=4). More than 200 cells per condition were counted in each experiment. ***P < 0.001 (Mann-Whitney U test, two-tailed). A.U., Arbitrary units. Black stars denote significant increases, whereas red stars denote significant decreases.

Concerning the S9.6 signal intensity in the nucleolus, DDX47-WT overexpression unlike mutant forms caused a significant increase in R-loop accumulation in siC control and siDDX47 cells. It is possible that an imbalance in the amount of this helicase could cause a negative effect, maybe leading to transcription impairment mainly at the nucleolus.

Then, we overexpressed DDX47 in cells that accumulate high levels of RNA-DNA hybrids by different mechanisms, such as depletion of RNA helicases, DDX23, SENTX, UAP56 or Fanconi Anemia repair factor FANCD2 (M. L. García-Rubio et al., 2015; Pérez-Calero et al., 2020; Skourti-Stathaki et al., 2011; Sreerama Chaitanya Sridhara et al., 2017). We measured R-loop accumulation via S9.6 IF immunofluorescence in cells transfected with an empty plasmid and the DDX47-WT plasmid. S9.6 IF assays show that DDX47 rescued RNA-DNA hybrid accumulation associated to depletion of FANCD2 and SENTX and DDX23 helicases, but not that caused by UAP56 depletion (Figure R24). Indeed, we observed that the viability of siUAP56 was severely affected upon overexpression of DDX47, suggesting a negative genetic interaction. We reasoned that probably DDX47, as RNA helicase with affinity for RNA and RNA-DNA hybrid, could bind to these structures and compete with UAP56 binding interfering with the function of this key helicase. In summary, DDX47-WT overexpression is able to supress *in vivo* R-loop accumulation in different mutants.

Taken together, our results suggest the highly conserved DDX47 helicase plays an important role in unwinding *in vitro* and resolving *in vivo* RNA-DNA hybrids and is necessary to prevent R-loop dependent genome instability.

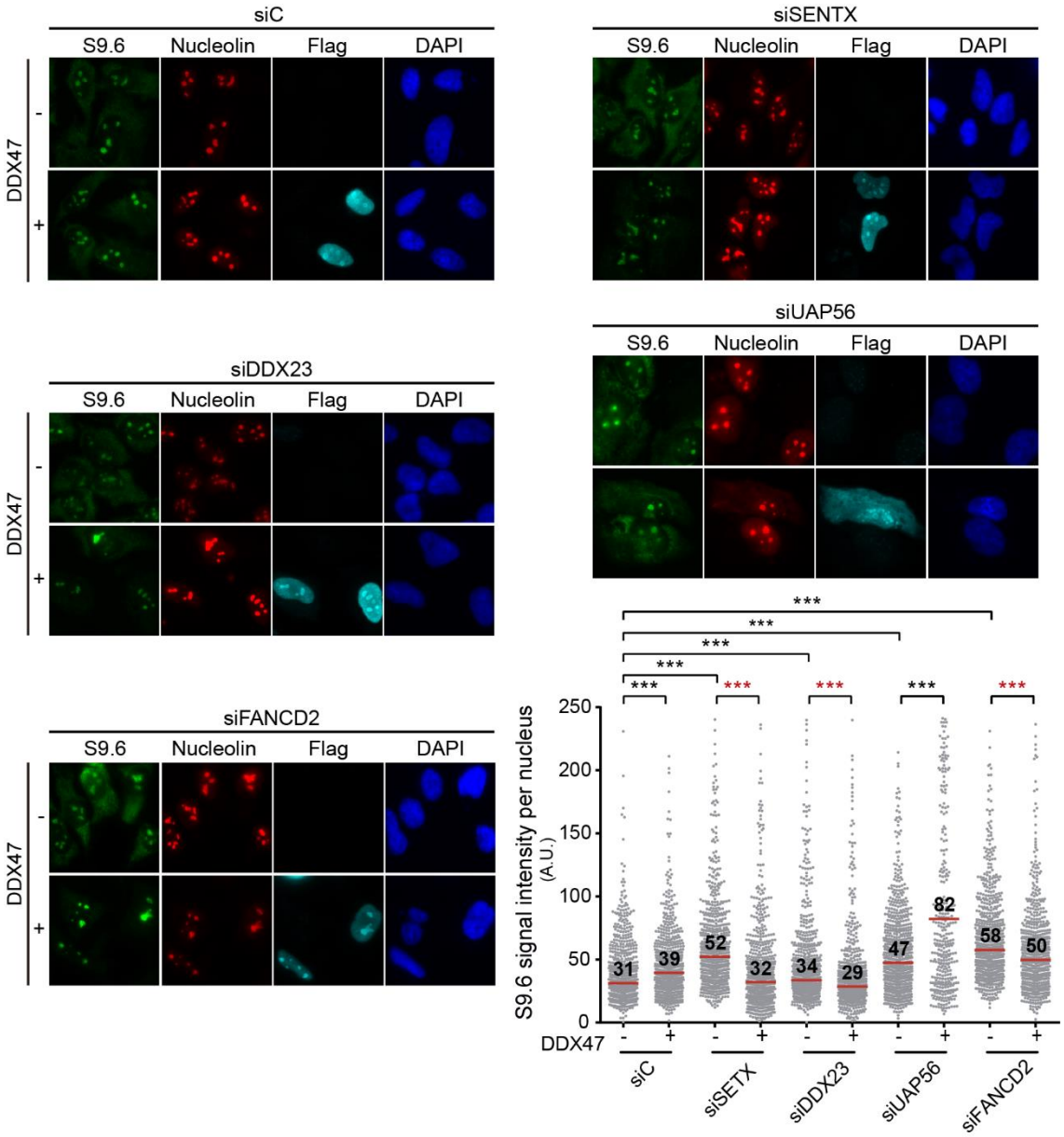


Figure R24. Analysis of R loop accumulation in SETX-, DDX23-, UAP56-, FANCD2-depleted cells upon DDX47 overexpression. Representative images and quantification of S9.6 immunofluorescence signal in HeLa transfected cells with siC, siSETX, siDDX23, siUAP56 and siFANCD2 and with pFLAG (-), pFLAG-DDX47 (+DDX47) for DDX47-WT overexpression (blue). The graph shows the median of S9.6 signal intensity in the nucleoplasm after nucleolar signal removal (n=3). More than 200 cells per condition was counted in each experiment. ***P < 0.001 (Mann-Whitney U test, two-tailed). A.U., Arbitrary units. Black stars denote significant increases, whereas red stars denote significant decreases.

DISCUSSION

In this thesis we have identified new factors related to R loop-mediated genome instability as a result of an siRNA high-throughput microscopy screening with the AIRD system (U2OS-TR-AID cells). Such genes are functionally related with chromatin remodelling and modification, transcription and RNA biogenesis, protein degradation, cell cycle control, and other cellular processes. Further analyses were performed in 7 selected genes to validate the putative role of these factors on R loop metabolism. We decided to focus on the two top hits: the DNA binding protein MeCP2 (methyl-CpG binding protein 2) the highest reproducible candidate, and the nucleolar RNA helicase DDX47, the top-hit with the highest R-loop levels. MeCP2 and DDX47 have been validated as factors involved in R loop homeostasis and required to avoid R-loop dependent genome instability, as determined by different assays in different cell lines (S9.6 IF, DRIP, γH2AX IF, comet assay). We have further characterized the role of DDX47 in RNA-DNA hybrid homeostasis and genome stability. We show that this helicase is recruited to chromatin at ribosomal DNA locus and RNAPII transcribed regions. DDX47 downregulation causes a reduction of nucleolar total area, impairs rDNA transcription, reduces slightly the occupancy of RNA polymerase I at rDNA locus and increases transcription-replication collisions. Importantly, we provided evidence that the highly conserved helicase DDX47 has a novel function as an RNA-DNA helicase and R loop resolvase, supported by both *in vitro* and *in vivo* analysis.

1. A NEW SCREENING TO INDENTIFY NEW FACTORS RELATED TO R-LOOP MEDIATED GENOME INSTABILITY

Given the relevance of R loop on the maintenance of cell homeostasis and genome integrity, different screenings have been performed to identify genes related to R-loop metabolism. Some of these works are based on genetic analysis to study the impact of RNA-DNA hybrids on different marks of DNA damage or chromosomal instability, such γH2AX foci, chromosomal rearrangements (GCR) or Rad52 foci in yeast and human cells (Paulsen et al., 2009; Stirling et al., 2012; Wahba et al., 2011). In some of these studies, RNAse H1 treatment was used as a tool to determine whether the genetic-instability observed phenotypes were DNA-RNA hybrid dependent. Other studies are based on the use of the antibody

S9.6, that recognizes RNA-DNA hybrids (Boguslawski et al., 1986), via cytological screenings or by biochemical approaches (Chan et al., 2014; Cristini et al., 2018; Nadel et al., 2015; Wang et al., 2018). Immunoprecipitation with S9.6 antibody, or pull-down with oligonucleotides containing R-loop structures, coupled to mass spectrometry, have led to the definition of a list of proteins that can interact with RNA-DNA hybrids (Cristini et al., 2018; Nadel et al., 2015; Wang et al., 2018). As a result, different RBP, RNA processing, DNA binding proteins and chromatin modification factors have been identified. Nevertheless, whether these factors have or do not have a direct role preventing R-loop accumulation is still unclear and precise a deeper analysis.

The screening described in this thesis is based on the action of AID on ssDNA, such as the displaced DNA strand of an R loop, deaminating Cs as a first step for the subsequent evolution into DNA double strand breaks (DSBs) catalyzed by other enzymes (Aguilera & Gómez-González, 2008; Basu et al., 2011). We have selected genes whose knock-down in combination with AID induction leads to an increase in DNA damage. Using this criteria, hits linked to both R loop homeostasis and genome instability are expected to be enriched. This is a difference with other screenings for R loop, as for example the RNA-DNA hybrid interactome (Cristini et al., 2018) that results in proteins associated to these structures independently of their putative function on the maintenance of genome integrity.

We identified factors previously implicated in R loop biology, including the helicase DHX9 that has been shown to prevent R loop-dependent DNA damage and is also required for transcription termination (Cristini et al., 2018); as well as TOPBP1, a protein required for DNA repair and checkpoint control identified in a screening for DDR factors involved in R loop homeostasis (Bagge et al., 2021; Barroso et al., 2019) providing confidence in our results. Specifically, in our screening with AID overexpression we found 46 hits related to processes such as RNA processing, regulation of RNA stability, ribonucleoprotein complex biogenesis, translational initiation cell cycle, and nucleobase-containing compound transport among others (Table R2).

Concerning the transcription process, some chromatin-related factors : MECP2 (methyl-CpG-binding protein 2) (that will be discussed below); HDAC8

(histone deacetylase 8) and CHRAC1 (chromatin accessibility complex protein 1) (Table R2), were identified in agreement with the nascent connection between chromatin changes and R loop homeostasis (Chédin, 2016). We have validated HDAC8 and MECP2 candidates with the use of S9.6 antibody (Figure R6). HDAC8 is a histone deacetylase necessary for cohesin recycling during the cell cycle and mutations in these gene are associated to Cornelia de Lange syndrome (Deardorff et al., 2012; Kaiser et al., 2014). In addition to control the acetylation of this and other protein substrates, HDAC8 is a class I HDAC that deacetylates histone H3 and H4 at nonspecific lysines, and recently has been shown to be a novel cofactor of SMAD3/4 complex playing a role in transcriptionally suppression (Tang et al., 2020). Thus, the increase in R loop in siHDAC8 cells could be related with a more relaxed and transcribed chromatin. Indeed, high chromatin accessibility caused by histone hyper-acetylation have been proposed to facilitate R loop generation. Depletion of the histone deacetylase mSin3a complex, both in yeast and human cells, or deacetylation inhibition by chemical compounds (TSA and SAGA) causes R loop accumulation and genome instability (Salas-Armenteros et al., 2017; Wahba et al., 2011). CHRAC1, another candidate of our screening, is a subunit of a complex that interacts with the remodelling factor ACF to facilitate nucleosome sliding (Eberharter et al., 2001). In the same line, the FACT complex (facilitates chromatin transcription) that works in transcription elongation and chromatin remodelling, has a role in the prevention of R-loop-mediated transcription-replication conflicts, likely associated with a specific chromatin organization (Herrera-Moyano et al., 2014). Moreover, recent results have opened new insights to understand the role of chromatin remodelers in the control of R loop homeostasis and its impact on genome integrity (Bayona-Feliu et al., 2021; Prendergast et al., 2020). Further analysis will be needed to know whether CHRAC1 plays a similar role in R loop and genome instability.

We found that some genes involved in transcription and RNA biogenesis as well as translation caused high levels of R loop dependent DNA damage highlighting transcription factors as MYOG, PHOX2A, ARNTL2 and NR2E3; some helicases as DDX42, DDX47 and DHX9; and translation initiation and elongation factors as EIF3A, EIF3C and EEF1D. The identification of transcription and RNA processing related factors is consistent with the importance of mRNP

biogenesis factors in the prevention of R loop accumulation (Gavaldá et al., 2013; Huertas & Aguilera, 2003; X. Li & Manley, 2005; Paulsen et al., 2009; Stirling et al., 2012). Interestingly, cells depleted of genes involved in protein degradation (PSMC5, PSMD7, STAMBP, SPINK5) exhibited high level of DNA damage upon AID induction (Ferdous et al., 2001; Geng et al., 2012). Accordingly, other proteasome subunits were identified in the RNA-DNA hybrid interactome (Cristini et al., 2018). Nevertheless, since the proteasome participates in numerous cellular processes, including transcription regulation, cell cycle progression, apoptosis, or DNA damage repair it is not clear whether these proteins could have a direct role in R loop homeostasis.

Finally, it should be notice that some factors previously related with R-loop genome instability were not identified by the AIRD system, as it would be expected. This could be due to high basal levels of DNA damage in the absence of these factors that did not allow to detect a significant increase upon AID induction; the double selection criteria for the screening; cell death associated to know-down of some factors; or technical issues derived from the high-throughput analysis protocol.

Our collection of siRNAs includes targeting 3205 human genes that locate both in nucleus and cytoplasm and are involved in a wide variety of processes, as apoptosis, nucleic acid binding, autophagy and others (Appendix 2). Some of the putative candidates and the process in which they participate have not been previously linked to R-loop dependent DNA damage, such as lipid metabolism, transporters, G-protein related and signalling, metabolic pathways among others. In spite to confirm these candidates, our data suggest a wide range of factors that could contribute may be indirectly to R-loop dependent DNA damage. Further analysis will be required to define the molecular mechanisms and biological meaning behind these factors.

In this thesis, we decided to focus our study in two candidates that showed different activities or properties on the nucleic acid metabolism: the methyl-CpG binding protein MECP2 and the helicase DDX47.

2. ROLE OF MECP2 IN THE MAINTENANCE OF GENOME INTEGRITY

Mutations in methyl-CpG binding protein 2 (MECP2) are the major cause of Rett syndrome which is a progressive neurologic developmental disorder that primarily affects females, since it is X-linked and subject to X inactivation (Amir et al., 1999; Good et al., 2021; Ip et al., 2018). MECP2 encodes a protein that binds DNA methylated and is highly expressed in neurons (near histone-levels) (Lewis et al., 1992; Meehan et al., 1992; Skene et al., 2010). Genome-wide analysis indicate that MECP2 binds widely across the genome and is enriched at mCG and mCA sites (Lin Chen et al., 2015). Classically MECP2 was proposed to participate in transcription repression by its ability to bind to methylated DNA and co-repressor complexes (Ip et al., 2018). However, MECP2 is a multi-talented protein that performs many functions, as a modulator of gene expression, transcriptional repression or activation, depending on its partners (Boxer et al., 2020); and some other crucial functions in chromatin architecture such as binding to pericentric heterochromatin regulating its condensation (Ip et al., 2018; C. H. Li et al., 2020). In addition, other mechanisms of gene expression regulation conferred by MeCP2 have been described. MECP2 is able to bind methylated miRNA loci controlling miRNA biogenesis and gene expression directly by interacting with DGCR8, a main component of the nuclear microRNA-processing machinery known as DiGeorge syndrome critical region 8, and interfering with Drosha (Cheng et al., 2014; Glaich et al., 2019). Another indirect MECP2 role in gene expression is linked with splicing, favouring splice junctions recognition and intron excision (Glaich et al., 2019; Osenberg et al., 2018). The activity of MECP2 in this process is mediated via interactions with spliceosome components in an RNA-dependent manner, indeed MECP2 protein has been shown to binds RNA *in vitro* (Jeffery & Nakielny, 2004; Young et al., 2005).

In this thesis we have provided evidence of a putative role of MECP2 in suppressing R-loop accumulation and R-loop dependent genome instability by different approaches (γH2AX IF, comet assay, DRIPs and S9.6 IF) in standard cancer cells such U2OS and HeLa. Here we will discuss different possible ways by which MECP2 could prevent R loop accumulation, and R-loop dependent genome instability phenotype (Figure D1):

One possible mechanism is linked to the capacity of MeCP2 to bind broadly across the genome acting as a modulator of transcription of specific genes through its interaction with DNA and with different transcription-repressor factors promoting a repressive chromatin environment (Figure D1A) (AW et al., 2020; Boxer et al., 2020; Ragione et al., 2016). Thus, we show that siMECP2 cells accumulate R loops in MeCP2-regulated genes such as PHLDA2 in addition to standard R loop accumulating genes as RLP13 and APOE (Figure R10B, right panel) (Meng et al., 2014). The hypothesis is that MECP2 depletion could favour R-loop formation because of an open chromatin state more accessible and prone to the formation of DNA-RNA hybrids. Among the interacting partners of MECP2 we highlight the histone deacetylase complex Sin3A, that is involved in co-transcriptional histone deacetylation and prevention of R-loop accumulation and genome instability (Salas-Armenteros et al., 2017). In addition, MECP2 also interacts with the chromatin remodelling complexes Brahma (Brm), a catalytic component of SWI/SNF- complex, and ATRX, both of them with a reported role in R-loop homeostasis and genome instability (Bayona-Feliu et al., 2021; D. T. Nguyen et al., 2017). Another interesting partner is MBD2, the methylated DNA binding component of the methyl-CpG-binding protein 1 (MeCP1) transcriptional repressor complex, identified as an R loop interactor in a biochemical screening (Cristini et al., 2018).

Regarding the role of MECP2 on modulating gene expression is interesting to notice that genome-wide analysis support the view of MeCP2 as a transcriptional regulator of transposons and repetitive elements by repressing inappropriate transcription (Muotri et al., 2010). Different studies in yeast and human cells indicate that these repetitive elements are R loop-prone sites (El Hage et al., 2014; Nadel et al., 2015; Zeng et al., 2021), thus it would be interesting to extend our DRIP analysis in siMECP2 cells to repetitive elements. Interestingly, MECP2 interacts with ATRX, a chromatin remodeler (Nan et al., 2007) that plays a role suppressing deleterious DNA secondary and R loops at telomeric repeats regions (D. T. Nguyen et al., 2017). Given that depletion of ATRX and MECP2 confer an increase in genome instability, it would be worth studying whether both proteins could act together to prevent R-loop-mediated genome instability (Huh et al., 2016; Leung et al., 2013).

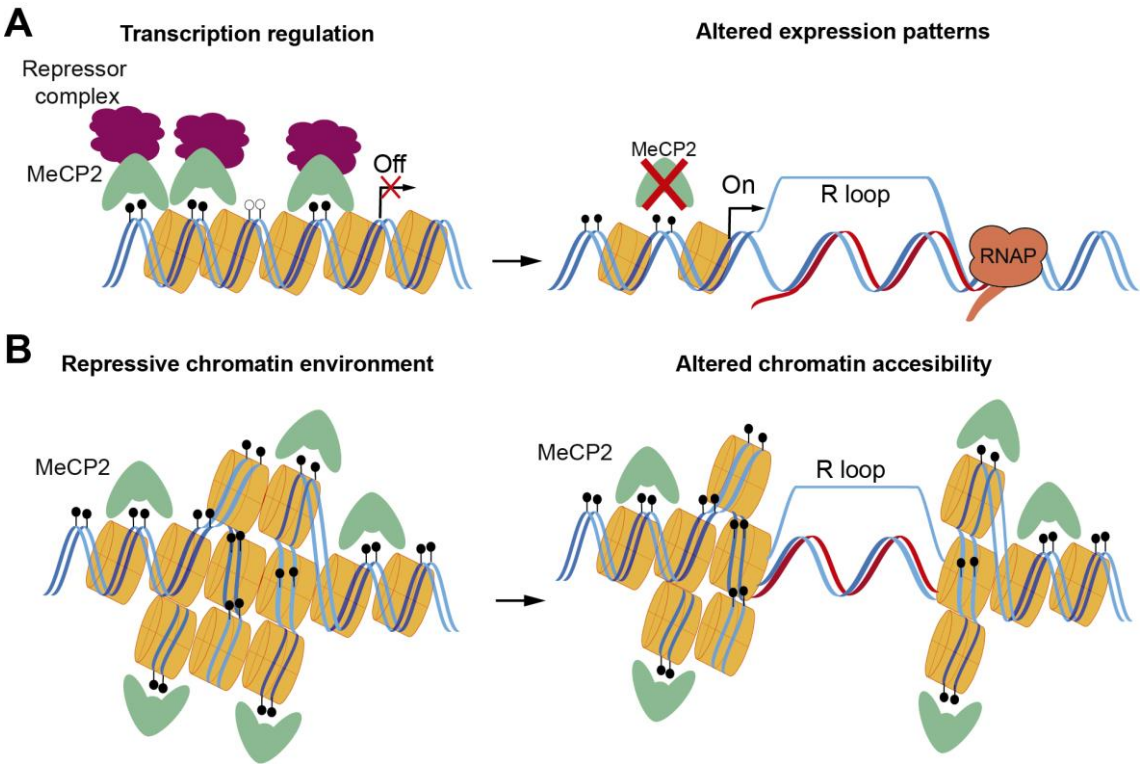


Figure D1. A model to explain the possible role of MeCP2 in genome integrity. (A) MeCP2 through its interaction with DNA and different transcription-repressor factors represses the expression of certain genes. MeCP2 depletion could lead to altered expression patterns and transcription deregulation which in turn, could lead to R loop formation. **(B)** MeCP2 through its capacity to interact with both DNA and nucleosomes and its crucial role in chromatin compaction promotes a repressive chromatin environment. MeCP2 depletion could leads to the unbalance between condensed and decondensed chromatin, favouring R-loop formation because of an open chromatin state more accessible and prone to the formation of DNA-RNA hybrids.

Genome wide data show that MeCP2 has an affinity to GC-rich chromatin, with a preferential binding profile to nucleosomal DNA (Rube et al., 2016). Furthermore, most of the MeCP2-bound promoters genome-wide are expressed genes being only 6% highly methylated, suggesting other roles of MeCP2 in addition to the silencing of methylated promoters (Yasui et al., 2007). Curiously, in human genome, the majority of unmethylated CGI promoters show significant strand asymmetry in the distribution of guanine and cytosines, known as GC skew, where long, stable, three-stranded nucleic acid structures are prone to be formed upon transcription (Liang Chen et al., 2017; Dumelie & Jaffrey, 2017; Ginno et al., 2012; Nadel et al., 2015; Sanz et al., 2016). For this, MeCP2 could have a role in regulating chromatin state at GC skew rich unmethylated promoters, so, when lost, an altered chromatin could be the responsible of R loop formation.

On the other hand, it is important to notice that the majority of MeCP2-binding sites are intergenic or intronic (Yasui et al., 2007). For this, another possible mechanism could be related to the function of MECP2 in alternative splicing regulation, though interaction with both splicing factors and epigenetic marks (Long et al., 2011; Maunakea et al., 2013; Young et al., 2005). It has been shown that MeCP2 binds to methylated DNA regions slowing down RNAPII elongation to enhance intron recognition and processing (Glaich et al., 2019; Wong et al., 2017). So, it is tempt to speculate that upon MECP2 depletion, the nascent RNA could be stickier, favouring the formation of RNA-DNA hybrids structures, according with the idea that a defective mRNP or RNAPII stalling would increase the possibility of hybridization of the mRNA with its DNA template leading R loop accumulation (García-Muse & Aguilera, 2019).

Finally, a third mechanism could be based on the capacity of MeCP2 to interact with both DNA and nucleosomes and its crucial role in chromatin compaction independent of DNA methylation (Figure D1B) (Georgel et al., 2003; Nikitina et al., 2007). MECP2 promotes condensation of pericentric heterocromatin in a RNA-dependent manner and it has been shown to compete with H1 for nucleosome binding (Brero et al., 2005; Ghosh et al., 2010; Good et al., 2021). Recently, it has been shown that MeCP2 and the chromatin remodeler ATRX are reciprocally dependent both for their expression and targeting to pericentric heterochromatin organization as chromocenters (Marano et al., 2019). It is tempt to speculate that the unbalance between condensed and decondensed chromatin could alter expression patterns upon MECP2 depletion and indirectly could cause R-loop accumulation, and genome instability (Figure D1).

The R loop accumulation and R-loop dependent DNA damage in siMECP2 cells reveals a novel connection between R loops and a factor that acts at the interface between transcription regulation and chromatin. The multi-faceted characteristic of this protein, its particular genome distribution and its described crosstalk with some proteins involved in the maintenance of genome stability open new perspectives on its biological role that would need to be explored in the future. Genome-wide analysis and the use of other cells lines with higher MECP2 expression such as the human SH-SY5Y neuronal cell line would contribute to expand the knowledge of this protein in a more physiological context.

3. ROLE OF DDX47 IN THE MAINTENANCE OF GENOME INTEGRITY

DDX47 is a conserved DEAD box RNA helicase that has been shown to be associated to pre-RNAs suggesting a role of this protein at early steps of pre-rRNA processing. This protein is mainly located in the nucleolus by its interaction with the nucleolar protein NOP132, but it has also been associated to ribosome biogenesis proteins and mRNA splicing factors, among others, as determined by pulldown analysis (Sekiguchi et al., 2006). The function in rRNA processing is conserved since, Rrp3 the yeast ortholog (identity 57%) has been reported to be an essential protein required for 18S rRNA production (Granneman et al., 2006; O'Day et al., 1996). In this thesis, we identified DDX47 in our screening as a gene that when depleted causes DNA damage that is increased by AID induction. Moreover, DDX47 is a conserved helicase that acts mainly in the nucleolus, in transcription and rRNA processing and preventing R loop accumulation and genome instability (Figure D2).

Our results indicated that DDX47 in addition to rRNA processing is an helicase involved in rDNA transcription, as determined by ChIP of RNAPI and DDX47 and EU labelling (Figure D2) (Figure R13D, R12B and R14). DDX47 depletion causes a reduction of nucleolar total area (Figure R13A), that could be a consequence of the defect in transcription. Indeed, it has been described that the state of transcription and the amount of rRNA modulate the kinetics of nucleoli formation (Berry et al., 2015). This phenotype was accompanied by an increase in the total amount of RNAPI and UBF, a factor necessary to active rDNA genes (Potapova & Gerton, 2019). This could be explained as a compensation mechanism of living cells to mitigate the observed transcription defect. Altogether data support that DDX47 is necessary for transcription and homeostasis at the nucleolus.

Since DDX47 leads to transcription defects, and genome instability we thought that this helicase could be necessary for a correct progression of polymerases and that its absence could cause transcription-associated genome instability. Indeed, we have observed in siDDX47 cells an increase of rDNA

transcription-replication collision events (Figure D2) (Figure R15). In agreement with these data, another study propose that DDX47 is recruited by FANCD2 and has a role preventing RNAPII transcription-replication conflicts (Okamoto et al., 2019). It would be necessary to study whether these collisions are suppressed by RNAse H1 overexpression, in order to determine whether this genome instability phenotype is mediated by R loop.

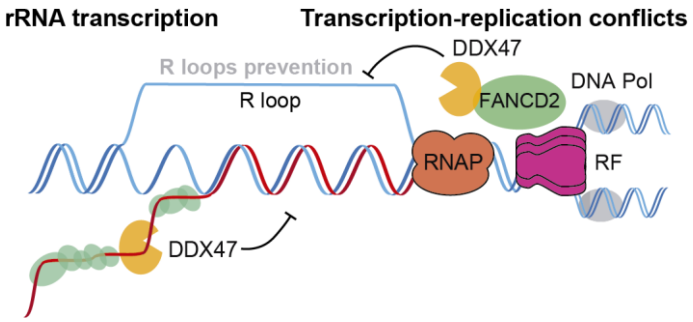


Figure D2. Role of DDX47 in rRNA transcription and transcription-replication conflicts. DDX47 could have a possible role in RNAPI-driven transcription since its depletion causes a reduction of nucleolar total area and rDNA transcription impairment at the nucleolus. DDX47 is also important to maintenance of genome integrity since its depletion leads to R loop accumulation and R-loop-mediated DNA damage as well as an increase of rDNA transcription-replication collision events in the nucleolus.

One interesting question is whether DDX47 has an unique and specific role, or there are helicases with redundant functions. Concerning other nucleolar helicases with similar functions, DHX33 is a nucleolar RNA helicase that is also important for the nucleolar integrity and is involved in mRNA translation initiation by promoting elongation-competent 80S ribosome assembly (C. Zhang et al., 2015). Nevertheless, there is any reported relation of this helicase with genome instability neither with R loops. It is worth to notice the multifaceted nucleolar RNA helicase DDX21 that is required for rRNA processing also plays an important role in resolving R loops and preserving genome instability both in rDNA locus and RNAPII-transcribed regions (Song et al., 2017). Similarly, the RNA-DNA hybrid accumulation phenotype in siDDX47 cells was observed in the nucleoplasm as well as in the nucleolus (Figure R23). Nevertheless, DDX21 has a wider role than DDX47, since in addition to RNA modification also promotes RNAPII transcription elongation (Calo et al., 2015). Interestingly, DDX21 interacts with DDX47 as

determined by pulldown analysis (Sekiguchi et al., 2006), whether these helicases could act in a coordinated manner is something to be explored.

Numerous RNA helicases have been reported to be involved in R-loop dynamics, following the classical SETX, RNA helicases highlight as UAP56, DDX23, DDX5, DHX9, DDX19, DDX21, AQR, DDX1 among others (García-Muse & Aguilera, 2019). However, for few cases, it has been tested a comparative analysis of the efficiency removing DNA-RNA hybrids versus RNA-RNA duplexes and the mechanism of action by which these helicases control R-loop homeostasis *in vivo* is still uncertain. It remains unclear whether these RNA helicases unwind DNA-RNA hybrids directly or whether their ability of remodelling or releasing mRNP could indirectly cause a reduction of R-loop accumulation.

We have shown that DDX47 is able to unwind RNA-RNA duplex in a ATP dependent manner (Figure R19) as well as its described yeast ortholog, Rrp3 (Garcia et al., 2012). We show that DDX47 binds substrates with a preferred orientation, in particular dsRNA with a 3' overhangs and similarly Rrp3 show an unwinding activity stimulated by either 5´ or 3´ single-strand extension. Our data demonstrate a novel role for the highly conserved human DDX47 in unwinding of RNA-DNA hybrids and structures that resemble an R loop *in vitro* (Figure D3) (Figure R20). In fact, DDX47 showed a preference for RNA-DNA hybrids rather than dsRNA duplexes. In order to test the helicase activity we assayed two different mutants, in the conserved motif I and II involved in ATP binding and ATP hydrolysis respectively (Cordin et al., 2006; Linder & Jankowsky, 2011). These mutations were previously assayed for the helicase UAP56 *in vitro* and *in vivo* analysis confirming the helicase activity of the protein (Pérez-Calero et al., 2020). Nevertheless, in the case of DDX47 both mutants showed a residual activity, so it is possible that other residues could contribute to this activity. A comparative structural analysis of some human DEAD-box RNA helicases highlight that the 6^{th} residue upstream of the conserved glutamine of the ATP binding sites is usually an aromatic residue, however DDX47 has a tryptophan in the corresponding position (Schütz et al., 2010), maybe this residue could have a biological meaning that may differentiate DDX47 function from other helicases. It would be interesting to study the effect of tryptophan mutation in DDX47 unwinding activity.

In parallel to the *in vitro* analysis, we also show that DDX47 overexpression in different genetic backgrounds, with high levels of R loops, supresses R loops accumulation, regardless the origin of these RNA-DNA hybrid structures. Thus, overexpression of DDX47 was able to suppress the phenotype of cells depleted of helicases involved in termination and splicing (SETX, DDX23) as well as that of cells depleted of components of Fanconi Anemia (FANCD2) confirming the relevance on its RNA-DNA unwinding activity *in vivo* (Figure D3) (Figure R24). Strikingly, DDX47 was not able to supress UAP56 depletion phenotype. Indeed, the viability of siUAP56 was severely affected upon overexpression of DDX47, suggesting a negative genetic interaction. In addition, the mutants DDX47-K74A and DDX47-E175A show a residual RNA-RNA and RNA-DNA unwinding activity unlike mutants in the same respective residues of UAP56, suggesting that other residues are involved (Figure R20) (Pérez-Calero et al., 2020). Probably DDX47, as RNA helicase with affinity for RNA and RNA-DNA hybrid, could bind to these structures and compete with UAP56 binding interfering with the function of this key helicase. This data, together the different efficiency of RNA-DNA unwinding activities of UAP56 and DDX47 (Figure R20) (Pérez-Calero et al., 2020) suggest that these helicase have different roles. UAP56 will be a general helicase with a global and co-transcriptional unwinding activity over the genome, whereas DDX47 will have a lower efficiency and impact, being more restricted to the nucleolus.

DDX47 dsRNA unwinding activity seems to be higher than that of UAP56 protein (Figure R19) (Pérez-Calero et al., 2020). Interestingly a comparative analysis of UAP56 and DDX49, the DDX47 ortholog, reveals that this helicase, that shares 46% of identity with DDX47, has a robust ATPase and RNA helicase activities of dsRNAs, significantly much higher than those of UAP56 (Awasthi et al., 2018). This activity could account for the different roles of DDX49, as an RNA helicase, that modulates translation by regulating mRNA export and the levels of pre-ribosomal (Awasthi et al., 2018). DDX47 and DDX49 could be more promiscuous performing their functions. In the case of DDX47, for which we have demonstrated an RNA-DNA hybrid unwinding activity, the ability to resolve R loops could be a consequence of its action as an RNA chaperone that resolves

RNA secondary structures and participates in RNP biogenesis adding/removing RBPs, and on other hand in its action as an RNA-DNA helicase (Figure D3).

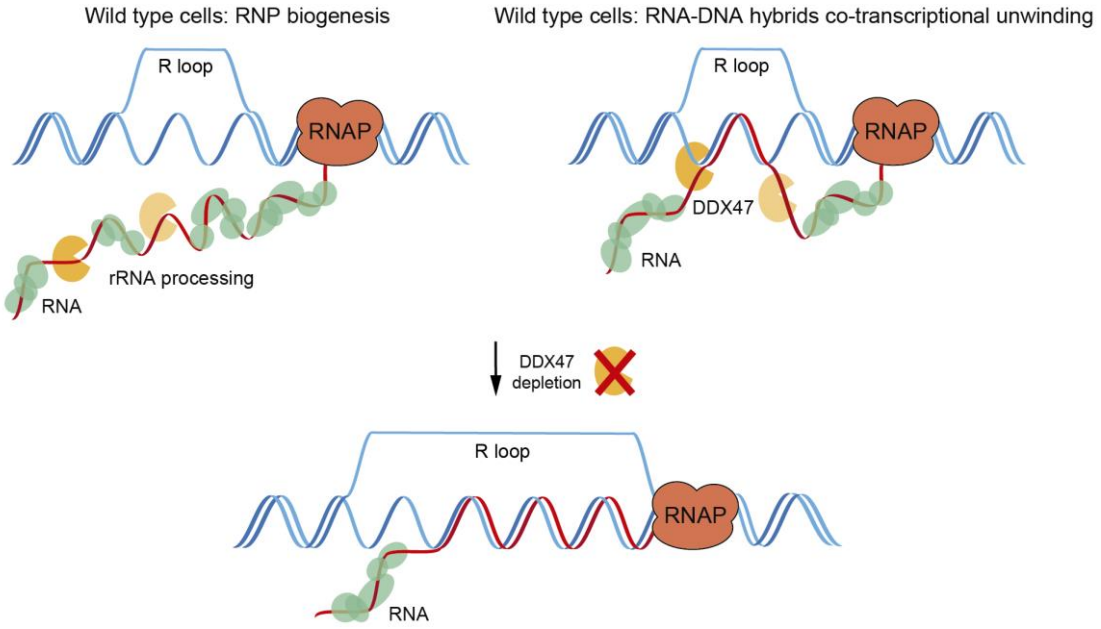


Figure D3. A dual role of DDX47 in R loop metabolism. Under normal circumstances, the co-transcriptional process of RNP biogenesis and a wide variety of RNA binding proteins are associated with the nascent mRNA to ensure an optimal structure of the RNP, reducing the capacity to hybridize back with the DNA template and preventing the R loop formation (left panel). In wild type cells, DDX47 is also able to unwind co-transcriptionally R loops using its R loop resolvase activity (right panel). In both cases, DDX47 depletion leads to R loop accumulation.

In yeast cells, DDX47 is essential and in human cells, the viability of siDDX47 was severely affected after depletion for more than 72 hours. The role of DDX47 in transcription and nucleolar homeostasis could be essential. As well, DDX47 as a part of a complex called small subunit processome (SSUP) that mediate 18S rRNA biogenesis and is required for the maintenance of pluripotent stem cells (You et al., 2015). Mutations and deregulation of several DEAD box proteins have been linked to genome instability and diseases, including cancer (Abdelhaleem, 2004; Cai et al., 2017; Sarkar & Ghosh, 2016). Using the Firebrowse web service, we observed that DDX47 is overexpressed in different cancer cells, this upregulation could contribute to mitigate the associated aberrant R loop accumulation and genome instability in order to overcome replication stress and proliferate; or may be a consequence of a higher

metabolism a more demanding conditions for rRNA synthesis and rRNA processing.

Here we have defined new factors as MeCP2 and DDX47 with different functions in R loop homeostasis and R-loop mediated genome instability. Given the connection between R loop and genetic instability, a hallmark of cancer cells, the study of these factors involved in R loop has an interest in biomedicine. Molecular mechanisms that guide cells to allow positive R loops and prevent the negative ones need to be elucidated in the future.

CONCLUSIONS/CONCLUSIONES

CONCLUSIONS

1- 46 human genes have been identified in an siRNA high-throughput screening covering 3205 different genes with a potential role in R loop and DNA damage homeostasis. Such genes are functionally related with chromatin remodelling and modification, transcription and RNA biogenesis, protein degradation, cell cycle control, and other cellular processes.

2- The two top hits identified were MeCP2, a Methyl-CpG Binding Protein, and DDX47, a nucleolar RNA helicase, both of which were validated and characterized further in this study.

3- Depletion of MECP2 leads to DNA damage and R-loop accumulation, identifying a novel connection between R loops and a factor that acts at the interface between transcription regulation and chromatin.

4- DDX47 depletion leads to R loop accumulation and R-loop-mediated DNA damage. Depletion of DDX47 increases the levels of rDNA transcription-replication collision events in the nucleolus, indicating that this helicase is important to the maintenance of genome integrity.

5- DDX47 depletion causes a reduction of nucleolar total area, transcription impairment at the nucleolus and a slight decrease of RNAPI occupancy at rDNA, suggesting a possible role of this helicase in RNAPI-driven transcription.

6- DDX47 is an RNA-DNA helicase and R-loop resolvase able to remove R loops regardless of their origin, as shown *in vitro* and *in vivo.*

CONCLUSIONES

1- Se han identificado 46 genes humanos en un escrutinio de alto rendimiento de siRNA que abarca 3205 genes diferentes con un papel potencial en la homeostasis de los bucles R (*R loops)* y el daño en el ADN. Estos genes están funcionalmente relacionados con la remodelación y modificación de la cromatina, la transcripción y la biogénesis del ARN, la degradación de proteínas, el control del ciclo celular y otros procesos celulares.

2- Los dos candidatos principales identificados fueron MeCP2, una proteína de unión a metil-CpG, y DDX47, una helicasa de ARN nucleolar, que fueron validados y caracterizados en este estudio.

3- La depleción de MECP2 causa un aumento de daño en el ADN y acumulación de *R loops*, identificándose así una nueva conexión entre los *R loops* y un factor que actúa en la interfaz entre la regulación de la transcripción y la cromatina.

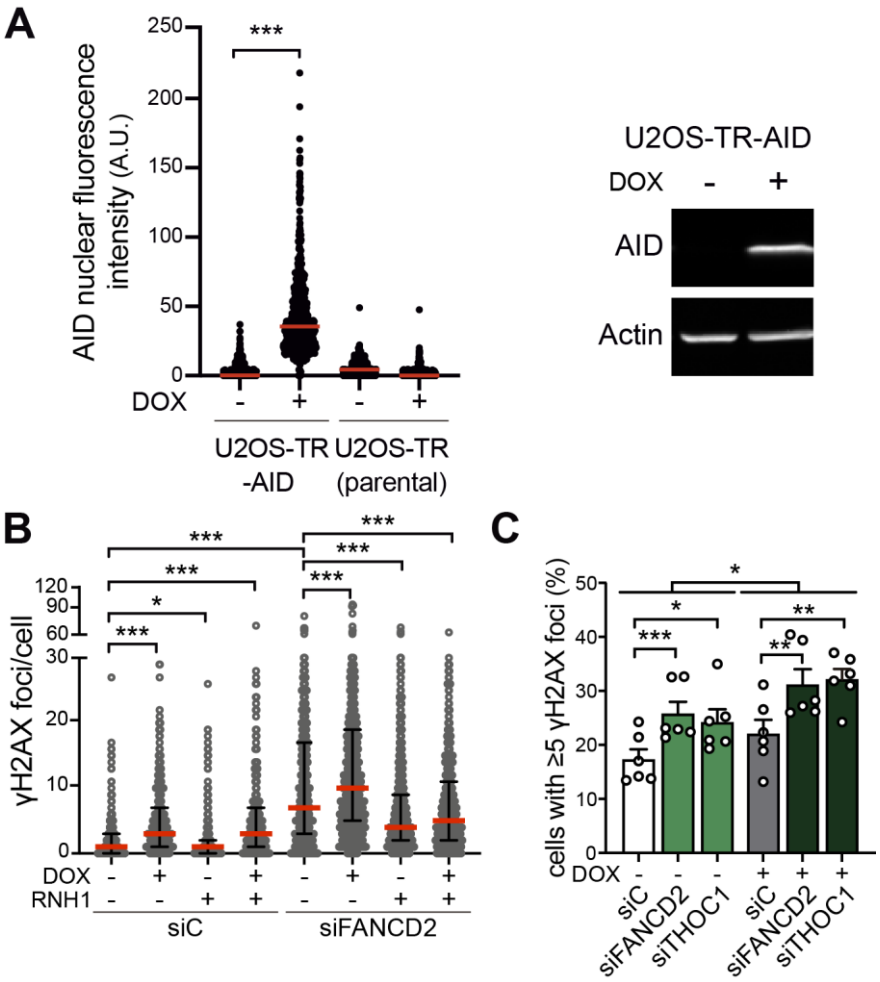
4- La depleción de DDX47 causa acumulación de *R loops* y daño del ADN mediado por *R loops.* La depleción de DDX47 aumenta los niveles de eventos de colisión transcripción-replicación en el rDNA, lo que indica que esta helicasa es importante para el mantenimiento de la integridad del genoma.

5- La depleción de DDX47 provoca una reducción del área total nucleolar, un deterioro de la transcripción en el nucléolo y una ligera disminución del reclutamiento de RNAPI en el rDNA, lo que sugiere un posible papel de esta helicasa en la transcripción llevada a cabo por la RNAPI.

6- DDX47 es una helicasa de ARN-ADN capaz de eliminar *R loops* independientemente de su origen, como se ha demostrado *in vitro* e *in vivo.*

APPENDIX

Appendix 1. Development of an inducible AID-based R-loop detection cell line. (A) Analysis of selected U2OS-TR-AID clone (#8) after 48h AID induction. Quantification of nuclear AID immunofluorescence signal is shown (left panel). AID inducible expression in the selected clone, as determined by western blot from total protein extracts (right panel). **(B)** Quantification of the number of γH2AX foci per cell in control (siC) and FANCD2-depleted (siFANCD2) U2OS-TR-AID cells, in the absence or presence of AID and with or without RNaseH1 overexpression, as indicated. The red line indicates the median (n=3). (A-B) The statistical significance of the difference was calculated with Mann-Whitney U-test; *P <0.05; ***P<0.001. **(C)** Validation of the U2OS-TR-AID cell system in high-throughput format with automated workflow and analysis. Percentages of cells with ≥5 γH2AX foci per cell in control cells (siC), cells depleted for known R-loop metabolism related factors (siFANCD2 and siTHOC1) in the absence or presence of AID. Data are plotted as mean ± SEM (n=2). *P <0.05; **P <0.01; ***P<0.001 as determined by paired two-tailed t-test.



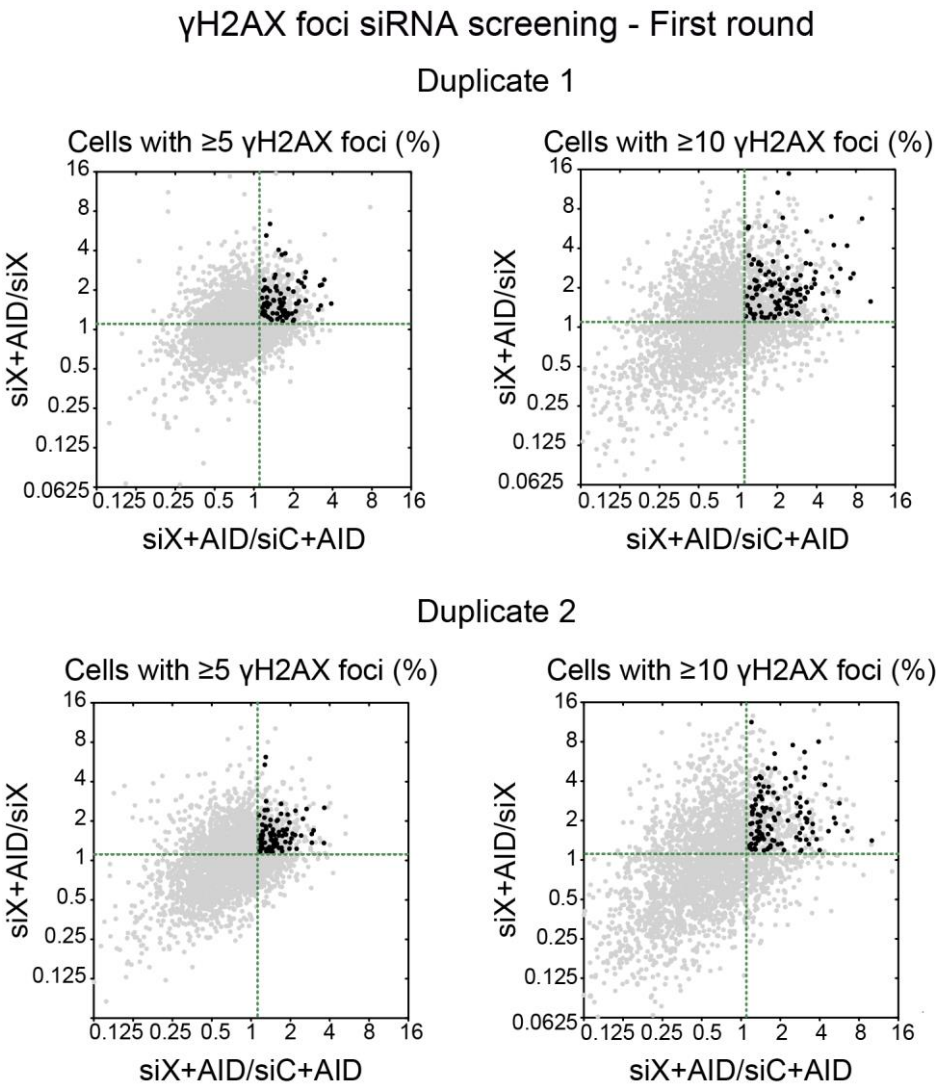
Appendix 2. List of 3205 siRNAs of Dharmacon-ON TARGET Plus-Druggable genome siRNA library. Note: the complete library comprises 4796 siRNAs. In order to improve the screening several cytoplasmic genes were excluded.

https://drive.google.com/file/d/1J48wbdHniDLEzFIgf_HDFy4WPRIihr7Z/view?usp=sharing

Appendix 3. γH2AX foci/cell analysis of 3205 siRNA library (First round). Data show the percentages of cells with ≥5 or ≥10 γH2AX foci with and without AID expression upon siRNA depletion. Two ratios for each siRNA are shown: one versus itself with and without AID expression (siX+DOX/siX), and the other one versus the median of siC with AID expression in its plate (siX+DOX/siC+DOX). The candidates selected (156 out of 3205) in this round of the screening are highlighted.

https://drive.google.com/file/d/1L09bRDq2x5d_j5bBwNvfj_nRxEXP73yf/view?usp=sharing

Appendix 4. Identification of high-throughput screening candidates (First round). Analysis of γH2AX immunofluorescence data (first round: 3205 siRNA in one replicate with a duplicate). The percentages of cells with ≥5 or ≥10 γH2AX foci when AID was expressed were used to calculate the indicated ratios, represented in both axis: Y, siRNA versus itself without AID expression; X, siRNA versus the median of siC with AID expression in its plate. Scatter plot of the ratios for each siRNA is shown. The green dotted lines show the cutoff (1.2) used to designate a siRNA as positive. Two ratios for each siRNA are represented by a dot: black dots: candidates; grey dots: others.



Appendix 5. γH2AX foci/cell analysis of validated candidates (Validation). Data show the percentages of cells with ≥5 or ≥10 γH2AX foci with and without AID expression upon siRNA depletion of the 144 candidate genes and their comparison ratios as explained in Appendix 3. The selected hits (46 genes) defined as those with both ratios ≥ 1.2, at least in two duplicates of one replicate, are highlighted.

https://drive.google.com/file/d/1H-_dRcFeA93MNjD5oOe3B7kckI4U7GmU/view?usp=sharing

MATERIALS AND METHODS

1. GROWTH MEDIA AND CONDITIONS

1.1. Bacteria cell culture

Bacteria were cultured in LB rich medium at 37 °C and supplemented with 100 µg/ml ampicillin or 25 µg/ml kanamycin or 25 µg/ml chloramphenicol for plasmid selection when it was necessary.

LB: 0.5% yeast extract, 1% bacto-tryptone, 1% NaCl (and 2% agar for solid medium).

1.2. Human cell culture

U2OS (ECACC, Sigma) and HeLa (ECACC, 93021013) cells were cultured in Dulbecco's modified Eagle's medium (DMEM; GIBCO, USA) supplemented with 10% heat-inactivated fetal bovine serum (SIGMA Aldrich, Germany), 2 mM L-glutamine and 1% antibiotic-antimycotic (Biowest, France). U2OS-TR-AID stable cell line carrying a tetracycline-regulated AID gene cassette was cultured under the same conditions but supplemented with a special Fetal Bovine Serum (FBS) South America, Tetracycline Free (S181T, Biowest). Cells were grown at 37ºC and 5% CO_2.

2. ANTIBIOTICS, DRUGS, INHIBITORS, ENZYMES AND ANTIBODIES

2.1. Antibiotics

- *Ampicillin* (Sigma): semisynthetic penicillin and β-lactam antibiotic that inhibits bacterial cell-wall synthesis. Used for plasmid selection in *Escherichia coli.* (Use: 100 µg/ml).
- *Kanamycin* (Sigma): aminoglycoside antibiotic that inhibits cell growth. It induces mistranslation and inhibits translocation during protein synthesis in *E. coli*. Used for plasmid selection in *E. coli*. (Use: 25 µg/ml).
- *Penicillin*, *streptomycin*, and *amphotericin B* (Biowest): penicillin is a bacterial cell-wall synthesis inhibitor. (Use: 60 µg/ml). Streptomycin is a prokaryote protein synthesis inhibitor by preventing the transition from initiation complex to chain-elongating ribosome and causes miscoding. (Use: 100 µg/ml).

Amphotericin B is used to prevent growth of bacteria, yeast and fungi in human cell culture. (Use: 0.25 µg/ml).

- *Doxycycline* (Sigma): synthetic oxytetracycline derivative that has been widely used for inducible protein expression in cultured mammalian cells. (Use: 6 µg/ml).
- *IPTG* (Isopropyl-β-D-thiogalactoside) (Roche): chemical analog of galactose that is commonly used in cloning procedures. It has been used for inducing transcription of the lac operon, and it is therefore used to induce recombinant protein expression where the gene is under the control of lac operator in *E. coli.* (Use: 0.2 mM).

2.2. Inhibitors

- *Complete Protease Inhibitor Cocktail* (Roche): mixture of several protease inhibitors including serine, cysteine, aspartic acid and metalloproteases. It was used according to manufacturer's recommendations.
- *Phenylmethanesulfonyl fluride,* (PMSF) (SIGMA): inhibitor of serine and cysteine proteases. (Use: 1 mM).
- *Diethyl pirocarbonate* (DEPC) (SIGMA): chemical used to inactivate RNAse enzymes. (Use: 0.1%).

2.3. Enzymes and antibodies

- *Phusion High-Fidelity DNA polymerase* (M0530S, New England Biolabs): *Pyrococcus*-like enzyme fused with a processivity-enhancing domain that increases fidelity and speed. Used for applications of long or difficult amplicons. It was used according to manufacturer's recommendations.
- *Q5 Hot Start High-Fidelity DNA Polymerase*: high-fidelity, thermostable, hot start DNA polymerase with 3´→ 5´ exonuclease activity, fused to a processivity-enhancing Sso7d DNA binding domain to improve speed, fidelity and reliability of performance. This polymerase supports robust DNA amplification and it has been used as a part of the Q5 Site-Directed Mutagenesis Kit (New England Biolabs).
- *iTaq universal SYBR Green supermix* (Bio-rad): 2x concentrated, ready-to-use reaction master mix optimized for dye-based quantitative PCR (qPCR) that contains antibody-mediated hot-start iTaq DNA polymerase, dNTPs,

$MgCl_2$, SYBR Green I dye, enhancers, stabilizers, and a blend of passive reference dyes (including ROX and fluorescein).

- *Proteinase K* (Roche): very efficient serine protease from *Pichia pastoris* with no pronounced cleavage specificity. (Use: 20 mg/ml).
- *Restriction enzymes* (New England and Takara): DNA endonucleases with specific DNA targets.
- *RNase A* (Roche): endoribonuclease that degrades single-stranded RNA. (Use: 10 mg/ml).
- *Lysozyme* (SIGMA): enzyme purified from chicken egg white that hydrolyzes peptidoglycans.
- *Dynabeads protein A/G* (Invitrogen): magnetic beads bind specifically to the Fc portion of IgG. Used for immunoprecipitation experiments (ChIP and DRIP).

Antibodies used are listed in Table M1 and Table M2 bellow.

Table M1. Primary antibodies used in this study.

Antibody	Source	Epitope	Reference	Use
β-Actin	Rabbit	C-terminal actin fragment attached to Multiple Antigen Peptide backbone	A2066 (Sigma)	WB (1:2000) TBS-T 5% milk
AID	Rabbit	Synthetic peptide corresponding to 14 amino acids near the C-terminus of Human AID	ab59361 (Abcam)	IF (1:100) (1 h)
DDX47	Rabbit	Probable ATP-dependent RNA helicase DDX47 recombinant protein epitope signature tag	HPA014855 (Sigma)	WB (1:250) TBS-T 5% milk IF (1:100) (4 h) PLA (1:100)
DDX47	Mouse	Monoclonal antibody specific for an epitope mapping between amino acids 395-429 near the C-terminus of DDX47 of human origin	sc-377333 (Santa Cruz)	IF (1:100) (o/n) ChIP (10 µg)
FLAG	Mouse	Monoclonal antibody against the FLAG epitope located on FLAG fusion proteins	F3165 (Sigma)	WB (1:5000) TBS-T 5% milk
FLAG	Rabbit	Polyclonal antibody recognizes the FLAG epitope located on FLAG fusion proteins	F7425 (Sigma)	IF (1:500) (4 h)
FLAG	Goat	Polyclonal to DDDDK tag (Binds to FLAG tag sequence)	ab1257 (Abcam)	IF (1:1000) (o/n)

Antibody	Source	Epitope	Reference	Use
Fibrillarin	Rabbit	Synthetic peptide	ab5821 (Abcam)	IF (1:1000) (o/n)
Histidine	-	Monoclonal antibody directed against six histidine residues, which are commonly used as an affinity tag for recombinant fusion proteins	GE27-4710-01 (Sigma)	WB (1:3000) TBS-T 5% milk
IgG	Rabbit	IgG is produced in rabbit using purified mouse IgG as the immunogen	M7023 (Sigma)	ChIP
IgG	Mouse	Mouse IgG is purified from normal mouse serum by fractionation	I8765 (Sigma)	ChIP
MECP2	Rabbit	Synthetic peptide corresponding to Mouse MeCP2 aa 1-15	ab2828 (Abcam)	WB (1:250) TBS-T 5% milk
MECP2	Rabbit	Synthetic peptide corresponding to Human MeCP2 aa 450 to the C-terminus (C terminal) conjugated to KLH	ChIP Grade ab195393 (Abcam)	ChIP (5 µg)
Nucleolin	Rabbit	Synthetic peptide conjugated to KLH, corresponding to N terminal amino acids 2-17 of human Nucleolin with a C-terminal added cysteine	ab50279 (Abcam)	IF (1:1000) (o/n)
PCNA	Rabbit	Synthetic peptide corresponding to Human PCNA aa 200 to the C-terminus (C terminal) conjugated to KLH	ab18197 (Abcam)	PLA (1:500)
Phospho Histone H2AX.X (Ser 139), Clone 2F3	Mouse	Monoclonal antibody. Synthetic peptide corresponding to amino acids 134-142 of human Histone H2A.X	613402 (Biolegend)	IF (1:1000) (1 h)
RNAPI (RPA194)	Mouse	Monoclonal antibody raised against amino acids 1-300 of RPA194 of human origin.	sc-48385 (Santa Cruz)	IF (1:100) (o/n) ChIP (3 µg) PLA (1:100) WB (1:500)
S9.6	Mouse	Antibody that detects RNA-DNA hybrids	Hybridoma cell Line HB-8730	DRIP (3 µg) IF (1:500) (o/n) PLA (1:1000)
UBF	Rabbit	The epitope recognized maps to a region between residue 546 and 596 of human UBF	A301-859A (Behtyl)	WB (1:250) TBS-T 5% milk
Vinculin	Mouse	Monoclonal antibody derived from the hVIN-1 hybridoma	V9264 (Sigma)	WB (1:10000) TBS-T 5% milk

KLH: Keyhole Limpet Haemocyanin; WB: Western blot; IF: immunofluorescence; IP: immunoprecipitation; ChIP: Chromatin immunoprecipitation; DRIP: RNA-DNA immunoprecipitation; PLA: Proximity Ligation Assay; TBS-T: TBS-0.05 % Tween-20; o/n: overnight.

Table M2. Secondary antibodies used in this study.

Specificity	Conjugation	Reference	Use
Rabbit	Horseradish peroxidase	A0545 (Sigma)	WB (1:5000)
Mouse	Horseradish peroxidase	A4416 (Sigma)	WB (1:5000)
Mouse	PLA probe (MINUS oligonucleotide)	DUO92004 (Olink Biosciences)	PLA (1:5)
Rabbit	PLA probe (PLUS oligonucleotide)	DUO92002 (Olink Biosciences)	PLA (1:5)
Mouse	Alexa Fluor 488	A11029 (Molecular Probes)	IF (1:1000)
Mouse	Alexa Fluor 488	A21200 (Molecular Probes)	IF (1:1000)
Mouse	Alexa Fluor 594	A21201 (Molecular Probes)	IF (1:1000)
Mouse	Alexa Fluor 647	A21463 (Molecular Probes)	IF (1:1000)
Rabbit	Alexa Fluor 488	A11008 (Molecular Probes)	IF (1:1000)
Rabbit	Alexa Fluor 555	A31572 (Molecular Probes)	IF (1:1000)
Rabbit	Alexa Fluor 568	A11011 (Molecular Probes)	IF (1:1000)
Goat	Alexa Fluor 647	A21447 (Molecular Probes)	IF (1:1000)

WB: Western blot; IF: immunofluorescence; PLA: Proximity Ligation Assay.

3. BACTERIA AND HUMAN CELL LINES

3.1. Escherichia coli strains

The DH5α strain (*F- endA1 gyr96 hsdR17 ΔlacU169(f80lacZΔM15) recA1 relA1 supE44 thi-1*) (Hanahan, 1983) was used to maintain and amplify DNA plasmids. Rosetta™ host strain (BL21 derivative) (*F- ompT hsdSB(rB- mB-) gal dcm (DE3) pRARE (CamR)*) was used for the expression of eukaryotic proteins (His6-DDX47) (BLB21 Rosetta Electrocompetent Cells -70954-3, Merck).

3.2. Human cell lines

Human cells used in this study are listed in the Table M3.

Table M3. Human cell lines used in this study.

Cell line	Description	Medium	Source
U2OS-TR-AID	Human Bone Osteosarcoma Epithelial Cells carrying a tetracycline regulated human cytidine deaminase AID gene cassette	DMEM	José Manuel Calderón Montaño. Aguilera's lab
U2OS	Human Bone Osteosarcoma Epithelial Cells	DMEM	ATCC
HeLa	Human Cervical Adenocarcinoma Epitelial Cells	DMEM	Sigma (ECACC) (12022001-1VL)

ATCC: American Type Culture Collection; ECACC: European Collection of Authenticated Cell Cultures.

U2OS-TR-AID stable cell line was used for siRNA library screening. Functional analyses were performed in both U2OS and HeLa cells.

4. PLASMIDS

Plasmids used are shown in Table M4.

Table M4. Plasmids used in this study.

Plasmid	Description	Resistance	Reference/Source
pCDNA3	Vector containing a P_{CMV} for expression in mammalian cells	Ampicillin	(ten Asbroek et al., 2002)
pCDNA3-RNaseH1	pcDNA3 containing the human RNase H1 gene under the P_{CMV}	Ampicillin	(ten Asbroek et al., 2002)
pFLAG-CMV-6A	Expression vector derivative of pCMV5 used to establish transient intracellular expression of FLAG- tagged proteins in mammalian cells	Ampicillin	E1900 (Sigma)
pCMV-DDX47-FLAG	Expression vector derivative of pCMV6-Entry containing full-length human DDX47-WT with C-terminal Myc-DDK-tag. Transient expression is driven by the CMV promoter-regulatory region	Kanamycin	RC209448 (Origene)
pCMV-DDX47-K74A-FLAG	Expression vector derivative of pCMV6-Entry containing DDX47-K74A mutation with C-terminal Myc-DDK-tag	Kanamycin	This study
pCMV-DDX47-E175A-FLAG	Expression vector derivative of pCMV6-Entry containing DDX47-E175A mutation with C-terminal Myc-DDK-tag	Kanamycin	This study
pDONR221-DDX47-WT	Vector derivative of pDONOR221 for use in Gateway® Technology. Vector containing a pUC origin for high plasmid yields and the DDX47-WT ORF flanked by Gateway att sites	Kanamycin	This study
pDONR221-DDX47-K74A	Vector derivative of pDONR221 containing DDX47-K74A mutation flanked by Gateway att sites	Kanamycin	This study
pDONR221-DDX47-E175A	Vector derivative of pDONR221 containing DDX47-E175A mutation flanked by Gateway att sites	Kanamycin	This study

Plasmid	Description	Resistance	Reference/Source
pT7-His6-DDX47-WT	Vector derivative of pET300/NT-DEST for rapid cloning with a Gateway® entry clone and subsequent high-level prokaryotic expression controlled by the strong bacteriophage T7 promoter. Vector containing DDX47-WT with an N-terminal 6xHis tag for protein purification	Ampicillin Chloramphenicol	This study
pT7-His6-DDX47-K74A	Vector derivative of pET300/NT-DEST containing DDX47-K74A mutation with an N-terminal 6xHis tag for protein purification	Ampicillin Chloramphenicol	This study
pT7-His6-DDX47-E175A	Vector derivative of pET300/NT-DEST containing DDX47-E175A mutation with an N-terminal 6xHis tag for protein purification	Ampicillin Chloramphenicol	This study

P_{CMV}: Cytomegalovirus promoter; ORF: Open reading frame.

5. BACTERIAL TRANSFORMATION AND HUMAN CELLS TRANSFECTION

5.1. Bacterial transformation

Transformation of bacteria with exogenous DNA was carried out according to standard heat shock transformation protocol (Sambrook et al., 1989).

5.2. Human cells transfection

All assays were performed 72 hours after small interfering RNA (siRNA) transfection and 24 or 48 hours after plasmid transfection.

5.2.1. siRNA transfection

Concerning the high-throughput microscopy screening, we analyzed a collection of siRNAs targeting 3205 human genes (3205 out of 4796 siRNAs of Dharmacon-ON TARGET Plus-Druggable genome siRNA library), in which four-siRNA pool for every targeted gene was used (Appendix 2).

Other siRNA used are shown in Table M5.

Table M5. siRNAs used in this study.

siRNA	Time	Source
ON-TARGETplus Non-targeting Pool (D-001810)	72 h	Dharmacon
ON-TARGETplus SMARTpool human FANCD2 (L-016376-00)	72 h	Dharmacon
HDAC8 siRNAs: 5'-GAC GGA AAU UUG AGC GUA U dTdT-3' 5'-GGA AUU GGC AAG UGU CUU A dTdT-3'	72 h	Dharmacon
MECP2 siRNAs: 5'-GGA AAG GAC UGA AGA CCU G dTdT-3' 5'-ACA CAU CCC UGG ACC CUA A dTdT-3'	72 h	Dharmacon
DDX42 siRNAs: 5'-CGU AAA GGG UAU UCG AGA U dTdT-3' 5'-GUU AAU AGA UCU CCG GCA U dTdT-3'	72 h	Dharmacon
DDX47 siRNAs: 5'-GGA UGA AGC CGA CCG AAU A dTdT-3' 5'-AGA AGA AAC GCU CGC GAG A dTdT-3'	72 h	Dharmacon
MYOG siRNAs: 5'-GGA UGA AGC CGA CCG AAU A dTdT-3' 5'-AGA AGA AAC GCU CGC GAG A dTdT-3'	72 h	Dharmacon
PSMC5 siRNAs: 5'-CCA AGA ACA UCA AGG UUA U dTdT-3' 5'-CAU ACG GAC UGU ACC UUU A dTdT-3'	72 h	Dharmacon
PSMD7 siRNAs: 5'-AGA CGA UUC UGU AUG GUU U dTdT-3' 5'-GAA AGU ACU UGA UGU AUC G dTdT-3'	72 h	Dharmacon
DDX47 UTR siRNA: 5'-CUU CGA CUU UGA UUC CUU G dTdT-3'	72 h	Sigma
ON-TARGETplus SMARTpool human SETX (L-021420-00)	72 h	Dharmacon
ON-TARGETplus SMARTpool human DDX23 (L-19861-01)	72 h	Dharmacon
ON-TARGETplus SMARTpool human UAP56 (L-003805-00)	72 h	Dharmacon

Cells were transfected with siRNA using DharmaFECT 1 (Dharmacon) at 40-50% confluence. Transfection in a well of 6-well plate was performed using the following protocol:

- *Mixture A* (final volume 100 µl): 50 µl serum free-medium (medium without Fetal Bovine Serum (FBS)), 36 µl H_2O, 9 µl 5X siRNA buffer (300 mM KCl, 30 mM HEPES-pH 7.5, 1.0 mM MgCl2) (Dharmacon) and 5 µl siRNA 20 µM (100 nM).
- *Mixture B* (final volume 100 µl): 95 µl serum free-medium and 5 µl DharmaFECT1.

Each mixture was incubated at room temperature (RT) for 5 min. Then, *Mixture A* is added over *Mixture B*, mixed and incubated for 20 min. Meanwhile, medium was replaced by 800 µl antibiotic-free complete medium (medium with FBS but without antibiotics). Transfection solution was added carefully drop by drop to the cell culture and incubated for 2 hours. Afterwards, 1ml of complete medium was added.

5.2.2. Plasmid transfection using Lipofectamine 2000 or Lipofectamine 3000

- For plasmid transfection using Lipofectamine 2000 (Invitrogen), cells were transfected at 80% confluence and were cultured in complete medium (2ml for 6-well plates). Transfection in a well of a 6-well plate was performed using the following protocol:

- *Mixture A* (final volume 200 µl): 2 µg DNA in Opti-MEM (Gibco).
- *Mixture B* (final volume 300 µl): 4 µl Lipofectamine 2000 in Opti-MEM.

Each mixture was incubated at RT for 5 min, mixed and incubated for 5 min at RT. Transfection solution was added carefully drop by drop to the cell culture.

- For plasmid transfection using Lipofectamine 3000 (Invitrogen), cells were transfected at 80% confluence and cell were cultured in complete medium (2 ml for 6-well plates). Transfection in a well of a 6-well plate was performed using the following protocol:

- *Mixture A* (final volume 200 µl): 2 µg DNA and 4 µl of Enhancer Reagent in Opti-MEM (Gibco).
- *Mixture B* (final volume 300 µl): 4 µl Lipofectamine 3000 in Opti-MEM.

Each mixture was incubated at RT for 5 min, mixed and incubated for 5 min at RT. Transfection solution was added carefully drop by drop to the cell culture.

6. siRNA SCREENING METHODS

ON TARGET Plus-Druggable genome siRNA library (Dharmacon, 77G-104655-05) containing four-siRNA pools targeting human genes considered potential targets for therapeutic drugs was used for the screening of genes controlling R-loops. 3205 out of 4796 siRNAs of the library were analyzed. For a general schematic workflow of high-throughput screening, see Figure M1.

siRNA library preparation, transfection and immunofluorescence protocol were carry out using Hamilton Microlab Star available at Genomic Unit at CABIMER. This high-throughput screening was performed in collaboration with Sonia Silva, Lola P. Camino and Jose Javier Marqueta-Gracia.

6.1. siRNA library preparation

siRNA pools were provided lyophilized (0.5 nmol of each siRNA pool) and distributed in 96-well plates. First, siRNAs were reconstituted to 20 µM with 25 µl of 1x siRNA buffer (60 mM KCl, 6 mM HEPES-pH 7.5, 0.2 mM $MgCl_2$) using the Hamilton Microlab STAR with an 8 tip disposable tip head. Then, siRNAs were diluted to 1µM in siRNA buffer. An aliquot of 5 µl of siRNAs was transferred from the collection plates to dilution plates with 95 µl of 1x siRNA buffer. 30 µl of diluted siRNAs (1 µM) were dispensed to new 96-well V-bottom plates containing 30 µl of Opti-MEM (Gibco) to prepare siRNA master mix-plates for transfection protocol.

Collection and dilution plates were immediately sealed using Aluminum Sealing Film (Axygen) and then frozen at -80 ºC so remaining siRNA could be used in future experiments.

6.2. Transfection protocol

A reverse transfection protocol was followed by plating the cells onto pre-plated siRNA-transfection mixtures.

First, 47 µl of each siRNA pool (0.5 µM) in Opti-MEM were transferred from siRNA master mix plates to a new single-use 96-well V-bottom plates, where 47 µl of the transfection reagent Lipofectamine 2000 (0.21 µl/well) in Opti-MEM was previously added. Then, the siRNA-transfection mixtures were gently mixed and incubated for 20 minutes at room temperature to allow siRNA-lipofectamine complex formation to occur. Meanwhile, transfection mixtures were prepared for positive control FANCD2 and UAP56 siRNA pools and the negative non-targeting control siRNA pool.

Next, using Hamilton Microlab Star, two transfection-plates (duplicate-plates) were prepared. 20 µl of each siRNA-transfection mixtures/well were transferred to 96-well tissue culture-treated (TC-treated) polystyrene U-bottom plates (Perkin Elmer) (20 µl/well). In parallel, positive and negative control siRNA -transfection

mixtures were added to 6 wells and 4 wells each of duplicate-plates. These transfection plates were immediately sealed and placed at -80 ºC until use.

For reverse transfection, U2OS-TR-AID cells in optimal growth phase are previously prepared. A final concentration of 33.3 nM siRNA was used to transfect 6000 cells/well. On the day of transfection, the transfection plates were thawed and 80 µl of U2OS-TR-AID cells per well (75,000 cells/ml in DMEM) were added to the assay plates. 4 hours after incubation, 50 µl/well of high-concentrate (3x) complete medium was added to adjust the normal concentration of the medium. For growth media and conditions, see Materials and Methods 1.1.

AID induction

U2OS-TR-AID is a stable cell line carrying a tetracycline-regulated AID gene cassette. Transcription was activated with doxycycline. This clone expresses high-levels of AID upon doxycycline addition, as previously validated by western blot and immunofluorescence analysis in our laboratory (Calderon, Jose). Thus, 24 hours post-transfection, for AID induction, U2OS-TR-AID cells were treated or not with doxycycline (6 µg/ml) and 48 hours later cells were harvested.

6.3. Fixing and staining

72 hours post-transfection, the reaction was stopped by manually tossing the media and fixing the cells in 4% formaldehyde in PBS for 10 min at RT. Cells were permeabilized with 70% ethanol for 5 min at -20 ºC, 5 min at 4 ºC and washed three times in PBS. The immunofluorescence protocol was performed using Hamilton Microlab Star. The plates were blocked by adding 100 µl blocking solution (3% bovine serum albumin (BSA) in PBS). After 1 h, the blocking solution was removed and 50 µl of anti-H2A.X Phospho(Ser 139) antibody (1:1000 in blocking solution; Biolegend 613402) was added and plates were incubated at RT for 1h. The plates were then washed three times. 50 µl of the secondary antibody goat-anti-mouse antibody Alexa-488 conjugated (1:1000 in blocking solution; Molecular Probes, A11029) was added, and the plates were incubated at RT for 1 h. The plates were then washed three times and 100 µl of Hoechst (AnaSpec Inc. Cat# 83219) was added in each well to stain the nucleus. Plates were sealed and stored at 4 ºC until imaging. As indicated before, the different

steps in the immunofluorescence protocol were performed using Hamilton Microlab Star.

6.4. Imaging

Data acquisition was performed with ImageXpress Micro Electron (Molecular Devices, 137239) using a 40x objective. The entire 96-well plate (Perkin Elmer) was scanned with a 20 nW 690 nm laser and set up for two channels of acquisition. γH2AX-Alexa-488 fluorescence was acquired using a 472/30 nm excitation filter, and Hoechst fluorescence was acquired using 377/50 nm excitation filter. Randomly fields were acquired and analyzed (≃ 200 cells/well). H2AX foci were quantified by automated scoring using MetaXpress software (version 4.0.0.24 Molecular Devices-granularity application).

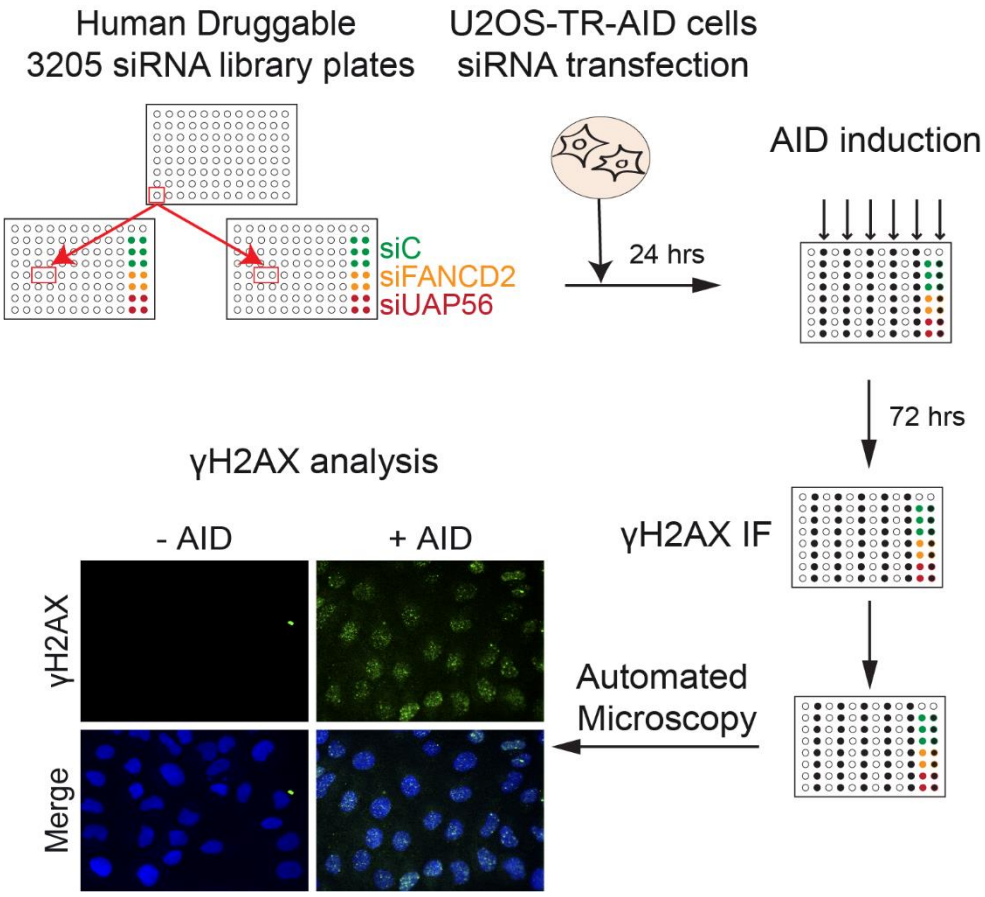


Figure M1. Workflow for high-throughput screening of factors involved in R-loop metabolism. Schematic workflow for testing the human druggable siRNA library (3205 siRNAs) for γH2AX foci formation in the absence or presence of AID. U2OS-TR-AID cell line are transfected and treated with doxycycline for 48h to induce AID expression. An example of γH2AX detection (green) by automated microscopy of cells with or without AID expression is shown. Nuclei were detected by DAPI (blue).

6.5. Data analysis and statistical analysis

Here we describe the procedure used to analyze the raw data and to determine which siRNA transfected cells conferred a H2AX increase after AID induction and were considered positive candidates.

We used MetaXpress in order to measure Granule Count that characterizes γH2AX foci associated to each siRNA and condition in each well. We calculated the percentage of cells with ≥5 γH2AX foci before and after AID induction of each siRNA depletion. The screening was performed in duplicate, and siRNAs whose depletion lead to an increase (≥1.2) in the percentage of cells with γH2AX foci after AID induction (siRNA +DOX), versus non-induced conditions (siRNA -DOX), that was higher (≥1.2) to that observed in control cells (siRNA +DOX versus siC +DOX) were selected as candidates. To take into account the variability of the data, per-plate and per-date, each condition was always compared versus non-induced conditions and control cells located in the same plate. The same procedure was performed to calculate the percentage of cells with ≥10 γH2AX foci after AID induction. The Z-score and the reproducibility score were calculated to take into consideration when candidate selection.

$$\mathrm{Z} - score = \frac{\% - \mu}{\sigma}$$

%= percentage of cells with ≥5 γH2AX foci / percentage of cells with ≥10 γH2AX foci.

μ= Mean of each plate.

σ= Standard deviation of each plate.

$$Reproducibility = 2^n$$

n= Number of times that a candidate meets the criteria described above.

For positive control ratios in the first round of the high-throughput screening, see Appendix 3.

7. SITE-DIRECT MUTAGENESIS

To generate the DDX47 mutant forms, we used the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) according manufacturer's recommendations. The protocol consists of a first exponential amplification (PCR) with the mutagenic primer pairs (Table M7), using the pCMV-DDX47-FLAG plasmid as template DNA. The primers were designed with the NEBaseChanger tool to optimize the sequence and the annealing temperature. Next, a Kinase, Ligase and DpnI (KLD) treatment for 5 min at RT was performed in order to phosphorylate, ligate the PCR product and remove the template. Competent cells were transformed with this reaction, transformants were checked by restriction enzymes digestion and finally plasmidic DNA was sequenced to confirm the presence of the mutation. For a schematic representation of mutations in the helicase core of DDX47, see Figure R17A.

8. DEVELOPMENT OF TOOLS FOR *IN VITRO* ANALYSIS AND *IN VIVO* OVEREXPRESSION OF DDX47

8.1. Tools for *in vitro* analysis

We used the Gateway cloning System (Invitrogen) to generate a vector for expression of DDX47 WT and DDX47 mutants in *E. coli*. Gateway technology takes advantage of the site-specific recombination properties of bacteriophage lambda to provide a highly efficient way to move the gene of interest into multiple vector systems. First, we amplified DDX47 ORF with specific primers for addition of attB sites. Then, we cloned DDX47 ORF sequences into a Gateway® entry vector pDONR221 to create an entry clone, and after we generated an expression clone by a recombination reaction between the entry clone and the destination vector. In this case, our destination vector was pET300/NT-DEST (Table M4) designed to allow high-level inducible expression from T7 promoter of recombinant proteins in *E. coli* (N-terminal 6xHis tag). Expression was induced in BL21 *E. coli* transformant exponential cultures by adding IPTG (0.2 mM) at exponential cultures for 3h (37ºC) or overnight (16ºC). The induction was tested by Coomassie blue staining and western blot (Materials and Methods 17).

8.2. Tools for *in vivo* overexpression

DDX47 overexpression analyses were carried out with the vector pCMV6-Entry containing full-length human DDX47-WT with C-terminal Myc-DDK tag that allows expression from pCMV promoter (pCMV-DDX47-FLAG) (RC209448, Origene) (Table M4). DDX47 mutants were generated with the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) as described below (Materials and Methods 7). Protein expression was confirmed by immunofluorescence and western blot analysis (Materials and Methods 12 and 17).

9. *IN VITRO* ANALYSIS

These experiments were performed in collaboration with Xiaoyu Xue lab in Yale University.

9.1. Purification of DDX47 wild-type and mutant proteins

The cDNAs that encode the wild type, K74A and E175A variants of DDX47 were introduced into the pET300/NT-DEST vector to add and N-terminal 6xHis tag to these proteins (Table M4; Materials and Methods 8.1.). The resulting DDX47 expression plasmids were introduced into *E. coli* BL21:DE3 Rosetta cells, which were grown at 37ºC to OD_{600nm} = 0.8, and protein expression was induced by the addition of 0.2 mM IPTG and incubation at 16ºC for 16 h. Cells were harvested by centrifugation and all the subsequent steps were carried out at 0-4ºC. For lysate preparation, a cell pellet (20 g, from 4 L of culture) was suspended in 100 ml K buffer with 300 mM KCl and 5 µg/ml each of the protease inhibitors aprotinin, chymostatin, leupeptin and pepstatin, and then subject to sonication (three 1 min pulses). The crude cell lysate was clarified by ultracentrifugation (100,000Xg for 90 min) and then mixed gently with 2 ml of Glutathione Sepharose 4B resin (GE) for 1.5 h. The resin was washed sequentially with 50 ml K buffer containing 1 M KCl, 50 ml K buffer containing 300 mM KCl, 20 ml K buffer containing 300 mM KCl and 1 mM ATP, 20 ml K buffer containing 300 mM KCl and 5 mM $MgCl_2$, and 2 x 50 ml K buffer containing 300 mM KCl. DDX47 was eluted with 12 ml K buffer containing 300 mM KCl and 10 mM reduced glutathione and concentrated to 1 ml (Amicon 10K concentrator, Millipore). The 6xHis tag was cleaved by incubating the concentrated protein pool with 100 µg of thrombin for 12 h. The reaction

mixture was diluted with 2 ml of K buffer and applied onto to a 1-ml Mono Q column (GE), which was washed with 5 ml K buffer plus 150 mM KCl and then developed with a 25-ml linear gradient from 150 to 650 mM KCl. The peak of DDX47, eluting at ~350 mM KCl, was collected, concentrated to 0.5 ml, and fractionated in a Superdex 200 gel filtration column (24 ml, GE) in K buffer containing 300 mM KCl. Fractions containing highly purified DDX47 (1 mg protein) were pooled, concentrated to 1 mg/ml, and stored in small aliquots at -80°C. The DDX47 K74A and E175A mutants were purified using the same procedure with a similar yield.

K buffer: 20 mM KH2PO4, pH 7.4, 10% glycerol, 0.5 mM EDTA, 0.01% Igepal, 1 mM DTT.

9.2. Nucleic acid unwinding assays

RNA-RNA duplexes, without or with a 5' or 3' overhang, were prepared as described (Shen et al., 2007). RNA-DNA hybrids without and with a 5' or 3' overhang were prepared by annealing oligonucleotides (with one of the oligonucleotides being labeled with ^{32}P) listed in Table M6. In the unwinding reaction, DDX47 (wild type or mutant at the indicated concentration) was incubated with 5 nM substrate in reaction buffer (1) and 100 nM of "trap" RNA or DNA (unlabeled version of the oligonucleotide that was labeled in the substrate) at 37°C (for the RNA-RNA substrates) or 30°C (for the RNA-DNA substrates) for 30 min or the indicated time. Reaction mixtures were deproteinized by treatment with SDS (0.1%) and proteinase K (0.5 mg/ml) for 10 min at 37°C and then resolved in 15% polyacrylamide gels in TAE buffer at 4°C. Gels were dried and subject to phosphorimaging analysis.

The 5' RNA-DNA flap structure that resembles a branch migratable R-loop structure was constructed as described (Schwab et al., 2015). DDX47 (wild type or mutant at the indicated concentration) was incubated with the substrate in reaction buffer (2) at 30°C for 20 min. Reaction mixtures were deproteinized before being resolved in 7% polyacrylamide gels in TAE buffer at 4°C and analyzed, as above.

Reaction buffer (1): 35 mM Tris-Cl, pH 7.5, 1 mM DTT, 3 mM ATP, 2 mM $MgCl_2$, 60 mM KCl.

Reaction buffer (2): 25 mM Hepes, pH 6.5, 1 mM DTT, 2 mM ATP, 2 mM $MgCl_2$, 60 mM KCl.

TAE buffer: 40 mM Tris, 20 mM Acetate acid and 1 mM EDTA.

Table M6. Oligonucleotides for unwinding assays used in this study.

Oligo Name	RNA or DNA	Length	Sequence
R13	RNA	13	5'GCUUUACGGUGCU3'
R13C	RNA	13	5'AGCACCGUAAAGC3'
R23-5'	RNA	23	5'AAAACAAAAUAGCACCGUAAAGC3'
R23-3'	RNA	23	5'GCUUUACGGUGCUUAAAACAAAA3'
D13	DNA	13	5'GCTTTACGGTGCT3'
D13C	DNA	13	5'AGCACCGTAAAGC3'
D23-5'	DNA	23	5'AAAACAAAATAGCACCGTAAAGC3'
XX1	DNA	60	5'ACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACA TCTTTGCCCACCTGCAGGTTCACCC3'
XX2	DNA	60	5'GGGTGAACCTGCAGGTGGGCAAAGATGTCCCAGC AAGGCACTGGTAGAATTCGGCAGCGT3'
R5'F	DNA	30	5'GGGUGAACCUGCAGGUGGGCAAAGAUGUCC3'

10. PROTEIN-PROTEIN INTERACTION METHODS

10.1. Proximity Ligation Assay (PLA)

For detection of protein-protein interactions *in situ*, PLA was performed following manufacturer's instruction with reagents from Duolink In Situ Starter Kit (Olink Biosciences). Cells were cultured on glass coverslips and fixed in 4% formaldehyde in PBS for 10 min at RT and permeabilized with 0.5% triton X-100 in PBS for 5 min at RT. Coverslips were blocked with PBS-3% BSA for 1 h at RT, incubated with primary antibodies diluted in PBS-3%BSA overnight at 4ºC (for antibodies dilution see Table M1). After the two primary antibodies incubation, coverslips were washed three times with PBS for 5 min and incubated with PLA probes for 1 h at 37 °C. After two washed with wash buffer A for 5 min, ligation reaction was performed for 30 min at 37 °C. Then, cells were washed twice in 5 ml wash buffer A for 2 min, incubated with the amplification reaction for 100 min at 37 °C in darkness, washed twice with buffer B for 10 min and once with 0.01X wash buffer B for 1 min. Finally, coverslips were dried in darkness, mounted with mounting medium with DAPI and images were acquired in a fluorescence microscope. For negative controls, everything was performed identically, except that only one of the primary antibodies was added.

For each coverslip:

- **PLA probes** (40 µl): 8 µl PLA probe anti-Mouse MINUS, 8 µl PLA probe anti-Rabbit PLUS, 24 µl PBS-3% BSA.
- **Ligation reaction** (40 µl): 8 µl of 5X Buffer ligase, 1 µl Ligase, 31 µl MQ H_2O.
- **Amplification reaction** (40 µl): 8 µl of 5X Buffer amplification, 0.5 µl Polymerase, 31.5 µl MQ H_2O.

PLA probes, buffers, enzymes and mounting media are provided in the Duolink In Situ Red Starter Kit.

11. TRANSCRIPTION ANALYSIS IN HUMAN CELLS

11.1. EU incorporation

In this assay the modified uridine analogue EU (5-ethynyluridine) is efficiently incorporated into the newly transcribed RNA *in vivo* during active synthesis and fluorescently labelled with a bright, photostable Alexa Fluor dye in a reaction that allows the visualization of RNA transcription sites (Jao & Salic, 2008). Click-iT RNA Imaging Kit (C10329, Invitrogen) was used following manufacturer's instructions.

Hela cells were seeded in 6-well plates and transfected with siRNA. After 72 h siRNA transfection, HeLa cells were incubated with EU at the final concentration of 1 mM (C10329, Invitrogen) in pre-warmed complete medium. We tested a range of EU incubations from 20 min to 120 min, with an intermediate point at 60 min. Subsequently, cells were fixed with 4% formaldehyde in PBS for 15 min at RT, washed three times with PBS, permeabilized with 0.5% Triton X-100 for 15 min, washed again with PBS. Afterwards, cells were incubated with 50 µl/coverslip of Click-iT reaction cocktail (42.8 µl of 1X Click-iT reaction buffer, 2 µl of 100 mM $CuSO_4$, 0.18 µl of Alexa Fluor azide, 5 µl of 1X reaction buffer additive) for 30 min at RT in darkness and washed once with 1 ml of Click-iT reaction rinse buffer. For antibody labeling, cells were blocked in 3% BSA in PBS for 1 h and incubated with Nucleolin primary antibody overnight at 4ºC. For antibody dilutions and incubation times see Table M1. Then, cells were washed three times in PBS and incubated with secondary antibodies conjugated with Alexa Fluor diluted (1:1000) in 3% BSA in PBS for 1 h at RT in darkness. (See

Table M1, for antibodies information). Next, coverslips were washed twice in PBS for 5 min and nuclei were counterstained with 1 µg/ml DAPI (2-(4-Amidinophenyl)-6-indolecarbamidine dihydrochloride) in PBS for 5 min, washed three times in PBS, once in water and mounted with a drop of ProLong Gold antifade reagent (Thermo). Random images were acquired with a 63x objective, and EU intensity was scored using the MetaMorph software (see Materials and Methods 14).

12. IMMUNOFLUORESCENCE

Cell were cultured on glass coverslips and, if needed, transfected as indicated in Materials and Methods 5.2. Cells were fixed in formaldehyde or methanol and specific target molecules were visualized in a fluorescence microscope after the incubation with the corresponding primary antibodies and with the subsequent fluorophore-conjugated secondary antibodies.

Type of cell fixation used:

- Formaldehyde fixation – Ethanol permeabilization: cells were fixed in 4% formaldehyde in PBS for 10 min at RT and permeabilized with 70% ethanol for 5 min at -20 ºC, 5 min at 4 ºC and washed three times in PBS. This fixation was used for immunofluorescence with γH2AX, AID, Nucleolin, DDX47, Fibrillarin and FLAG antibodies.
- Formaldehyde fixation – Triton permeabilization: cells were fixed in 4% formaldehyde as previously described, and permeabilized with 0.5% triton X-100 in PBS for 5 min. This fixation was used for immunofluorescence with Nucleolin and RNAPI antibodies.
- Methanol fixation: Cells were fixed in cold absolute methanol for 8 min at -20 °C and washed three times in PBS. This fixation was used for immunofluorescence with S9.6, FLAG and Nucleolin antibodies.

Immunofluorescence: Blocking, incubation with primary and secondary antibodies, DAPI staining and mounting conditions to prepare for imaging.

Cells were blocking using with 3% bovine serum albumin (BSA) in PBS. Afterwards, the coverslips were incubated with primary antibodies diluted in 3% BSA in PBS. For antibody dilutions and incubation times see Table M1. Then,

cells were washed three times in PBS and incubated with secondary antibodies conjugated with Alexa Fluor diluted (1:1000) in 3% BSA in PBS for 1 h at RT in darkness. Next, coverslips were washed twice in PBS for 5 min before and after the staining of the DNA with 1 µg/ml DAPI (2-(4-Amidinophenyl)-6-indolecarbamidine dihydrochloride) for 5 min. Coverslips were washed in water and a drop of ProLong Gold Antifade reagent (Thermo) was used for mounting. For S9.6 immunofluorescence see Materials and Methods 13.3.2.

13. GENOME INSTABILITY ANALYSIS

13.1. Analysis of γH2AX foci

Cells were cultured on glass coverslips and transfected with siRNA as indicated in Materials and Methods 5.2., fixed in 4% formaldehyde and permeabilized with 70% ethanol. Immunofluorescence was performed with anti-H2A.X Phospho(Ser 139) antibody (613402 Biolegend) (see Materials and Methods 12 and Table M1). More than 200 cells from each experiment were analyzed (see Materials and Methods 14).

13.2. Alkaline single cell gel electrophoresis (Alkaline comet assay)

Comet assay was performed using a commercial kit (Trevigen 4250-050-ES, Gaithersburg, MD, USA) following the manufacturer's protocol, 72 h after siRNA transfection. In the case of comet assay experiments overexpressing RNase H1, cells were transfected with siRNA for 72 h and with pcDNA3 (- RNase H1) or pcDNA3-RNase H1 (+ RNase H1) 48 h before the experiment.

Cells were collected using accutase, washed and resuspended in ice cold 1X PBS (200 000 cells per ml), combined with low melting agarose, immobilized on CometSlides (30 min at 4 °C, until agarose is solidified) and lysed for 30 min at 4 °C. Then, DNA was unwound and denatured in freshly prepared alkaline unwinding solution pH>13 for 20 min at RT and electrophoresis was performed in prechilled alkaline electrophoresis solution pH>13 at 21 V for 30 min. Next, slides were immersed twice in dH_2O for 5 min each, then in 70% ethanol for 5 min and dried at RT. DNA was stained with SYBR Gold Nucleic Acid Gel Stain at 4 °C for 5 min.

Lysis solution: Kit CometAssay ES II (Trevigen 4250-050-ES).

Alkaline unwinding solution: 0.4 g of NaOH, 250 µl 200 mM/pH10 EDTA were diluted up to 50 ml with H_2O.

Alkaline electrophoresis solution: 8g of NaOH, 2 ml of 500 mM/pH8 EDTA were diluted up to 1 l with H_2O.

For comet assays analyses at least three independent experiments were performed. More than 200 cells were scored in each experiment (see Materials and Methods 14).

13.3. RNA-DNA hybrids detection

13.3.1. RNA-DNA hybrids immunoprecipitation (DRIP-qPCR)

DRIP assays were performed by immunoprecipitating DNA–RNA hybrids using the S9.6 antibody from gently extracted and enzymatically digested DNA, treated or not with RNase H (New England Biolabs, USA) in vitro as described (M. García-Rubio et al., 2018; M. L. García-Rubio et al., 2015; Herrera-Moyano et al., 2014), with minor modifications. After 72 h of siRNA transfection, pellet from a confluent 10-cm plate of cells was collected using accutase, washed in PBS and resuspended in 800 µl of 1X TE. Then, 20.75 µl SDS 20% and 2.5 µl proteinase K (20 mg/ml) were added and pellet was incubated at 37 °C overnight. DNA was extracted gently with phenol-chloroform in phase lock tubes (VWR, USA). Precipitated DNA was spooled on a glass rod, washed 2 times with 70% EtOH, resuspended gently in 1X TE and digested overnight with 50 U of HindIII, EcoRI, BsrGI, XbaI and SspI. For the negative control, half of the DNA (16 µg) was treated with 5 µl RNase H overnight. In parallel, S9.6 antibody (3 µg/sample) was incubated overnight at 4ºC with Dynabeads Protein A (Invitrogen) (30 µl/sample) in 1X binding buffer (final volume: 60 µl/sample). 5 µg of the digested DNA, untreated or treated with RNase H, were bound to S9.6 antibody-dynabeads complexes during 2h at 4ºC for immunoprecipitation. Next, the beads were washed 3 times with 1X binding buffer. DNA was eluted in 180 µl elution buffer, treated 45 min with 7 µl proteinase K at 55ºC and cleaned with the NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, USA). 1 µg of the digested DNA, untreated or treated with RNase H, was used for input DNA of each condition, in which similar proteinase K treatment and purification were performed.

Quantitative PCR (qPCR) of immunoprecipitated DNA (IP) fragments and input DNA was performed on a 7500 Fast Real-Time PCR System (Applied Biosystems, Carlsbad,CA). Primers used are listed in Table M7.

10X Binding buffer: 100 mM $NaPO_4$ pH 7.0, 1.4 M NaCl, 0.5% triton X-100. **Elution buffer**: 50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS.

DRIP quantification and normalization: Input and immunoprecipitated (IP) were eluted in 150 µl of 1X TE. 2 µl of input and IP were used for qPCR. Changes in the abundance of RNA-DNA hybrids in each region were determined as the percentage of input recovered for each immunoprecipitated sample using the equation $2^{[\ Ct_{IP} - (Ct_{INPUT} - \log 2\ DF)]}$. Notes: (Ct -threshold cycle); (DF - dilution factor).

13.3.2. S9.6 immunofluorescence

Cells were cultured on glass coverslips and fixed in methanol (see Materials and Methods 12). When required, for RNase treatments cells were incubated in their respective commercial buffers at 1x containing 40 U/ml RNase III (AM2290, Ambion) or 60 U/ml RNase H (M0297S, NEB), for 30 min at 37°C. Next, for S9.6 immunofluorescences, coverslips were blocked in 3% BSA in PBS for 5 h at 4ºC and incubated with anti-S9.6 (Hybridoma cell Line HB-8730, 1:500) and anti-nucleolin (ab50279, 1:1000) or anti-FLAG (ab1257, 1:1000) primary antibodies diluted in 3% BSA in PBS o/n at 4 °C, washed in PBS three times and incubated with the subsequent secondary antibodies conjugated with Alexa Fluor diluted in 3% BSA in PBS (1:1000) for 1 h at RT in darkness. Washed twice for 5 min, DAPI staining and mounting as described above. More than 200 cells from each experiment were scored (see Materials and Methods 14).

14. MICROSCOPY IMAGES ACQUISITION, DATA ANALYSIS AND STADISTICAL ANALYSIS

14.1. Fluorescence microscopy

Data acquisition was performed with LAS AX (Leica) equipped with a DFC390 camera. A 63x objective was used for immunofluorescence (γH2AX, AID, RNAPI, Fibrillarin, Nucleolin, DDX47 IF, PLA, S9.6, FLAG and Nucleolin IF) and a 10x objective was used for comet assays.

14.2. Data analysis

- γH2AX foci and RNAPI intensity measurements were analyzed and processed with the MetaMorph v7.5.1.0. software using the *granularity* application.
- S9.6 signal intensity per nucleus was analyzed and processed with the MetaMorph v7.5.1.0. software using the *multi wavelength cell scoring* application. The S9.6 signal corresponding to the nucleolus area was previously removed using the nucleolin signal and *granularity* application. The S9.6 signal intensity corresponding to nucleolus per cell was analyzed using the nucleolus area and *granularity* application.
- FLAG and PLA signal intensity per nucleus were analyzed and processed with the MetaMorph v7.5.1.0. software using the *multi wavelength cell scoring* application.
- Nucleolar total area measurements were analyzed and processed with the MetaMorph v7.5.1.0. software using the *granularity* application and nucleolin immunofluorescence signal.
- Comet assays tail moments were analyzed using TriTek CometScore Professional (version 1.0.1.60) software. Tail moment (TM) reflects both the tail length (TL) and the fraction of DNA in the comet tail (TM=%DNA in tail x TL/100).
- EU signal intensity corresponding to nucleolus per cell was analyzed using the nucleolus area and *granularity* application with the MetaMorph v7.5.1.0. software. EU signal intensity per nucleus was analyzed and processed using the multi wavelength cell scoring application, where EU signal corresponding to the nucleolus area was previously determined using the nucleolin signal and *granularity* application.

For all experiments, at least three biological repeats (n) were performed. More than 200 cells were scored in each repeat.

14.3. Statistical analysis

- γH2AX foci number per cell, PLA signal intensity and S9.6 intensity per nucleus: Graphs show the median of the measurements from at least three biological repeats. Data were analyzed with GraphPad Prism software. For statistical analysis, Mann-Whitney U-test, two tailed was performed and P

value < 0.05 was considered as statistically significant. (***, P < 0.001; **, P < 0.01; *, P < 0.05).

- DRIP: Graphs show changes in the abundance of RNA-DNA hybrids in each region were determined as the percentage of input recovered for each immunoprecipitated sample normalized with respect to the siC control from at least three biological repeats. Data were analyzed with EXCEL program and GraphPad Prism software. For statistical analysis, Student's t-test, one tailed was performed and a P value < 0.05 was considered as statistically significant.
- ChIP: Graphs show values represent the percentage of the precipitated DNA (IP) to input DNA (Input) from at least three biological repeats. Data were analyzed with EXCEL program and GraphPad Prism software. For statistical analysis, Student's t-test, one tailed was performed and a P value < 0.05 was considered as statistically significant.
- Comet assay: Graphs shows the median of tail moment from at least three biological repeats. Data were analyzed with EXCEL program and GraphPad Prism software. For statistical analysis, Mann-Whitney U-test, two tailed was performed.
- Nucleolar total area per nucleus, RNAPI intensity and EU experiments. Graph shows the mean of medians of measurements from at least three biological repeats. Data were analyzed with GraphPad Prism software. For statistical analysis, Student's t-test, two tailed was performed and a P value < 0.05 was considered as statistically significant. (***, P < 0.001; **, P < 0.01; *, P < 0.05).

15. CHROMATIN IMMUNOPRECIPITATION (ChIP) ASSAY

After 72 h of siRNA transfection, HeLa cells were crosslinked for 10 min with 11% formaldehyde crosslinking solution, resuspended in 2.5 ml of cell lysis buffer 1, then centrifuged and 1 ml of nuclei lysis buffer 2 was added. Chromatin was sonicated on the maximum intensity setting, with eight pulses of 30 s on and 30 s off in Bioruptor (Diagenode), to obtain approx. 500 bp fragments. For each immunoprecipitation, 30 µg of chromatin were diluted up to 1100 µl with IP buffer. 100 µl and 1000 µl of diluted chromatin were used for input and

immunoprecipitation, respectively. Chromatin was incubated overnight at 4 ºC with ChIP-grade antibodies (Table M1). A negative control with IgG rabbit or mouse antibody was used to calculate the background signal. Chromatin-antibody complexed were immunoprecipitated for 2 h with 30 µl of Dynabeads Protein A (rabbit)/G (mouse) (Invitrogen) at 4 ºC and washed once with wash buffer 1, once with wash buffer 2, once with wash buffer 3 and twice with 1X TE. Input and immunoprecipitate (IP) were then un-crosslinked in TE - 1% SDS at 65 ºC overnight, treated with proteinase K for 1 h at 37 ºC and phenol-chloroform purified. Finally, DNA was resuspended in 50 µl of MQ H_2O. Primer pairs used for amplification are listed Table M7. Samples were run in 7500 Fast Real-time PCR system (Applied Biosystem). Results were analyzed with 7500 System Software V2.0.6.

Crosslinking solution - formaldehyde: 50 mM HEPES pH 8, 0.1 M NaCl, 1 mM EDTA pH 8, 0.5 mM EGTA. Formaldehyde was added fresh, in the appropriate concentration to obtain a final concentration of 1% formaldehyde after the addition to the medium.

Cell lysis buffer 1: 5 mM PIPES pH 8, 85 mM KCl, 0.5% NP-40, 1 mM PMSF and protease inhibitor cocktail.

Nuclei lysis buffer 2: 1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1, 1 mM PMSF and protease inhibitor cocktail.

IP buffer: 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 167 mM NaCl.

Wash buffer 1: 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl.

Wash buffer 2: 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 500 mM NaCl.

Wash buffer 3: 0.25 M LiCl, 1% NP-40, 1% Sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.1.

ChIP quantification and normalization: Input and immunoprecipitate (IP) were eluted in 50 µl of MQ H_2O. 2 µl of 1:10 dilutions of the Input and 2 µl of IP were used for qPCR. A calibration curve with five 5-fold serial dilutions of a standard

DNA sample was calculated for taking into account the amplification efficiency (based on the slope of the standard curve) of each qPCR reaction. Considering amplification (AF) and dilution (DF) factor $(AF^{(40-Ct)})DF$, we calculate the absolute quantification of each sample. IP/Input ratios in the different regions were calculated and multiplied by 100. The IP signal of the IgG (negative control) was considered as background.

16. POLYMERASE CHAIN REACTION (PCR) ANALYSIS

16.1. Non-quantitative PCR

PCRs carried out to amplify DNA fragment to be cloned were performed following standard and manufacture's protocols with the polymerases described in Materials and Methods 2.3. DNA primers used are listed in Table M7.

16.2. Quantitative PCR analysis

Real-time quantitative PCRs (qPCRs) were performed on a 7500 Fast Real-Time PCR system (Applied Biosystems, Carlsbad, CA). For PCRs, 6 µl H_2O, 2 µl primer mixture (each 10 µM), 2 µl DNA and 10 µl SYBR® green qPCR Mix (Bio-rad) were used. The following PCR reaction was used: 1 cycle (10 minutes 95 °C), 40 cycles (15 s 95 °C and 1 minute 65 °C) and 1 dissociation cycle (15 s 95 °C, 1 minute 65 °C, 15 s 95 °C and 15 s 60 °C). DNA primers were designed using Primer express 3.0 Software (Applied Biosystems) and are listed in Table M7. qPCR primers were validated by qPCR by establishing that each pair of primers had the same amplification efficiency (the slope of the 10-fold serial dilutions of a calibration curve was between -3.3 and -3.4).

16.2.1. Reverse Transcription quantitative PCR (RT-PCR) analysis

Relative qPCRs were used to determine the relative mRNA levels in human cells. cDNA was obtained from total RNA extracted using RNeasy Mini Kit (Qiagen) (1 µg) by reverse transcription using QuantiTect Reverse transcription (Qiagen), where an optimized blend of oligo-dT and random primers is used. mRNA expression values were normalized to mRNA expression of the Hypozanthine phosphoRibosylTransferase (HPRT) or Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) housekeeping genes.

16.2.2. Primer pairs used for amplification

Primer pairs used for amplification are listed Table M7.

Table M7. DNA primers used in this study.

Non-quantitative PCR primers		
Primer	**Sequence 5' to 3'**	**Use**
VP 1.5 Fwd	GGACTTTCCAAAATGTCG	Sequencing of prey inserts
XL39 Rv	ATTAGGACAAGGCTGGTGGG	Sequencing of prey inserts
DDX47 Fwd	GAGACAGAGGTTGACAAGATCC	Sequencing of prey inserts
DDX47 Rv	AGTAGCAAAGCTGTTCTCTGG	Sequencing of prey inserts
T7 promoter	TAATACGACTCACTATAGGG	Sequencing of prey inserts
T7 terminator	GCTAGTTATTGCTCAGCGG	Sequencing of prey inserts
M13 Fwd	GTAAAACGACGGCCAGT	Sequencing of prey inserts
M13 Rv	CAGGAAACAGCTATGACC	Sequencing of prey inserts
K74A Fwd K74A Rv	TGGCTCTGGAGCGACAGGCGCC GTTTCTGCAAGCCCAATG	Plasmid construction: pCMV-DDX47-K74A-FLAG
E175A Fwd E175A Rv	GGTCATGGATGCAGCCGACCGA AAGTATTTGAGAGCTCTCAAGTTGAAAC	Plasmid construction: pCMV-DDX47-E175A-FLAG
DDX47 gateway Fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTGGC GGCACCCGAGGAACACGATTCTCCG	Plasmid construction: pT7-His6- DDX47-WT/K74A/175A
DDX47 gateway Rv	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTAC TAACGGCCTTTCCGCTTCTTCATTTTTCC	

Quantitative PCR primers		
Primer	**Sequence 5' to 3'**	**Use**
HDAC8 Fwd HDAC8 Rv	TGGGAGGAGGAGGCTATAACC CCGGTCAAGTATGTCCAGCAT	Relative mRNA expression
MECP2 Fwd MECP2 Rv	CGCTCTGCTGGGAAGTATGAT CCACTTTAGAGCGAAAGGCTTTT	Relative mRNA expression
DDX42 Fwd DDX42 Rv	CAGCAGATCCATGCAGAATGTAA ACGGCCACTGATCGAAGATT	Relative mRNA expression
DDX47 Fwd DDX47 Rv	TCCTACCCATTCCAAGGATTACA CGGAGCGCCCAGCTCTA	Relative mRNA expression
PSMC5 Fwd PSMC5 Rv	CGAGAACGGCGAGTCCAT TGACCTTGGCTACTGCCATCT	Relative mRNA expression
PSMD7 Fwd PSMD7 Rv	CAACCGAATCGGCAAGGTT TGCCATGACCCCAAAAGC	Relative mRNA expression

Quantitative PCR primers		
Primer	**Sequence 5' to 3'**	**Use**
APOE Fwd APOE Rv	AAGCTGGAGGAGCAGGCC ACTGGCGCTGCATGTCTTC	Relative mRNA expression
RPL13 Fwd RPL13 Rv	GGGAGCAAGGAAAGGGTCTTA ACAATTCTCCGAGTGCTTTCAAG	Relative mRNA expression
HPRT Fwd HPRT Rv	GGACTAATTATGGACAGGACTG TCCAGCAGGTCAGCAAAGAA	Relative mRNA expression
GAPDH Fwd GAPDH Rv	TGCACCACCAACTGCTTAGC GGCATGGACTGTGGTCATGAG	Relative mRNA expression
APOE Fwd APOE Rv	GGGAGCCCTATAATTGGACAAGT CCCGACTGCGCTTCTCA	DRIP ChIP
RPL13 Fwd RPL13 Rv	GCTTCCAGCACAGGACAGGTAT CACCCACTACCCGAGTTCAAG	DRIP ChIP
18S Fwd 18S Rv	GGCGTCCCCCAACTTCTTA GGGCATCACAGACCTGTTATTG	DRIP ChIP
28S Fws 28S Rv	GAATCCGCTAAGGAGTGTGTAACA CTCCAGCGCCATCCATTT	DRIP ChIP
5'ETS Fws 5'ETS Rv	CAGGCGTTCTCGTCTCCG CACCACATCGATCGAAGAGC	ChIP
IGS Fws IGS RV	GTTGACGTACAGGGTGGACTG GGAAGTTGTCTTCACGCCTGA	ChIP
PHLDA2 Fwd PHLDA2 Rv	CAGCGGAAGTCGATCTCCTT GAGCGCACGGGCAAGTAC	DRIP ChIP
YIF1A Fws YIF1A Rv	TCTCCCAATTGGCCACTGA TGTTGCCTTCTGCCGAATC	DRIP ChIP

17. PROTEIN ANALYSIS

17.1. Bacteria cell protein extraction

Pellet of bacteria was collected at 12 000 g for 5 min at 4ºC and resuspended using 1X Loading Buffer. Prior to gel loading, samples were boiled for 5 min and centrifuged 5 min at 3000 rpm at RT.

17.2. Human cell protein extraction

Pellet of HeLa cells was collected using accutase, washed in cold PBS and kept in ice. Proteins were extracted by adding 1X Lammeli buffer (10 µl/75 000 cells) with gently pipetting up and down. Samples were centrifuged 5 min at 3000 rpm. The lysate was sonicated on the maximum intensity setting, with pulses of 30 s on and 30 s off for 5 min in Bioruptor (Diagenode). Prior to gel loading samples were boiled for 5 min and centrifuged 3 min at 3000 rpm at RT.

4X Lammeli buffer: 200 mM Tris-HCl, 40% glycerol, 8% SDS, 0.4% Bromophenol Blue, 400 mM β-mercaptoethanol.

17.3. SDS-PAGE

Proteins were separated in 29:1 acrylamide:bis-acrylamide gels with concentrations appropriate to the molecular size of the proteins of interest or in 4-20% gradient SDSPAGE CriterionTM TGX TM Precast Gels (BioRad) and SDS-PAGE was performed according to previously described method (Laemmli, 1970). Electrophoresis were performed in a Mini-PROTEAN 3 Cell in Running Buffer at 100 V. Page Ruler Prestained Protein Ladder (26617, ThermoFisher) and PageRuler Plus Prestained Protein Ladder (26620, ThermoFisher) were used as protein markers.

Running buffer: 25 mM Tris base pH 8.3, 194 mM glycine, 0.1% SDS.

17.4. Western Blot analysis

For Western blot, proteins were wet-transferred using Trans-Blot system (Biorad) for 1 h at 300 mA in 1X Transfer Buffer with 20% methanol or o/n at 30 V at 4 °C. Membranes were stained with Ponceau S (0.1% w/v Ponceau -SIGMA- in 5% acetic acid) to check protein loading and correct transference.

5X Transfer buffer: 6 g/L Tris base, 28.8 g/L glycine, 0.5% SDS.

17.5. Non-fluorescence WB

Proteins were transferred to a nitrocellulose membrane (Hybond-ECL, GE Healthcare). Membranes were blocked with 1X TBS - 0.05% Tween 20 - 5% milk for 1 h. Primary antibodies were incubated o/n at 4ºC at the indicated concentrations (Table M1) in 1x TBS - 0.05% Tween 20 - 5% milk. After three washes of 5 min each one in 1x TBS – 0.05% Tween 20, membranes were incubated with the corresponding secondary antibodies conjugated with the horseradish peroxidase at the indicated concentrations (Table M2) in 1x TBS, 0.05% Tween-20 with 5% milk for 1 h at room temperature and washed again. Finally, SuperSignal West Pico Plus (Thermo) was used for chemiluminescence detection. Protein levels were calculated with Chemidoc Imaging system using as loading controls: vinculin or actin antibodies.

17.6. Expression of His6-DDX47-WT and mutant forms

BLB21 Rosetta Electrocompetent Cells (70954-3, Merck) were electroporated with 0.5 µg of pT7-His6-DDX47-WT, K94A and E175A plasmids, selected with ampicillin (75 µg/ml) and chloramphenicol (34 µg/ml). Plates were incubated at

37 ºC overnight. A single colony was inoculated into 5 ml LB + ampicillin (75 µg/ml) + chloramphenicol (34 µg/ml) culture and incubated at 37 ºC with shaking at 200 rpm overnight. Overnight cultures were diluted to an optical density (OD_{600nm}) of 0.4. At this moment, we induced His6-DDX47 expression by adding a final concentration of 0.2 mM IPTG to the culture. His6-DDX47 was expressed at 37 ºC for 3h or at 16ºC overnight with 200 rpm shaking. For bacteria cells protein extraction and western blot analysis, see Materials and Methods 17.1. For protein staining, acrylamide gels were incubated with Coomassie Brilliant Blue for 1h with gentle shaking at RT and then destained with Coomassie destaining solution for 15 min.

Coomassie solution: 2.5 g Coomassie Brilliant Blue, 400 ml methanol, 100 ml glacial acetic acid and 500 ml MQ H_2O.

Coomassie destaining solution: 40 ml methanol, 100 ml glacial acetic acid and 500 ml MQ H_2O.

REFERENCES

Abdelhaleem, M. (2004). Do human RNA helicases have a role in cancer? In *Biochimica et Biophysica Acta - Reviews on Cancer* (Vol. 1704, Issue 1, pp. 37–46). Biochim Biophys Acta. https://doi.org/10.1016/j.bbcan.2004.05.001

Aguilera, A. (2002). The connection between transcription and genomic instability. In *EMBO Journal* (Vol. 21, Issue 3, pp. 195–201). EMBO J. https://doi.org/10.1093/emboj/21.3.195

Aguilera, A., & García-Muse, T. (2012). R Loops: From Transcription Byproducts to Threats to Genome Stability. In *Molecular Cell* (Vol. 46, Issue 2, pp. 115–124). Cell Press. https://doi.org/10.1016/j.molcel.2012.04.009

Aguilera, A., & García-Muse, T. (2013). Causes of genome instability. *Annual Review of Genetics*, *47*(July), 1–32. https://doi.org/10.1146/annurev-genet-111212-133232

Aguilera, A., & Gómez-González, B. (2008). Genome instability: A mechanistic view of its causes and consequences. *Nature Reviews Genetics*, *9*(3), 204–217. https://doi.org/10.1038/nrg2268

Akamatsu, Y., & Kobayashi, T. (2015). The Human RNA Polymerase I Transcription Terminator Complex Acts as a Replication Fork Barrier That Coordinates the Progress of Replication with rRNA Transcription Activity. *Molecular and Cellular Biology*, *35*(10), 1871–1881. https://doi.org/10.1128/mcb.01521-14

Amir, R. E., Van Den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl- CpG-binding protein 2. *Nature Genetics*, *23*(2), 185–188. https://doi.org/10.1038/13810

Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., & Niehrs, C. (2019). GADD45A binds R-loops and recruits TET1 to CpG island promoters. In *Nature Genetics* (Vol. 51, Issue 2, pp. 217–223). Nature Publishing Group. https://doi.org/10.1038/s41588-018-0306-6

AW, C., DY, W., JR, M., DL, C., G, Z., & HW, G. (2020). MeCP2 Represses Enhancers through Chromosome Topology-Associated DNA Methylation. *Molecular Cell*, *77*(2), 279-293.e8. https://doi.org/10.1016/J.MOLCEL.2019.10.033

Awasthi, S., Verma, M., Mahesh, A., Khan, M. I. K., Govindaraju, G., Rajavelu, A., Chavali, P. L., Chavali, S., & Dhayalan, A. (2018). DDX49 is an RNA helicase that affects translation by regulating mRNA export and the levels of pre-ribosomal RNA. *Nucleic Acids Research*, *46*(12), 6304–6317. https://doi.org/10.1093/nar/gky231

Bagge, J., Oestergaard, V. H., & Lisby, M. (2021). Functions of TopBP1 in preserving genome integrity during mitosis. *Seminars in Cell & Developmental Biology*, *113*, 57–64. https://doi.org/10.1016/J.SEMCDB.2020.08.009

Barroso, S., Herrera- Moyano, E., Muñoz, S., García- Rubio, M., Gómez-

González, B., & Aguilera, A. (2019). The DNA damage response acts as a safeguard against harmful DNA–RNA hybrids of different origins. *EMBO Reports*, *20*(9). https://doi.org/10.15252/embr.201847250

Basu, U., Meng, F. L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C. R., Wesemann, D. R., Januszyk, K., Gregory, R. I., Deng, H., Lima, C. D., & Alt, F. W. (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell*, *144*(3), 353–363. https://doi.org/10.1016/j.cell.2011.01.001

Bayona-Feliu, A., Barroso, S., Muñoz, S., & Aguilera, A. (2021). The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated transcription–replication conflicts. *Nature Genetics*. https://doi.org/10.1038/s41588-021-00867-2

Beletskii, A., & Bhagwat, A. S. (1996). Transcription-induced mutations: Increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13919–13924. https://doi.org/10.1073/pnas.93.24.13919

Beletskii, A., & Bhagwat, A. S. (1998). Correlation between transcription and C to T mutations in the non-transcribed DNA strand. *Biological Chemistry*, *379*(4–5), 549–551.

Belotserkovskii, B. P., Mirkin, S. M., & Hanawalt, P. C. (2013). DNA sequences that interfere with transcription: Implications for genome function and stability. In *Chemical Reviews* (Vol. 113, Issue 11, pp. 8620–8637). Chem Rev. https://doi.org/10.1021/cr400078y

Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. In *Nature Reviews Genetics* (Vol. 15, Issue 3, pp. 163–175). Nat Rev Genet. https://doi.org/10.1038/nrg3662

Bermejo, R., Lai, M. S., & Foiani, M. (2012). Preventing Replication Stress to Maintain Genome Stability: Resolving Conflicts between Replication and Transcription. In *Molecular Cell* (Vol. 45, Issue 6, pp. 710–718). Mol Cell. https://doi.org/10.1016/j.molcel.2012.03.001

Berry, J., Weber, S. C., Vaidya, N., Haataja, M., Brangwynne, C. P., & Weitz, D. A. (2015). RNA transcription modulates phase transition-driven nuclear body assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(38), E5237–E5245. https://doi.org/10.1073/pnas.1509317112

Bhatia, V., Barroso, S. I., García-Rubio, M. L., Tumini, E., Herrera-Moyano, E., & Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*, *511*(7509), 362–365. https://doi.org/10.1038/nature13374

Boguslawski, S. J., Smith, D. E., Michalak, M. A., Mickelson, K. E., Yehle, C. O., Patterson, W. L., & Carrico, R. J. (1986). Characterization of monoclonal antibody to DNA · RNA and its application to immunodetection of hybrids.

Journal of Immunological Methods, *89*(1), 123–130. https://doi.org/10.1016/0022-1759(86)90040-2

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., & Guil, S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(18), 5785–5790. https://doi.org/10.1073/pnas.1421197112

Boxer, L. D., Renthal, W., Greben, A. W., Griffith, E. C., Bonev, B., & Greenberg Correspondence, M. E. (2020). MeCP2 Represses the Rate of Transcriptional Initiation of Highly Methylated Long Genes. *Molecular Cell*, *77*, 294–309. https://doi.org/10.1016/j.molcel.2019.10.032

Brero, A., Easwaran, H. P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H., & Cardoso, M. C. (2005). Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *Journal of Cell Biology*, *169*(5), 733–743. https://doi.org/10.1083/jcb.200502062

Brock, R. D. (1971). Differential mutation of the beta-galactosidase gene of Escherichia coli. *Mutation Research*, *11*(2), 181–186.

Cai, W., Chen, Z. X., Rane, G., Singh, S. S., Choo, Z., Wang, C., Yuan, Y., Tan, T. Z., Arfuso, F., Yap, C. T., Pongor, L. S., Yang, H., Lee, M. B., Goh, B. C., Sethi, G., Benoukraf, T., Tergaonkar, V., & Kumar, A. P. (2017). Wanted DEAD/H or alive: Helicases winding up in cancers. In *Journal of the National Cancer Institute* (Vol. 109, Issue 6). Oxford University Press. https://doi.org/10.1093/jnci/djw278

Calo, E., Flynn, R. A., Martin, L., Spitale, R. C., Chang, H. Y., & Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature*, *518*(7538), 249. https://doi.org/10.1038/NATURE13923

Castellano-Pozo, M., García-Muse, T., & Aguilera, A. (2012). The Caenorhabditis elegans THO Complex Is Required for the Mitotic Cell Cycle and Development. *PLoS ONE*, *7*(12). https://doi.org/10.1371/journal.pone.0052447

Castellano-Pozo, M., Santos-Pereira, J., Rondón, A. G., Barroso, S., Andújar, E., Pérez-Alegre, M., García-Muse, T., & Aguilera, A. (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Molecular Cell*, *52*(4), 583–590. https://doi.org/10.1016/j.molcel.2013.10.006

Cerritelli, S. M., & Crouch, R. J. (2009). Ribonuclease H: The enzymes in eukaryotes. In *FEBS Journal* (Vol. 276, Issue 6, pp. 1494–1505). FEBS J. https://doi.org/10.1111/j.1742-4658.2009.06908.x

Cerritelli, S. M., & Crouch, R. J. (2019). RNases H: Multiple roles in maintaining genome integrity. In *DNA Repair* (Vol. 84). Elsevier B.V. https://doi.org/10.1016/j.dnarep.2019.102742

Chan, Y. A., Aristizabal, M. J., Lu, P. Y. T., Luo, Z., Hamza, A., Kobor, M. S., Stirling, P. C., & Hieter, P. (2014). Genome-Wide Profiling of Yeast

DNA:RNA Hybrid Prone Sites with DRIP-Chip. *PLoS Genetics*, *10*(4), 1004288. https://doi.org/10.1371/journal.pgen.1004288

Chaudhuri, J., & Alt, F. W. (2004). Class-switch recombination: Interplay of transcription, DNA deamination and DNA repair. In *Nature Reviews Immunology* (Vol. 4, Issue 7, pp. 541–552). Nature Publishing Group. https://doi.org/10.1038/nri1395

Chávez, S., & Aguilera, A. (1997). The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes and Development*, *11*(24), 3459–3470. https://doi.org/10.1101/gad.11.24.3459

Chédin, F. (2016). Nascent Connections: R-Loops and Chromatin Patterning. In *Trends in Genetics* (Vol. 32, Issue 12, pp. 828–838). Elsevier Ltd. https://doi.org/10.1016/j.tig.2016.10.002

Chen, Liang, Chen, J.-Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., Zhou, Y., Zhang, D.-E., & Fu, X.-D. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Molecular Cell*, *68*(4), 745. https://doi.org/10.1016/J.MOLCEL.2017.10.008

Chen, Lin, Chen, K., Lavery, L. A., Baker, S. A., Shaw, C. A., Li, W., & Zoghbi, H. Y. (2015). MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), 5509–5514. https://doi.org/10.1073/pnas.1505909112

Chen, P. B., Chen, H. V., Acharya, D., Rando, O. J., & Fazzio, T. G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nature Structural and Molecular Biology*, *22*(12), 999–1007. https://doi.org/10.1038/nsmb.3122

Cheng, T.-L., Wang, Z., Liao, Q., Zhu, Y., Zhou, W.-H., Xu, W., & Qiu, Z. (2014). MeCP2 Suppresses Nuclear MicroRNA Processing and Dendritic Growth by Regulating the DGCR8/Drosha Complex. *Developmental Cell*, *28*(5), 547–560. https://doi.org/10.1016/J.DEVCEL.2014.01.032

Chon, H., Sparks, J. L., Rychlik, M., Nowotny, M., Burgers, P. M., Crouch, R. J., & Cerritelli, S. M. (2013). RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Research*, *41*(5), 3130–3143. https://doi.org/10.1093/nar/gkt027

Ciccia, A., & Elledge, S. J. (2010). The DNA Damage Response: Making It Safe to Play with Knives. In *Molecular Cell* (Vol. 40, Issue 2, pp. 179–204). Mol Cell. https://doi.org/10.1016/j.molcel.2010.09.019

Cloutier, S. C., Wang, S., Ma, W. K., Al Husini, N., Dhoondia, Z., Ansari, A., Pascuzzi, P. E., & Tran, E. J. (2016). Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation. *Molecular Cell*, *61*(3), 393–404. https://doi.org/10.1016/j.molcel.2015.12.024

Cordin, O., Banroques, J., Tanner, N. K., & Linder, P. (2006). The DEAD-box

protein family of RNA helicases. In *Gene* (Vol. 367, Issues 1–2, pp. 17–37). Elsevier. https://doi.org/10.1016/j.gene.2005.10.019

Cristini, A., Groh, M., Kristiansen, M. S., & Gromak, N. (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Reports*, *23*(6), 1891–1905. https://doi.org/10.1016/j.celrep.2018.04.025

Crossley, M. P., Bocek, M., & Cimprich, K. A. (2019). R-Loops as Cellular Regulators and Genomic Threats. In *Molecular Cell* (Vol. 73, Issue 3, pp. 398–411). Cell Press. https://doi.org/10.1016/j.molcel.2019.01.024

Deardorff, M. A., Bando, M., Nakato, R., Watrin, E., Itoh, T., Minamino, M., Saitoh, K., Komata, M., Katou, Y., Clark, D., Cole, K. E., De Baere, E., Decroos, C., Di Donato, N., Ernst, S., Francey, L. J., Gyftodimou, Y., Hirashima, K., Hullings, M., … Shirahige, K. (2012). HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature*, *489*(7415), 313–317. https://doi.org/10.1038/nature11316

Domínguez-Sánchez, M. S., Barroso, S., Gómez-González, B., Luna, R., & Aguilera, A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genetics*, *7*(12), 1002386. https://doi.org/10.1371/journal.pgen.1002386

Drolet, M., Phoenix, P., Menzel, R., Massé, E., Liu, L. F., & Crouch, R. J. (1995). Overexpression of RNase H partially complements the growth defect of an Escherichia coli ΔtopA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(8), 3526–3530. https://doi.org/10.1073/pnas.92.8.3526

Duch, A., Felipe-Abrio, I., Barroso, S., Yaakov, G., García-Rubio, M., Aguilera, A., De Nadal, E., & Posas, F. (2013). Coordinated control of replication and transcription by a SAPK protects genomic integrity. *Nature*, *493*(7430), 116–121. https://doi.org/10.1038/nature11675

Dumelie, J. G., & Jaffrey, S. R. (2017). Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *ELife*, *6*. https://doi.org/10.7554/eLife.28306

Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., & Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes & Development*, *18*(13), 1618–1629. https://doi.org/10.1101/GAD.1200804

Eberharter, A., Ferrari, S., Längst, G., Straub, T., Imhof, A., Varga-Weisz, P., Wilm, M., & Becker, P. B. (2001). Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. *EMBO Journal*, *20*(14), 3781–3788. https://doi.org/10.1093/emboj/20.14.3781

El Hage, A., French, S. L., Beyer, A. L., & Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes and Development*, *24*(14), 1546–1558. https://doi.org/10.1101/gad.573310

El Hage, A., Webb, S., Kerr, A., & Tollervey, D. (2014). Genome-Wide Distribution of RNA-DNA Hybrids Identifies RNase H Targets in tRNA Genes, Retrotransposons and Mitochondria. *PLoS Genetics*, *10*(10), 1004716. https://doi.org/10.1371/journal.pgen.1004716

Enikanolaiye, A., Ruston, J., Zeng, R., Taylor, C., Schrock, M., Buchovecky, C. M., Shendure, J., Acar, E., & Justice, M. J. (2020). Suppressor mutations in Mecp2-null mice implicate the DNA damage response in Rett syndrome pathology. *Genome Research*, *30*(4), 540–552. https://doi.org/10.1101/gr.258400.119

Fatica, A., & Tollervey, D. (2002). Making ribosomes. In *Current Opinion in Cell Biology* (Vol. 14, Issue 3, pp. 313–318). Elsevier Ltd. https://doi.org/10.1016/S0955-0674(02)00336-8

Ferdous, A., Gonzalez, F., Sun, L., Kodadek, T., & Johnston, S. A. (2001). The 19S regulatory particle of the proteasome is required for efficient transcription elongation by RNA polymerase II. *Molecular Cell*, *7*(5), 981–991. https://doi.org/10.1016/S1097-2765(01)00250-7

Gaillard, H., & Aguilera, A. (2016). Transcription as a Threat to Genome Integrity. *Annual Review of Biochemistry*, *85*(March), 291–317. https://doi.org/10.1146/annurev-biochem-060815-014908

Gaillard, H., García-Muse, T., & Aguilera, A. (2015). *Replication stress and cancer*. https://doi.org/10.1038/nrc3916

Gaillard, H., Herrera-Moyano, E., & Aguilera, A. (2013). Transcription-associated genome instability. In *Chemical Reviews* (Vol. 113, Issue 11, pp. 8638–8661). https://doi.org/10.1021/cr400017y

Gamalinda, M., & Woolford, J. L. (2015). *Paradigms of ribosome synthesis: Lessons learned from ribosomal proteins*. https://doi.org/10.4161/21690731.2014.975018

Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J. L., & Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes and Development*, *25*(19), 2041–2056. https://doi.org/10.1101/gad.17010011

García-Benítez, F., Gaillard, H., & Aguilera, A. (2017). Physical proximity of chromatin to nuclear pores prevents harmful R loop accumulation contributing to maintain genome stability. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(41), 10942–10947. https://doi.org/10.1073/pnas.1707845114

García-Muse, T., & Aguilera, A. (2019). R Loops: From Physiological to Pathological Roles. In *Cell* (Vol. 179, Issue 3, pp. 604–618). Cell Press. https://doi.org/10.1016/j.cell.2019.08.055

García-Pichardo, D., Cañas, J. C., García-Rubio, M. L., Gómez-González, B., Rondón, A. G., & Aguilera, A. (2017). Histone Mutants Separate R Loop Formation from Genome Instability Induction. *Molecular Cell*, *66*(5), 597-609.e5. https://doi.org/10.1016/j.molcel.2017.05.014

García-Rubio, M., Barroso, S. I., & Aguilera, A. (2018). Detection of DNA-RNA hybrids in vivo. In *Methods in Molecular Biology* (Vol. 1672, pp. 347–361). Humana Press Inc. https://doi.org/10.1007/978-1-4939-7306-4_24

García-Rubio, M. L., Pérez-Calero, C., Barroso, S. I., Tumini, E., Herrera-Moyano, E., Rosado, I. V., & Aguilera, A. (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genetics*, *11*(11), 1005674. https://doi.org/10.1371/JOURNAL.PGEN.1005674

Garcia, I., Albring, M. J., & Uhlenbeck, O. C. (2012). Duplex destabilization by four ribosomal DEAD-box proteins. *Biochemistry*, *51*(50), 10109–10118. https://doi.org/10.1021/bi301172s

Gavaldá, S., Gallardo, M., Luna, R., & Aguilera, A. (2013). R-Loop Mediated Transcription-Associated Recombination in trf4Δ Mutants Reveals New Links between RNA Surveillance and Genome Integrity. *PLoS ONE*, *8*(6), e65541. https://doi.org/10.1371/journal.pone.0065541

Geng, F., Wenzel, S., & Tansey, W. P. (2012). Ubiquitin and proteasomes in transcription. *Annual Review of Biochemistry*, *81*, 177–201. https://doi.org/10.1146/annurev-biochem-052110-120012

Georgel, P. T., Horowitz-Scherer, R. A., Adkins, N., Woodcock, C. L., Wade, P. A., & Hansen, J. C. (2003). Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *Journal of Biological Chemistry*, *278*(34), 32181–32188. https://doi.org/10.1074/jbc.M305308200

Ghosh, R. P., Horowitz-Scherer, R. A., Nikitina, T., Shlyakhtenko, L. S., & Woodcock, C. L. (2010). MeCP2 Binds Cooperatively to Its Substrate and Competes with Histone H1 for Chromatin Binding Sites. *Molecular and Cellular Biology*, *30*(19), 4656–4670. https://doi.org/10.1128/mcb.00379-10

Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I., & Chédin, F. (2013). GC skew at the 59 and 39 ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Research*, *23*(10), 1590–1600. https://doi.org/10.1101/gr.158436.113

Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., & Chédin, F. (2012). R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Molecular Cell*, *45*(6), 814–825. https://doi.org/10.1016/j.molcel.2012.01.017

Glaich, O., Parikh, S., Bell, R. E., Mekahel, K., Donyo, M., Leader, Y., Shayevitch, R., Sheinboim, D., Yannai, S., Hollander, D., Melamed, Z., Lev-Maor, G., Ast, G., & Levy, C. (2019). DNA methylation directs microRNA biogenesis in mammalian cells. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-13527-1

Gómez-González, B., & Aguilera, A. (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(20), 8409–8414. https://doi.org/10.1073/pnas.0702836104

Gómez-González, B., & Aguilera, A. (2009). R-loops do not accumulate in

transcription-defective hpr1-101 mutants: Implications for the functional role of THO/TREX. *Nucleic Acids Research*, *37*(13), 4315–4321. https://doi.org/10.1093/nar/gkp385

Gómez-González, B., & Aguilera, A. (2019). Transcription-mediated replication hindrance: A major driver of genome instability. In *Genes and Development* (Vol. 33, Issues 15–16, pp. 1008–1026). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/gad.324517.119

Gómez-González, B., & Aguilera, A. (2020). Looping the (R) Loop in DSB Repair via RNA Methylation. *Molecular Cell*, *79*(3), 361–362. https://doi.org/10.1016/J.MOLCEL.2020.07.015

Gómez-González, B., García-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marín, A., Foiani, M., & Aguilera, A. (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO Journal*, *30*(15), 3106–3119. https://doi.org/10.1038/emboj.2011.206

González-Aguilera, C., Tous, C., Gómez-González, B., Huertas, P., Luna, R., & Aguilera, A. (2008). The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Molecular Biology of the Cell*, *19*(10), 4310–4318. https://doi.org/10.1091/mbc.E08-04-0355

Good, K. V., Vincent, J. B., & Ausió, J. (2021). MeCP2: The Genetic Driver of Rett Syndrome Epigenetics. In *Frontiers in Genetics* (Vol. 12). Frontiers Media S.A. https://doi.org/10.3389/fgene.2021.620859

Granneman, S., Bernstein, K. A., Bleichert, F., & Baserga, S. J. (2006). Comprehensive Mutational Analysis of Yeast DEXD/H Box RNA Helicases Required for Small Ribosomal Subunit Synthesis. *Molecular and Cellular Biology*, *26*(4), 1183–1194. https://doi.org/10.1128/mcb.26.4.1183-1194.2006

Grunseich, C., Wang, I. X., Watts, J. A., Burdick, J. T., Guber, R. D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K., Schindler, A. B., Edwards, N., Ray-Chaudhury, A., Yao, J., Lehky, T., Piszczek, G., Crain, B., Fischbeck, K. H., & Cheung, V. G. (2018). Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. *Molecular Cell*, *69*(3), 426-437.e7. https://doi.org/10.1016/j.molcel.2017.12.030

Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T., & Cimprich, K. A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell*, *170*(4), 774-786.e19. https://doi.org/10.1016/j.cell.2017.07.043

Hamperl, S., & Cimprich, K. A. (2016). Conflict Resolution in the Genome: How Transcription and Replication Make It Work. In *Cell* (Vol. 167, Issue 6, pp. 1455–1467). Cell Press. https://doi.org/10.1016/j.cell.2016.09.053

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology*, *166*(4), 557–580.

https://doi.org/10.1016/S0022-2836(83)80284-8

Hanawalt, P. C., & Spivak, G. (2008). Transcription-coupled DNA repair: Two decades of progress and surprises. In *Nature Reviews Molecular Cell Biology* (Vol. 9, Issue 12, pp. 958–970). Nat Rev Mol Cell Biol. https://doi.org/10.1038/nrm2549

Hartono, S. R., Malapert, A., Legros, P., Bernard, P., Chédin, F., & Vanoosthuyse, V. (2018). The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. *Journal of Molecular Biology*, *430*(3), 272–284. https://doi.org/10.1016/j.jmb.2017.12.016

Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., McKinney, K. M., Eaton, M. L., Kellis, M., Hill, S. J., Parmigiani, G., Proudfoot, N. J., & Livingston, D. M. (2015). BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Molecular Cell*, *57*(4), 636–647. https://doi.org/10.1016/j.molcel.2015.01.011

Helmrich, A., Ballarino, M., Nudler, E., & Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. In *Nature Structural and Molecular Biology* (Vol. 20, Issue 4, pp. 412–418). https://doi.org/10.1038/nsmb.2543

Hendriks, G., Calléja, F., Besaratinia, A., Vrieling, H., Pfeifer, G. P., Mullenders, L. H. F., Jansen, J. G., & de Wind, N. (2010). Transcription-Dependent Cytosine Deamination Is a Novel Mechanism in Ultraviolet Light-Induced Mutagenesis. *Current Biology*, *20*(2), 170–175. https://doi.org/10.1016/j.cub.2009.11.061

Herman, R. K., & Dworkin, N. B. (1971). Effect of gene induction on the rate of mutagenesis by ICR-191 in Escherichia coli. *Journal of Bacteriology*, *106*(2), 543–550. https://doi.org/10.1128/jb.106.2.543-550.1971

Herrera-Moyano, E., Mergui, X., García-Rubio, M. L., Barroso, S., & Aguilera, A. (2014). The yeast and human FACT chromatinreorganizing complexes solve R-loopmediated transcription-replication conflicts. *Genes and Development*, *28*(7), 735–748. https://doi.org/10.1101/gad.234070.113

Hodroj, D., Serhal, K., & Maiorano, D. (2017). Ddx19 links mRNA nuclear export with progression of transcription and replication and suppresses genomic instability upon DNA damage in proliferating cells. In *Nucleus* (Vol. 8, Issue 5, pp. 489–495). Taylor and Francis Inc. https://doi.org/10.1080/19491034.2017.1348448

Hoeijmakers, J. H. J. (2009). *DNA Damage, Aging, and Cancer*. 1475–1485.

Huertas, P., & Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular Cell*, *12*(3), 711–721. https://doi.org/10.1016/j.molcel.2003.08.010

Huh, M. S., Ivanochko, D., Hashem, L. E., Curtin, M., Delorme, M., Goodall, E., Yan, K., & Picketts, D. J. (2016). Stalled replication forks within

heterochromatin require ATRX for protection. *Cell Death and Disease*, *7*(5), e2220. https://doi.org/10.1038/cddis.2016.121

Ide, S., Miyazaki, T., Maki, H., & Kobayashi, T. (2010). Abundance of ribosomal RNA gene copies maintains genome integrity. *Science*, *327*(5966), 693–696. https://doi.org/10.1126/science.1179044

Ip, J. P. K., Mellios, N., & Sur, M. (2018). Rett syndrome: Insights into genetic, molecular and circuit mechanisms. *Nature Reviews Neuroscience*, *19*(6), 368–382. https://doi.org/10.1038/s41583-018-0006-3

Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. In *Nature* (Vol. 461, Issue 7267, pp. 1071–1078). Europe PMC Funders. https://doi.org/10.1038/nature08467

Jao, C. Y., & Salic, A. (2008). Exploring RNA transcription and turnover in vivo by using click chemistry. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(41), 15779–15784. https://doi.org/10.1073/pnas.0808480105

Jeffery, L., & Nakielny, S. (2004). Components of the DNA Methylation System of Chromatin Control Are RNA-binding Proteins*. *Journal of Biological Chemistry*, *279*, 49479–49487. https://doi.org/10.1074/jbc.M409070200

Jeggo, P. A., Pearl, L. H., & Carr, A. M. (2016). DNA repair, genome stability and cancer: A historical perspective. *Nature Reviews Cancer*, *16*(1), 35–42. https://doi.org/10.1038/nrc.2015.4

Jiang, F., Taylor, D. W., Chen, J. S., Kornfeld, J. E., Zhou, K., Thompson, A. J., Nogales, E., & Doudna, J. A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, *351*(6275), 867–871. https://doi.org/10.1126/science.aad8282

Jinks-Robertson, S., & Bhagwat, A. S. (2014). Transcription-Associated mutagenesis. *Annual Review of Genetics*, *48*, 341–359. https://doi.org/10.1146/annurev-genet-120213-092015

Kaiser, F. J., Ansari, M., Braunholz, D., Gil-Rodríguez, M. C., Decroos, C., Wilde, J. J., Fincher, C. T., Kaur, M., Bando, M., Amor, D. J., Atwal, P. S., Bahlo, M., Bowman, C. M., Bradley, J. J., Brunner, H. G., Clark, D., Campo, M. Del, Di Donato, N., Diakumis, P., … Deardorff, M. A. (2014). Loss-of-function HDAC8 mutations cause a phenotypic spectrum of Cornelia de Lange syndrome-like features, ocular hypertelorism, large fontanelle and X-linked inheritance. *Human Molecular Genetics*, *23*(11), 2888–2900. https://doi.org/10.1093/hmg/ddu002

Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. In *EMBO Journal* (Vol. 20, Issue 14, pp. 3617–3622). EMBO J. https://doi.org/10.1093/emboj/20.14.3617

Kobayashi, T. (2003). The Replication Fork Barrier Site Forms a Unique Structure with Fob1p and Inhibits the Replication Fork. *Molecular and Cellular Biology*, *23*(24), 9178–9188. https://doi.org/10.1128/mcb.23.24.9178-9188.2003

Kobayashi, T. (2014). Ribosomal RNA gene repeats, their stability and cellular senescence. In *Proceedings of the Japan Academy Series B: Physical and Biological Sciences* (Vol. 90, Issue 4, pp. 119–129). Japan Academy. https://doi.org/10.2183/pjab.90.119

Kobayashi, T., Heck, D. J., Nomura, M., & Horiuchi, T. (1998). Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: Requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes and Development*, *12*(24), 3821–3830. https://doi.org/10.1101/gad.12.24.3821

Kogoma, T. (1997). Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiology and Molecular Biology Reviews : MMBR*, *61*(2), 212–238. https://doi.org/10.1128/.61.2.212-238.1997

Kouzine, F., Wojtowicz, D., Baranello, L., Yamane, A., Nelson, S., Resch, W., Kieffer-Kwon, K. R., Benham, C. J., Casellas, R., Przytycka, T. M., & Levens, D. (2017). Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Systems*, *4*(3), 344-356.e7. https://doi.org/10.1016/j.cels.2017.01.013

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, *227*(5259), 680–685. https://doi.org/10.1038/227680a0

Lang, K. S., Hall, A. N., Merrikh, C. N., Ragheb, M., Tabakh, H., Pollock, A. J., Woodward, J. J., Dreifus, J. E., & Merrikh, H. (2017). Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. *Cell*, *170*(4), 787-799.e18. https://doi.org/10.1016/j.cell.2017.07.044

Leung, J. W. C., Ghosal, G., Wang, W., Shen, X., Wang, J., Li, L., & Chen, J. (2013). Alpha thalassemia/mental retardation syndrome X-linked gene product ATRX is required for proper replication restart and cellular resistance to replication stress. *Journal of Biological Chemistry*, *288*(9), 6342–6350. https://doi.org/10.1074/jbc.M112.411603

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F., & Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA. *Cell*, *69*(6), 905–914. https://doi.org/10.1016/0092-8674(92)90610-O

Li, C. H., Coffey, E. L., Dall, A., Hannett, N. M., Tang, X., Henninger, J. E., Platt, J. M., Oksuz, O., Zamudio, A. V, Afeyan, L. K., Schuijers, J., Shawn Liu, X., Markoulaki, S., Lungjangwa, T., LeRoy, G., Svoboda, D. S., Wogram, E., Ihn Lee, T., Jaenisch, R., & Young, R. A. (2020). MeCP2 links heterochromatin condensates and neurodevelopmental disease. *440* | *Nature* |, *586*. https://doi.org/10.1038/s41586-020-2574-4

Li, L., Germain, D. R., Poon, H.-Y., Hildebrandt, M. R., Monckton, E. A., McDonald, D., Hendzel, M. J., & Godbout, R. (2016). DEAD Box 1 Facilitates Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks. *Molecular and Cellular Biology*, *36*(22), 2794–2810.

https://doi.org/10.1128/mcb.00415-16

Li, X., & Manley, J. L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell*, *122*(3), 365–378. https://doi.org/10.1016/j.cell.2005.06.008

Linder, P., & Jankowsky, E. (2011). From unwinding to clamping ĝ€" the DEAD box RNA helicase family. In *Nature Reviews Molecular Cell Biology* (Vol. 12, Issue 8, pp. 505–516). Nature Publishing Group. https://doi.org/10.1038/nrm3154

Liu, B., & Alberts, B. M. (1995). Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science*, *267*(5201), 1131–1137. https://doi.org/10.1126/science.7855590

Long, S. W., Ooi, J. Y. Y., Yau, P. M., & Jones, P. L. (2011). A brain-derived MeCP2 complex supports a role for MeCP2 in RNA processing. *Bioscience Reports*, *31*(5), 333–343. https://doi.org/10.1042/BSR20100124

Luna, R., Jimeno, S., Marín, M., Huertas, P., García-Rubio, M., & Aguilera, A. (2005). Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. *Molecular Cell*, *18*(6), 711–722. https://doi.org/10.1016/j.molcel.2005.05.001

Luna, R., Rondón, A. G., Pérez-Calero, C., Salas-Armenteros, I., & Aguilera, A. (2019). The THO complex as a paradigm for the prevention of cotranscriptional R-Loops. *Cold Spring Harbor Symposia on Quantitative Biology*, *84*, 105–114. https://doi.org/10.1101/sqb.2019.84.039594

Madireddy, A., Kosiyatrakul, S. T., Boisvert, R. A., Herrera-Moyano, E., García-Rubio, M. L., Gerhardt, J., Vuono, E. A., Owen, N., Yan, Z., Olson, S., Aguilera, A., Howlett, N. G., & Schildkraut, C. L. (2016). FANCD2 Facilitates Replication through Common Fragile Sites. *Molecular Cell*, *64*(2), 388–404. https://doi.org/10.1016/j.molcel.2016.09.017

Marano, D., Fioriniello, S., Fiorillo, F., Gibbons, R. J., D'esposito, M., & Ragione, F. Della. (2019). ATRX contributes to MECP2-mediated pericentric heterochromatin organization during neural differentiation. *International Journal of Molecular Sciences*, *20*(21), 5371. https://doi.org/10.3390/ijms20215371

Marnef, A., & Legube, G. (2021). R-loops as Janus-faced modulators of DNA repair. *Nature Cell Biology 2021 23:4*, *23*(4), 305–313. https://doi.org/10.1038/s41556-021-00663-4

Maunakea, A. K., Chepelev, I., Cui, K., & Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Research*, *23*(11), 1256–1269. https://doi.org/10.1038/cr.2013.110

Mayer, C., Neubert, M., & Grummt, I. (2008). The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. *EMBO Reports*, *9*(8), 774–780. https://doi.org/10.1038/embor.2008.109

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., & Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, *41*(Web Server issue). https://doi.org/10.1093/nar/gkt376

Meehan, R., Lewis, J. D., & Bird, A. P. (1992). Characterization of MECP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Research*, *20*(19), 5085–5092. https://doi.org/10.1093/nar/20.19.5085

Meng, G., Lv, Y. F., Dai, H., Zhang, X., & Guo, Q. N. (2014). Epigenetic silencing of methyl-CpG-binding protein 2 gene affects proliferation, invasion, migration, and apoptosis of human osteosarcoma cells. *Tumor Biology*, *35*(12), 11819–11827. https://doi.org/10.1007/s13277-014-2336-8

Mischo, H. E., Gómez-González, B., Grzechnik, P., Rondón, A. G., Wei, W., Steinmetz, L., Aguilera, A., & Proudfoot, N. J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. *Molecular Cell*, *41*(1), 21–32. https://doi.org/10.1016/j.molcel.2010.12.007

Morales, J. C., Richard, P., Patidar, P. L., Motea, E. A., Dang, T. T., Manley, J. L., & Boothman, D. A. (2016). XRN2 Links Transcription Termination to DNA Damage and Replication Stress. *PLoS Genetics*, *12*(7), e1006107. https://doi.org/10.1371/journal.pgen.1006107

Muotri, A. R., Marchetto, M. C. N., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., & Gage, F. H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature*, *468*(7322), 443–446. https://doi.org/10.1038/nature09544

Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N. A., Ó Broin, P., Sato, H., Zhang, Z., Jeddeloh, J., Montagna, C., Golden, A., Seoighe, C., & Greally, J. M. (2015). RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics and Chromatin*, *8*(1), 46. https://doi.org/10.1186/s13072-015-0040-6

Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M. J., Shu, X., Kriaucionis, S., & Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(8), 2709–2714. https://doi.org/10.1073/pnas.0608056104

Nguyen, D. T., Voon, H. P. J., Xella, B., Scott, C., Clynes, D., Babbs, C., Ayyub, H., Kerry, J., Sharpe, J. A., Sloane-Stanley, J. A., Butler, S., Fisher, C. A., Gray, N. E., Jenuwein, T., Higgs, D. R., & Gibbons, R. J. (2017). The chromatin remodelling factor ATRX suppresses R-loops in transcribed telomeric repeats . *EMBO Reports*, *18*(6), 914–928. https://doi.org/10.15252/embr.201643078

Nguyen, V. C., Clelland, B. W., Hockman, D. J., Kujat-Choy, S. L., Mewhort, H. E., & Schultz, M. C. (2010). Replication stress checkpoint signaling controls tRNA gene transcription. *Nature Structural and Molecular Biology*, *17*(8), 976–981. https://doi.org/10.1038/nsmb.1857

Niedernhofer, L. J., Gurkar, A. U., Wang, Y., Vijg, J., Hoeijmakers, J. H. J., & Robbins, P. D. (2018). Nuclear Genomic Instability and Aging. In *Annual Review of Biochemistry* (Vol. 87, pp. 295–322). Annual Reviews Inc. https://doi.org/10.1146/annurev-biochem-062917-012239

Nikitina, T., Shi, X., Ghosh, R. P., Horowitz-Scherer, R. A., Hansen, J. C., & Woodcock, C. L. (2007). Multiple Modes of Interaction between the Methylated DNA Binding Protein MeCP2 and Chromatin. *Molecular and Cellular Biology*, *27*(3), 864–877. https://doi.org/10.1128/mcb.01593-06

O'Day, C. L., Chavanikamannil, F., & Abelson, J. (1996). 18S rRNA processing requires the RNA helicase-like protein Rrp3. *Nucleic Acids Research*, *24*(16), 3201–3207. https://doi.org/10.1093/nar/24.16.3201

Oestergaard, V. H., & Lisby, M. (2017). Transcription-replication conflicts at chromosomal fragile sites—consequences in M phase and beyond. In *Chromosoma* (Vol. 126, Issue 2, pp. 213–222). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s00412-016-0617-2

Okamoto, Y., Abe, M., Itaya, A., Tomida, J., Ishiai, M., Takaori-Kondo, A., Taoka, M., Isobe, T., & Takata, M. (2019). FANCD2 protects genome stability by recruiting RNA processing enzymes to resolve R-loops during mild replication stress. *FEBS Journal*, *286*(1), 139–150. https://doi.org/10.1111/febs.14700

Osenberg, S., Karten, A., Sun, J., Li, J., Charkowick, S., Felice, C. A., Kritzer, M., Vu Chuong Nguyen, M., Yu, P., & Ballas, N. (2018). Activity-dependent aberrations in gene expression and alternative splicing in a mouse model of Rett syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(23), E5363–E5372. https://doi.org/10.1073/pnas.1722546115

Panov, K. I., Friedrich, J. K., Russell, J., & Zomerdijk, J. C. B. M. (2006). UBF activates RNA polymerase I transcription by stimulating promoter escape. *EMBO Journal*, *25*(14), 3310–3322. https://doi.org/10.1038/sj.emboj.7601221

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., Meyer, T., & Cimprich, K. A. (2009). A Genome-wide siRNA Screen Reveals Diverse Cellular Processes and Pathways that Mediate Genome Stability. *Molecular Cell*, *35*(2), 228–239. https://doi.org/10.1016/j.molcel.2009.06.021

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z. P., Economides, A. N., Bradner, J. E., Rabadan, R., & Basu, U. (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell*, *161*(4), 774–789. https://doi.org/10.1016/j.cell.2015.04.034

Pérez-Calero, C., Bayona-Feliu, A., Xue, X., Barroso, S. I., Muñoz, S., González-Basallote, V. M., Sung, P., & Aguilera, A. (2020). UAP56/DDX39B is a major cotranscriptional RNA–DNA helicase that

unwinds harmful R loops genome-wide. *Genes and Development*, *34*(13–14), 1–15. https://doi.org/10.1101/GAD.336024.119

Potapova, T. A., & Gerton, J. L. (2019). Ribosomal DNA and the nucleolus in the context of genome organization. In *Chromosome Research* (Vol. 27, Issues 1–2, pp. 109–127). Springer Netherlands. https://doi.org/10.1007/s10577-018-9600-5

Prado, F., & Aguilera, A. (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO Journal*, *24*(6), 1267–1276. https://doi.org/10.1038/sj.emboj.7600602

Prendergast, L., McClurg, U. L., Hristova, R., Berlinguer-Palmini, R., Greener, S., Veitch, K., Hernandez, I., Pasero, P., Rico, D., Higgins, J. M. G., Gospodinov, A., & Papamichos-Chronakis, M. (2020). Resolution of R-loops by INO80 promotes DNA replication and maintains cancer cell proliferation and viability. *Nature Communications*, *11*(1), 1–18. https://doi.org/10.1038/s41467-020-18306-x

Promonet, A., Padioleau, I., Liu, Y., Sanz, L., Biernacka, A., Schmitz, A. L., Skrzypczak, M., Sarrazin, A., Mettling, C., Rowicka, M., Ginalski, K., Chedin, F., Chen, C. L., Lin, Y. L., & Pasero, P. (2020). Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. *Nature Communications*, *11*(1), 1–12. https://doi.org/10.1038/s41467-020-17858-2

Proudfoot, N. J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. In *Science* (Vol. 352, Issue 6291). American Association for the Advancement of Science. https://doi.org/10.1126/science.aad9926

Proudfoot, N. J., Furger, A., & Dye, M. J. (2002). Integrating mRNA processing with transcription. In *Cell* (Vol. 108, Issue 4, pp. 501–512). Elsevier B.V. https://doi.org/10.1016/S0092-8674(02)00617-7

Ragione, F. Della, Vacca, M., Fioriniello, S., Pepe, G., & D'Esposito, M. (2016). MECP2, a multi-talented modulator of chromatin architecture. *Briefings in Functional Genomics*, *15*(6), 420–431. https://doi.org/10.1093/bfgp/elw023

Raška, I., Koberna, K., Malínský, J., Fidlerová, H., & Mašata, M. (2004). The nucleolus and transcription of ribosomal genes. In *Biology of the Cell* (Vol. 96, Issue 8, pp. 579–594). Wiley-Blackwell Publishing Ltd. https://doi.org/10.1016/j.biolcel.2004.04.015

Rondón, A. G., & Aguilera, A. (2019). What causes an RNA-DNA hybrid to compromise genome integrity? In *DNA Repair* (Vol. 81). Elsevier B.V. https://doi.org/10.1016/j.dnarep.2019.102660

Rondón, A. G., Jimeno, S., & Aguilera, A. (2010). The interface between transcription and mRNP export: From THO to THSC/TREX-2. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* (Vol. 1799, Issue 8, pp. 533–538). Biochim Biophys Acta. https://doi.org/10.1016/j.bbagrm.2010.06.002

Rube, H. T., Lee, W., Hejna, M., Chen, H., Yasui, D. H., Hess, J. F., Lasalle, J.

M., Song, J. S., & Gong, Q. (2016). Sequence features accurately predict genome-wide MeCP2 binding in vivo. *Nature Communications*, *7*(1), 1–12. https://doi.org/10.1038/ncomms11025

Ruiz, J. F., Gómez-González, B., & Aguilera, A. (2011). AID induces double-strand breaks at immunoglobulin switch regions and c-MYC causing chromosomal translocations in yeast THO mutants. *PLoS Genetics*, *7*(2), 1002009. https://doi.org/10.1371/journal.pgen.1002009

Salas-Armenteros, I., Barroso, S. I., Rondón, A. G., Pérez, M., Andújar, E., Luna, R., & Aguilera, A. (2019). Depletion of the MFAP1/SPP381 Splicing Factor Causes R-Loop-Independent Genome Instability. *Cell Reports*, *28*(6), 1551-1563.e7. https://doi.org/10.1016/j.celrep.2019.07.010

Salas-Armenteros, I., Pérez-Calero, C., Bayona-Feliu, A., Tumini, E., Luna, R., & Aguilera, A. (2017). Human THO –Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability . *The EMBO Journal*, *36*(23), 3532–3547. https://doi.org/10.15252/embj.201797208

Salim, D., & Gerton, J. L. (2019). Ribosomal DNA instability and genome adaptability. *Chromosome Research*, *27*(1–2), 73–87. https://doi.org/10.1007/s10577-018-9599-7

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Molecular cloning: a laboratory manual. *Molecular Cloning: A Laboratory Manual.*, *Ed. 2*.

Santoro, R., Li, J., & Grummt, I. (2002). The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nature Genetics*, *32*(3), 393–396. https://doi.org/10.1038/ng1010

Santos-Pereira, J. M., & Aguilera, A. (2015). R loops: New modulators of genome dynamics and function. In *Nature Reviews Genetics* (Vol. 16, Issue 10, pp. 583–597). Nature Publishing Group. https://doi.org/10.1038/nrg3961

Sanz, L. A., & Chédin, F. (2019). High-resolution, strand-specific R-loop mapping via S9.6-based DNA–RNA immunoprecipitation and high-throughput sequencing. *Nature Protocols*, *14*(6), 1734–1755. https://doi.org/10.1038/s41596-019-0159-1

Sanz, L. A., Hartono, S. R., Lim, Y. W., Steyaert, S., Rajpurkar, A., Ginno, P. A., Xu, X., & Chédin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Molecular Cell*, *63*(1), 167–178. https://doi.org/10.1016/j.molcel.2016.05.032

Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G. P., Heron, M., Williams, H., Söding, J., Stewart, A., & Svejstrup, J. Q. (2014). RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell*, *157*(5), 1037–1049. https://doi.org/10.1016/j.cell.2014.03.048

Sarkar, M., & Ghosh, M. K. (2016). DEAD box RNA helicases: Crucial regulators of gene expression and oncogenesis. In *Frontiers in Bioscience* -

Landmark (Vol. 21, Issue 2, pp. 225–250). Frontiers in Bioscience. https://doi.org/10.2741/4386

Schütz, P., Karlberg, T., van den Berg, S., Collins, R., Lehtiö, L., Högbom, M., Holmberg-Schiavone, L., Tempel, W., Park, H. W., Hammarström, M., Moche, M., Thorsell, A. G., & Schüler, H. (2010). Comparative structural analysis of human DEAD-Box RNA helicases. *PLoS ONE*, *5*(9), 1–11. https://doi.org/10.1371/journal.pone.0012791

Schwab, R. A., Nieminuszczy, J., Shah, F., Langton, J., Lopez Martinez, D., Liang, C. C., Cohn, M. A., Gibbons, R. J., Deans, A. J., & Niedzwiedz, W. (2015). The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Molecular Cell*, *60*(3), 351–361. https://doi.org/10.1016/j.molcel.2015.09.012

Sekiguchi, T., Hayano, T., Yanagida, M., Takahashi, N., & Nishimoto, T. (2006). NOP132 is required for proper nucleolus localization of DEAD-box RNA helicase DDX47. *Nucleic Acids Research*, *34*(16), 4593–4608. https://doi.org/10.1093/nar/gkl603

Selth, L. A., Sigurdsson, S., & Svejstrup, J. Q. (2010). Transcript elongation by RNA polymerase II. *Annual Review of Biochemistry*, *79*, 271–293. https://doi.org/10.1146/annurev.biochem.78.062807.091425

Sessa, G., Gómez-González, B., Silva, S., Pérez-Calero, C., Beaurepere, R., Barroso, S., Martineau, S., Martin, C., Ehlén, Å., Martínez, J. S., Lombard, B., Loew, D., Vagner, S., Aguilera, A., & Carreira, A. (2021). BRCA2 promotes DNA-RNA hybrid resolution by DDX5 helicase at DNA breaks to facilitate their repair‡. *The EMBO Journal*, *40*(7). https://doi.org/10.15252/embj.2020106018

Shen, J., Zhang, L., & Zhao, R. (2007). Biochemical characterization of the ATPase and helicase activity of UAP56, an essential pre-mRNA splicing and mRNA export factor. *Journal of Biological Chemistry*, *282*(31), 22544–22550. https://doi.org/10.1074/jbc.M702304200

Shivji, M. K. K., Renaudin, X., Williams, H., & Venkitaraman, A. R. (2018). BRCA2 Regulates Transcription Elongation by RNA Polymerase II to Prevent R-Loop Accumulation Correspondence Article BRCA2 Regulates Transcription Elongation by RNA Polymerase II to Prevent R-Loop Accumulation. *CellReports*, *22*, 1031–1039. https://doi.org/10.1016/j.celrep.2017.12.086

Silva, S., Camino, L. P., & Aguilera, A. (2018). Human mitochondrial degradosome prevents harmful mitochondrial R loops and mitochondrial genome instability. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(43), 11024–11029. https://doi.org/10.1073/pnas.1807258115

Singh, D. K., Pandita, R. K., Singh, M., Chakraborty, S., Hambarde, S., Ramnarain, D., Charaka, V., Ahmed, K. M., Hunt, C. R., & Pandita, T. K. (2018). MOF Suppresses Replication Stress and Contributes to Resolution of Stalled Replication Forks. *Molecular and Cellular Biology*, *38*(6). https://doi.org/10.1128/mcb.00484-17

Singleton, M. R., Dillingham, M. S., & Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. In *Annual Review of Biochemistry* (Vol. 76, pp. 23–50). Annual Reviews. https://doi.org/10.1146/annurev.biochem.76.052305.115300

Skene, P. J., Illingworth, R. S., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Andrews, R., & Bird, A. P. (2010). Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State. *Molecular Cell*, *37*(4), 457–468. https://doi.org/10.1016/j.molcel.2010.01.030

Skourti-Stathaki, K., & Proudfoot, N. J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. In *Genes and Development* (Vol. 28, Issue 13, pp. 1384–1396). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/gad.242990.114

Skourti-Stathaki, K., Proudfoot, N. J., & Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Molecular Cell*, *42*(6), 794–805. https://doi.org/10.1016/j.molcel.2011.04.026

Sollier, J., Stork, C. T., García-Rubio, M. L., Paulsen, R. D., Aguilera, A., & Cimprich, K. A. (2014). Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Molecular Cell*, *56*(6), 777–785. https://doi.org/10.1016/j.molcel.2014.10.020

Song, C., Hotz-Wagenblatt, A., Voit, R., & Grummt, I. (2017). SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes and Development*, *31*(13), 1370–1381. https://doi.org/10.1101/gad.300624.117

Sreerama Chaitanya Sridhara, A., Carvalho, S., Rita Grosso, A., Marcela Gallego-Paez, L., & Carmo-Fonseca, M. (2017). Transcription Dynamics Prevent RNA-Mediated Genomic Instability through SRPK2-Dependent DDX23 Phosphorylation. *Cell Reports*, *18*, 334–343. https://doi.org/10.1016/j.celrep.2016.12.050

Stirling, P. C., Chan, Y. A., Minaker, S. W., Aristizabal, M. J., Barrett, I., Sipahimalani, P., Kobor, M. S., & Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes and Development*, *26*(2), 163–175. https://doi.org/10.1101/gad.179721.111

Stork, C. T., Bocek, M., Crossley, M. P., Sollier, J., Sanz, L. A., Chédin, F., Swigut, T., & Cimprich, K. A. (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *ELife*, *5*(AUGUST). https://doi.org/10.7554/eLife.17548

Stuckey, R., García-Rodríguez, N., Aguilera, A., & Wellinger, R. E. (2015). Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(18), 5779–5784. https://doi.org/10.1073/pnas.1501769112

Sulli, G., Di Micco, R., & Di Fagagna, F. D. A. (2012). Crosstalk between

chromatin state and DNA damage response in cellular senescence and cancer. In *Nature Reviews Cancer* (Vol. 12, Issue 10, pp. 709–720). https://doi.org/10.1038/nrc3344

Takeuchi, Y., Horiuchi, T., & Kobayashi, T. (2003). Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes and Development*, *17*(12), 1497–1506. https://doi.org/10.1101/gad.1085403

Tang, X., Li, G., Su, F., Cai, Y., Shi, L., Meng, Y., Liu, Z., Sun, J., Wang, M., Qian, M., Wang, Z., Xu, X., Cheng, Y. X., Zhu, W. G., & Liu, B. (2020). HDAC8 cooperates with SMAD3/4 complex to suppress SIRT7 and promote cell survival and migration. *Nucleic Acids Research*, *48*(6), 2912–2923. https://doi.org/10.1093/nar/gkaa039

Tehranchi, A. K., Blankschien, M. D., Zhang, Y., Halliday, J. A., Srivatsan, A., Peng, J., Herman, C., & Wang, J. D. (2010). The Transcription Factor DksA Prevents Conflicts between DNA Replication and Transcription Machinery. *Cell*, *141*(4), 595–605. https://doi.org/10.1016/j.cell.2010.03.036

Tous, C., & Aguilera, A. (2007). Impairment of transcription elongation by R-loops in vitro. *Biochemical and Biophysical Research Communications*, *360*(2), 428–432. https://doi.org/10.1016/j.bbrc.2007.06.098

Trautinger, B. W., Jaktaji, R. P., Rusakova, E., & Lloyd, R. G. (2005). RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Molecular Cell*, *19*(2), 247–258. https://doi.org/10.1016/j.molcel.2005.06.004

Tschochner, H., & Hurt, E. (2003). Pre-ribosomes on the road from the nucleolus to the cytoplasm. In *Trends in Cell Biology* (Vol. 13, Issue 5, pp. 255–263). Elsevier Ltd. https://doi.org/10.1016/S0962-8924(03)00054-0

Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., Pommier, Y., Tazi, J., Coquelle, A., & Pasero, P. (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature Cell Biology*, *11*(11), 1315–1324. https://doi.org/10.1038/ncb1984

Valentini, M., & Linder, P. (2021). Happy Birthday: 30 Years of RNA Helicases. In *Methods in Molecular Biology* (Vol. 2209, pp. 17–34). Humana Press Inc. https://doi.org/10.1007/978-1-0716-0935-4_2

van Attikum, H., & Gasser, S. M. (2009). Crosstalk between histone modifications during the DNA damage response. In *Trends in Cell Biology* (Vol. 19, Issue 5, pp. 207–217). Trends Cell Biol. https://doi.org/10.1016/j.tcb.2009.03.001

Vydzhak, O., Luke, B., & Schindler, N. (2020). Non-coding RNAs at the Eukaryotic rDNA Locus: RNA–DNA Hybrids and Beyond. In *Journal of Molecular Biology* (Vol. 432, Issue 15, pp. 4287–4304). Academic Press. https://doi.org/10.1016/j.jmb.2020.05.011

Wahba, L., Amon, J. D., Koshland, D., & Vuica-Ross, M. (2011). RNase H and Multiple RNA Biogenesis Factors Cooperate to Prevent RNA:DNA Hybrids from Generating Genome Instability. *Molecular Cell*, *44*(6), 978–988.

https://doi.org/10.1016/j.molcel.2011.10.017

Wahba, L., Costantino, L., Tan, F. J., Zimmer, A., & Koshland, D. (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes and Development*, *30*(11), 1327–1338. https://doi.org/10.1101/gad.280834.116

Wang, I. X., Grunseich, C., Fox, J., Burdick, J., Zhu, Z., Ravazian, N., Hafner, M., & Cheung, V. G. (2018). Human proteins that interact with RNA/DNA hybrids. *Genome Research*, *28*(9), 1405–1414. https://doi.org/10.1101/gr.237362.118

Wellinger, R. E., Prado, F., & Aguilera, A. (2006). Replication Fork Progression Is Impaired by Transcription in Hyperrecombinant Yeast Cells Lacking a Functional THO Complex. *Molecular and Cellular Biology*, *26*(8), 3327–3334. https://doi.org/10.1128/mcb.26.8.3327-3334.2006

Wong, J. J. L., Gao, D., Nguyen, T. V., Kwok, C. T., Van Geldermalsen, M., Middleton, R., Pinello, N., Thoeng, A., Nagarajah, R., Holst, J., Ritchie, W., & Rasko, J. E. J. (2017). Intron retention is regulated by altered MeCP2-mediated splicing factor recruitment. *Nature Communications*, *8*(1), 1–13. https://doi.org/10.1038/ncomms15134

Xu, B., & Clayton, D. A. (1996). RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: An implication for RNA-DNA hybrids serving as primers. *EMBO Journal*, *15*(12), 3135–3143. https://doi.org/10.1002/j.1460-2075.1996.tb00676.x

Yan, Q., Shields, E. J., Bonasio, R., Sarma Correspondence, K., & Sarma, K. (2019). Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach Cell Reports Resource Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach. *Cell Reports*, *29*. https://doi.org/10.1016/j.celrep.2019.09.052

Yang, Q., Del Campo, M., Lambowitz, A. M., & Jankowsky, E. (2007). DEAD-Box Proteins Unwind Duplexes by Local Strand Separation. *Molecular Cell*, *28*(2), 253–263. https://doi.org/10.1016/j.molcel.2007.08.016

Yasuhara, T., Kato, R., Hagiwara, Y., Shiotani, B., Yamauchi, M., Nakada, S., Shibata, A., & Miyagawa, K. (2018). Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. *Cell*, *175*(2), 558-570.e11. https://doi.org/10.1016/j.cell.2018.08.056

Yasui, D. H., Peddada, S., Bieda, M. C., Vallero, R. O., Hogart, A., Nagarajan, R. P., Thatcher, K. N., Farnham, P. J., & LaSalle, J. M. (2007). Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19416–19421. https://doi.org/10.1073/pnas.0707442104

You, K. T., Park, J., & Kim, V. N. (2015). Role of the small subunit processome in the maintenance of pluripotent stem cells. *Genes and Development*, *29*(19), 2004–2009. https://doi.org/10.1101/gad.267112.115

Young, J. I., Hong, E. P., Castle, J. C., Crespo-Barreto, J., Bowman, A. B., Rose, M. F., Kang, D., Richman, R., Johnson, J. M., Berget, S., & Zoghbi, H. Y. (2005). Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(49), 17551–17558. https://doi.org/10.1073/pnas.0507856102

Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T. E., & Lieber, M. R. (2003). *R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells*. https://doi.org/10.1038/ni919

Yu, K., Roy, D., Bayramyan, M., Haworth, I. S., & Lieber, M. R. (2005). Fine-Structure Analysis of Activation-Induced Deaminase Accessibility to Class Switch Region R-Loops. *Molecular and Cellular Biology*, *25*(5), 1730–1736. https://doi.org/10.1128/mcb.25.5.1730-1736.2005

Zeng, C., Onoguchi, M., & Hamada, M. (2021). Association analysis of repetitive elements and R-loop formation across species. *Mobile DNA 2021 12:1*, *12*(1), 1–11. https://doi.org/10.1186/S13100-021-00231-5

Zhang, C., You, Y., Wang, J., & Weber, X. (2015). The DHX33 RNA helicase promotes mRNA translation initiation. *Molecular and Cellular Biology*, *35*. https://doi.org/10.1128/MCB.00315-15

Zhang, X., Chiang, H. C., Wang, Y., Zhang, C., Smith, S., Zhao, X., Nair, S. J., Michalek, J., Jatoi, I., Lautner, M., Oliver, B., Wang, H., Petit, A., Soler, T., Brunet, J., Mateo, F., Angel Pujana, M., Poggi, E., Chaldekas, K., … Li, R. (2017). Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. *Nature Communications*, *8*(1), 1–12. https://doi.org/10.1038/ncomms15908