Papel de la helicasa UAP56/DDX39B y otros factores asociados al metabolismo del ARN en la integridad del genoma

Carmen Pérez Calero

Tesis doctoral

Universidad de Sevilla

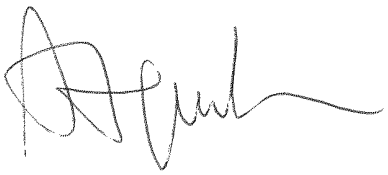
2019

Papel de la helicasa UAP56/DDX39B y otros factores asociados al metabolismo del ARN en la integridad del genoma

Trabajo realizado en el Departamento de Genética, Facultad de Biología, y en el departamento de Biología del Genoma, CABIMER, de la Universidad de Sevilla, para optar al grado de doctora en la Universidad de Sevilla por la licenciada Carmen Pérez Calero.

Sevilla, 2019

La doctoranda, **El director de tesis,**



Carmen Pérez Calero **Andrés Aguilera López**

TABLE OF CONTENTS

INDEX OF FIGURES

INTRODUCTION

MATERIALS AND METHODS

INDEX OF TABLES

ABBREVIATIONS

A	Alanine
AID	Activation-Induced cytidine Deaminase
AREX	Alternative mRNA Export complex
AS	Alternative splicing
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine triphosphate
ATRIP	ATR Interacting Protein
A.U.	Arbitrary Units
bp	base pairs
BRCA	Breast Cancer
BrdU	Bromodeoxyuridine
CBC	Cap-Binding Complex
CBP	Cap-Binding Protein
CDC25	Cell Division Cycle 25
cDNA	complementary DNA
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation-sequencing
CIN	Chromosome Instability
CldU	Cloro-deoxyuridine
CNVs	Copy Number Variants
Co-IP	Co-immunoprecipitation
CTD	Carboxyl-Terminal Domain
DDR	DNA Damage Response
DDX	DEAD box
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
DRIP	RNA-DNA immunoprecipitation
DRIP-seq	RNA-DNA immunoprecipitation-sequencing
DRIPc-seq	RNA-DNA immunoprecipitation followed by cDNA conversion-sequencing

dsRNA	Double-stranded RNA
E	Glutamic acid
EGFP	Enhanced Green Fluorescent Protein
EJC	Exon Junction Complex
EMT	Epithelial-mesenchymal transition
GCR	Gross chromosomal rearrangements
FA	Fanconi Anemia
FACS	Fluorescence-Activated Cell Sorting
FACT	Facilitates Chromatin Transcription
G+C	Guanine and Cytosine
GFP	Green Fluorescence Protein
H3S10-P	Histone H3 Phosphorylation at serine 10
HBD	Hybrid-binding domain
HDAC	Histone Deacetylase
hnRNP	heterogeneous nuclear Ribonucleoprotein
HR	Homologous recombination
HU	Hydroxyurea
IdU	Iodo-deoxyuridine
IF	Immunofluorescence
Ig	Immunoglobulin
K	Lysine
Kb	Kilobases
KDa	Kilodalton
lncRNA	long non-coding RNA
LOH	Loss Of Heterozygosity
MCM	Mini chromosome maintenance
mESC	mouse embryonic stem cells
MRN	Mre11-Rad50-Nbs1
mRNA	messenger RNA
mRNP	messenger Ribonucleoprotein particle
NES	Nuclear export signal
NHEJ	Non-homologous end joining

NPC	Nuclear Pore Complex
nt	Nucleotides
NTS	Non-transcribed strand
NXF1	RNA export factor 1
OHT	Hydroxytamoxifen
ORF	Open reading frame
PCR	Polymerase Chain Reaction
PIKK	Phosphatidylinositol 3 like kinase
Pol	Polymerase
PRO-seq	Precision nuclear run-on-sequencing
qPCR	quantitative Polymerase Chain Reaction
RBP	RNA-binding protein
rDNA	Ribosomal DNA
RF	Replication fork
RFB	Replication fork barrier
RNAi	RNA interference
RNAP	RNA polymerase
RNase	Ribonuclease
RPA	Replication protein A
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SEM	Standard Error of the Mean
siRNA	small interfering RNA
snRNP	small nucleolar ribonucleoprotein
SR	Serine/arginine-rich protein
SSB	Single-Strand Break
ssDNA	single-stranded DNA
TAM	Transcription-associated mutation
TAR	Transcription-associated recombination
TC-NER	Transcription couple nucleotide excision repair
TCR	Transcription couple Repair
Top	Topoisomerase
TREX	Transcription-Export complex

TS	Transcribed Strand
TSA	Trichostatin A
UV	Ultraviolet
U2AF	U2 Small Nuclear RNA Auxiliary Factor
WT	Wild Type

RESUMEN

Para la transmisión fidedigna de la información genética es indispensable preservar la integridad del genoma. Las células han desarrollado procesos complejos altamente regulados que velan por la estabilidad del mismo, evitando o resolviendo problemas que puedan comprometerla. La causa de dicha inestabilidad genética no sólo radica en la acción de agentes genotóxicos externos, sino también en aquellos derivados del propio metabolismo celular, resultado de fallos en los propios procesos endógenos de la célula como la transcripción, replicación y recombinación. La manifestación de la inestabilidad genética se presenta generalmente en forma de mutaciones y reordenaciones cromosómicas, características asociadas a la predisposición a envejecimiento y procesos tumorales.

Paradójicamente, uno de los procesos más esenciales para la supervivencia de la célula, la transcripción, puede constituir a su vez una de las fuentes más importantes de inestabilidad genética. Esto se debe principalmente a la aparición de ADN de cadena sencilla durante su desarrollo, que es más susceptible a daños que la doble cadena. Asimismo, la propia maquinaria de transcripción puede suponer un obstáculo para otro proceso esencial en la célula como es la replicación, pudiendo derivar todo ello en un aumento de roturas y recombinación en el ADN. Durante la transcripción, el ARN naciente puede hibridar con la hebra molde de ADN, de manera que la hebra no transcrita se mantiene en forma de cadena sencilla. Estas estructuras generadas se conocen como bucles R (*R loops*) y están compuestos por un híbrido de ADN-ARN y la cadena sencilla de ADN desplazada. Si bien se ha demostrado que los *R loops* pueden desempeñar importantes funciones fisiológicas, cuando se forman o acumulan de manera incontrolada pueden suponer una amenaza para la integridad del genoma.

El ARNm naciente ha de ser correctamente empaquetado en forma de ribonucleoproteínas mensajeras (mRNPs) para la elongación de la transcripción,

la integridad y el procesamiento del ARNm, previo a su transporte al citoplasma. Dicho ARNm, como sustrato potencial en la formación de *R loops*, ha de ser pertinentemente procesado, empaquetado y protegido para evitar que hibride con el ADN molde. Es por ello que, durante su procesamiento, existen numerosos factores que se unen al ARNm y contribuyen a proteger el genoma de la formación de *R loops*. En esta tesis nos hemos centrado en factores que tienen un papel en la biogénesis de las mRNPs y en particular en la helicasa UAP56/DDX39B perteneciente a la familia de helicasas DEAD/H box. Esta proteína, además de desempeñar un papel en el acoplamiento del madurosoma (*spliceosome*), interviene en la interfaz de la transcripción con la biogénesis y transporte de mRNPs. En concreto, UAP56 interacciona con el complejo THO, involucrado en la formación de la mRNP, y promedia el transporte de la mayoría de los mRNAs a través del reclutamiento del factor ALY/REF. El silenciamiento de UAP56 genera una alta inestabilidad genética en las células. Esto se atribuye a defectos en la formación de la mRNP, pero se desconoce si este fenotipo está mediado por *R loops.*

En la presente tesis hemos querido profundizar en los mecanismos por los cuales UAP56 contribuye al mantenimiento de la integridad del genoma. Hemos corroborado que el silenciamiento de UAP56 causa inestabilidad genética, determinada por un incremento de roturas en el ADN y descubierto nuevos fenotipos asociados como defectos en la replicación. Además, hemos desvelado que tanto la inestabilidad como son los defectos en replicación están mediados por *R loops*. En colaboración con el laboratorio del Dr. Patrick Sung en la Universidad de Yale, hemos descubierto que UAP56 es una helicasa de ADN-ARN capaz de resolver *R loops in vitro*. *In vivo*, la sobreexpresión de UAP56 suprime los fenotipos de inestabilidad mediados por *R loops* característicos de diversos mutantes, confirmando su actividad helicasa. Por último, hemos determinado en todo el genoma las regiones más propicias a acumular *R loops* y a donde se une preferentemente UAP56, así como el transcriptoma resultante del silenciamiento de UAP56.

Debido al creciente número de estudios que relacionan el silenciamiento de diferentes helicasas de la familia DEAD/H box con la acumulación de *R loops*, hemos comparado a nivel de todo el genoma los sitios de acumulación de *R loops* tras el silenciamiento de UAP56 y DDX5 para investigar si ambas proteínas presentan una actividad redundante.

Finalmente, hemos estudiado la contribución del factor de transcripción Snail1 a la integridad del genoma. Tradicionalmente, el estudio de esta proteína se ha enmarcado en el contexto de la transición epitelio-mesenquima (EMT), gracias a la cual las células epiteliales adquieren las características propias de las células mesenquimales como la pérdida de los contactos intracelulares e interacción con la membrana basal. Aunque, este proceso es esencial para el desarrollo, se ha demostrado que se activa de manera anormal en los procesos cancerosos. Su activación permite a las células adquirir una mayor capacidad de invasión y metástasis. Un factor crucial encargado de desencadenar el inicio de esta transición es Snail1. En este trabajo hemos demostrado que el silenciamiento de Snail1 causa inestabilidad genética y defectos en la replicación mediados por *R loops*. Estos datos son particularmente relevantes al sugerir que pueda existir una posible conexión entre los *R loops* y la inestabilidad asociada a cáncer, dado el papel de Snail en determinados procesos cancerígenos.

INTRODUCTION

1. GENOME INSTABILITY

Cells are able to maintain order in a disordered world due to a large extent to the genetic information encompassed in the genome and stored in form of DNA molecule. Since the prime objective for cells is to faithfully transmit this genetic material in each cell division to its offspring, cells have developed a wide variety of highly regulated coordinated processes to guarantee genome stability. However, DNA is constantly assaulted by endogenous and environmental agents that might propitiate genome changes. This paradoxical situation drives to genetic variability, which although can have detrimental consequences, is also the basis for evolution (Aguilera and Gomez-Gonzalez, 2008). Thereby, genome instability can be a source of genetic diversity as is the case of immunoglobulin diversification. Nonetheless, under certain circumstances such as the ones that affect a proper DNA repair and/or replication, genome alterations can compromise genome integrity leading to an increased genome instability, which is a hallmark of aging, hereditary genetic diseases and cancer-related disorders (Gaillard and Aguilera, 2016; Negrini et al., 2010).

Such changes in the genome can be the consequence of lesions in the DNA generated by different sources of DNA damage, including endogenous metabolites from cellular processes like transcription or replication, and exogenous sources, such as external genotoxic agents. Indeed, DNA can undergo distinct type of lesions that cover abasic sites, bases mismatch, DNA adducts, inter- and intra-strand crosslinks, single-stranded DNA (ssDNA) gaps and double-strand breaks (DSBs). Therefore, genome instability comprises different form of mutations from point mutations to microsatellite contractions and expansions and chromosomal changes, such as chromosome instability (CIN), gross chromosomal rearrangements (GCRs), copy number variants (CNVs), loss of heterozygosity (LOH) and hyper-recombination (Aguilera and Garcia-Muse, 2013).

To counteract this threat to genome integrity, cells have evolved several systems based on the sensing and repair of DNA damages and coupling to the

cell cycle to warrant proper cell proliferation or, alternatively apoptosis, which are collectively termed DNA Damage Response (DDR) (Gaillard and Aguilera, 2016).

1.1. DNA damage response

The DNA damage response (DDR) is based on a complex signally network that detects DNA damage and replication stress and orchestrates DNA repair and cell cycle progression in order to counteract DNA damage. This system comprises a wide variety of DNA repair mechanisms, given the diversity of DNA-lesion types. The DDR signaling cascades are triggered by the recognition of specific DNA lesions by sensor proteins such as the MRN (Mre11 Rad50 Nb1) sensor complex that detects DSBs and RPA (replication protein A). These proteins signal the accumulation of ssDNA generated during replication stress, which are sensed by the ataxia telangiectasia mutated (Tel1/ATM) and ataxia telangiectasia and Rad3 related (Mec1/ATR) via its partner protein ATRIP (ATR-interacting partner) (as named in yeast/humans), respectively. Afterwards, ATM/ATR phosphorylates various proteins that trigger cell cycle progression arrest in coordination with DNA repair pathways to preserve genome integrity (Sulli et al., 2012). Among them, there are important mediators such as the histone variant H2AX, p53-binding protein (53BP1) or BRCA1. The DDR signaling can be spread from the damage locus thanks to the phosphorylation of the histone variant H2AX on Ser139 (known as γH2AX) by ATM/ATR or the activation of the downstream kinases including the checkpoint kinases 1 and 2 (Chk1/CHK1 and Rad53/CHK2) phosphorylated by ATR and ATM, respectively. Finally, all these complex signaling pathways converge on downstream effectors such as p53 and the cell division cycle 25 (CDC25) phosphatases. At the end, cells have different alternatives: cell cycle can be arrested to permit DNA repair before proliferation and cell cycle resumption or, if the DNA damage is persistent or irreparable, cell apoptosis and senescence pathways can be activated (Sulli et al., 2012).

1.2. Replication

DNA replication is essential for genome duplication, and then, for transmission of genetic information to the offspring. Each time a cell divides, only once during S

phase, billions of nucleotides must be accurately copied in coordination with cell cycle. For this purpose, replisomes are required to maintain an accurate rate through the chromatinized DNA template and overcome different impediments such as DNA lesions, torsional stress, non-B DNA structures or the transcription machinery itself that can hamper replication fork (RF) progression (Gaillard et al., 2015; Gomez-Gonzalez and Aguilera, 2019). However, these obstacles can alter the proper function of the replisome leading to replication stress, which in turn could drive to persistent RF stalling and then to, the collapse of the replication machinery causing replisome disassembly, ssDNA gaps and DNA breaks threatening genetic integrity (Aguilera and Garcia-Muse, 2013; Aguilera and Gomez-Gonzalez, 2008; Kim and Jinks-Robertson, 2012). To counteract this situation, cells have evolved refined mitotic and S phase checkpoint pathways that ensure DNA integrity and chromosome transmission (Gaillard et al., 2015).

1.3. Transcription-associated genome instability

Gene expression encompasses different processes from transcription to the proper processing of the messenger RNA (mRNA), export and translation of that RNA into a protein in the cytoplasm (Gaillard et al., 2013). During transcription the two DNA strands separate locally within the RNA polymerase (RNAP) to form the transcription bubble in which the RNAP uses one of the DNA strands as template (transcribed strand) to generate a complementary RNA chain, forming an RNA-DNA hybrid. Meanwhile, the non-transcribed strand remains unpaired as ssDNA. Furthermore, transcription generates DNA topological changes and even chromatin remodeling changes required to permit the movement of the transcription machinery through the DNA (Selth et al., 2010). Numerous studies from the last decades have revealed that transcription represent also a source of DNA variability. Since any DNA lesions or secondary structure with the potential to stall a DNA polymerase, could also block an RNA polymerase, transcription has been found to be responsible for high levels of mutation (Transcription-associated mutation, TAM) and recombination (Transcription-associated recombination, TAR) (Gaillard and Aguilera, 2016).

Genome instability associated with transcription can be adduced to several factors. For instance, the transient exposure of a ssDNA after the RNAPII passage increase its vulnerability toward mutagenic agents (Figure I1A). This is consistent with previous reports showing a synergistic effect between the transcriptional stated of a DNA regions and its susceptibility to DNA damaging agents (Garcia-Rubio et al., 2003). Moreover, during this process the DNA is also subjected to topological changes that nucleic acids must accomplish to allow the movement of the transcription machinery, which leads to positive and negative supercoiling ahead and behind the RNAPII, respectively. Such changes can facilitate the transient formation of the ssDNA. This stretched ssDNA could be damaged (Bermejo et al., 2009), but it can also facilitate the appearance of non-B structures such as RNA-DNA hybrids (known as R loops when formed outside the transcription bubble), hairpins, triplex DNA or G-quadruplexes (Figure I1B), among others. All such structures that could lead to genome instability due to its potential to block the replisome (Aguilera and Garcia-Muse, 2013; Gaillard and Aguilera, 2016; Kim and Jinks-Robertson, 2012).

In other cases, the RNA itself can facilitate aberrant structures such as RNA-DNA hybrids or R loops. For instance, if the nascent mRNA is not properly packaged into a messenger ribonucleoprotein particle (mRNP), as in the case of mRNA biogenesis mutant, it can invade the DNA duplex (Huertas and Aguilera, 2003)(Figure I1B). However, the mechanisms of TAR are not completely understood. Since homologous recombination (HR) is the main pathway responsible for repair of DNA breaks occurring preferentially during replication, mounting evidence suggest that TAR is the consequence of transcription-replication collisions which can cause RF collapse (Aguilera, 2002; Gaillard and Aguilera, 2013; Garcia-Muse and Aguilera, 2016).

1.4. Transcription-replication conflicts

Since transcription and replication are crucial for cell survival and proliferation and they use the same DNA as a template, collisions between both machineries are unavoidable when both encounter each other at the same DNA region in

certain occasions in prokaryotes and eukaryotes (Gomez-Gonzalez and Aguilera, 2019). These two processes are significantly different: whereas the RNA polymerase embraces both DNA strands during transcription, the replicative machinery uses two DNA polymerases embracing each one each ssDNA. Furthermore, while several RNA polymerases can simultaneously transcribe the same gene, replisomes always move alone once per cell cycle. Additionally, in prokaryotes the fact that DNA replication and transcription exhibit different rates (Helmrich et al., 2013) make these conflicts to occur often (Merrikh et al., 2012). Therefore, how the replisome advance along the double-stranded DNA despite the presence of the transcription machinery becomes an abiding question. Moreover, as mentioned above, RNA polymerases are considered one of the main impediments to the progression of the RFs (Bermejo et al., 2012; Deshpande and Newlon, 1996; Liu and Alberts, 1995). And of utmost importance, these encounters are prone to occur at transcribed sites driven by different RNAPs (Gaillard et al., 2013; Gottipati et al., 2008; Prado and Aguilera, 2005) representing a common issue for both prokaryotes and eukaryotes (Helmrich et al., 2011; Prado and Aguilera, 2005).

In any case, these collisions that compromise genome integrity not always require a physical contact between both machineries. Given that both processes modify topology, chromatin organization and structure of the DNA, different mechanism have arose by which transcription compromise genome integrity in a replication-mediated manner (Gomez-Gonzalez and Aguilera, 2019). Even though transcription-replication conflicts can be co-directional or head-on, regardless the directionality of the conflict, replisomes are impeded to progress through an RNA polymerase. However, experimental evidence suggests that head-on collisions are more deleterious (Figure I1C) (Prado and Aguilera, 2005).

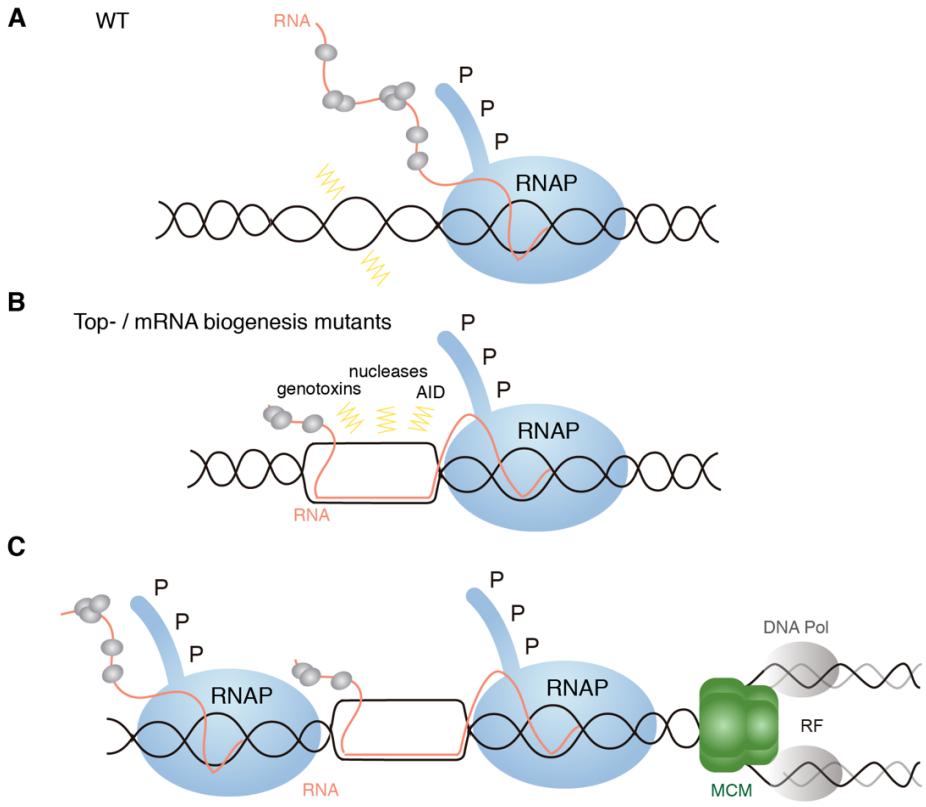


Figure I1. Transcription-associated genome instability.
Transcription could compromise genome integrity by itself. (A) Local positive supercoiling ahead the RNAPII and negative supercoiling behind it make the DNA more vulnerable to damage. (B) Non B-structures such as R loops can emerge when the nascent RNA hybridizes with the template DNA. Under circumstances with high level of R loops, as is the case of mRNP biogenesis mutants such as THO, the displaced ssDNA is more vulnerable to damage by genotoxins or nucleases. (C) Transcription-replication collisions as a result of replication progression caused by the RNA polymerase. If these encounters are head-on (if both machineries progress in the opposite direction), they are more deleterious than when they are co-directional (both machineries advance in the same direction). MCM, minichromosome maintenance complex; DNA pol, DNA polymerase. Figure adapted from (Gaillard et al., 2013).

Cells have evolved different strategies to minimize the harmful impact of such collisions. In bacteria, highly transcribed and essential genes are orientated co-directionally with replication to avoid head-on collisions (Brewer, 1988). In the human genome, this co-orientation bias is also present (Petryk et al., 2016). However, the considerable complexity of eukaryotic genomes demand further solutions. For instance, in budding yeast a RFB downstream of rDNA loci prevents head-on collisions with the RNA polymerase I by blocking RF (Linskens and Huberman, 1988). In general, transcription and replication in eukaryotic cells have some spatial and temporal separation but some highly expressed genes are transcribed during the S-phase (Wei et al., 1998). Interestingly, those

mechanisms and factors that contribute to eliminate the RNAPII stalling also help to suppress the consequences of these encounters. This is the case of the RECQL5 DNA helicase, which decreases the elongation rate of the RNAPII preventing its stalling (Saponaro et al., 2014).

2. RNA METABOLISM AND GENOME STABILITY

2.1. Transcription and RNA processing

In eukaryotic gene expression, the nascent pre-mRNA undergoes maturation by processing factors and protein-packaging lead to an export-competent mRNP particle apt to its translation in the cytoplasm. These mRNA processing steps comprise 5′-capping, splicing and 3′-end processing. During these events, numerous RBPs are loaded co-transcriptionally to the nascent RNA. The processing steps and mRNP export are interlinked and they influence one another's specificity and efficiency but them are also tightly linked to transcription. This fact allows the regulation of the process at different steps during transcription and makes the composition of the mRNP a dynamic process (Bentley, 2014; Proudfoot et al., 2002).

The first processing step that occurs when the RNA emerge from the RNAPII is the addition of a cap structure to the 5' end of all the mRNAs by capping enzymes that binds to the phosphorylated Carboxy-Terminal Domain (CTD) of the RNPII largest subunit. This initial structure is then recognized by the Cap Binding Complex (CBC) (Figure I2). This first step initiates in the assembly of the mRNP and is believed to play a major role in the stabilization of the mRNA and required for further processes that occur on the mRNA molecule (splicing, transcription termination export, nuclear mRNA decay, translation, non-sense-mediated decay and decapping) (Aguilera, 2005a; Gonatopoulos-Pournatzis and Cowling, 2014; Proudfoot et al., 2002)

Splicing is also an important step to which pre-mRNAs are subjected in order to obtain a mature mRNP. The pre-mRNA is spliced by different splicing factors that eliminate introns and join the remaining exons to obtain the mature

mRNA. Splicing factors associate rapidly with the nascent RNA and introns are removed co-transcriptionally, but some of them can be eliminated after transcription (Kornblihtt et al., 2004). Splicing occurs in eukaryotes from yeast to human. However, whereas introns in the budding yeast are scarce and the larger part are found in ribosomal protein genes, In humans almost all RNAPII transcribed genes contain introns (Izquierdo and Valcarcel, 2006; Shkreta and Chabot, 2015). The splicing process is catalyzed by a dynamic ribonucleoprotein structure called spliceosome, whose formation involves the stepwise assembly of five uridine (U)-rich small nuclear ribonucleoproteins (snRNPs) (U1, U2, U4, U5 and U6), along with many associated protein cofactors (Will and Luhrmann, 2011). The spliceosome recognizes short consensus sequences at the nuclear pre-mRNA introns: the 5' splice site, the branch site and the 3' splice site (Gersappe et al., 1999). Nuclear pre-mRNA introns are removed by two consecutive transesterification reactions that are necessary to excise introns and join together the remained exons. Sequential assembly of the spliceosome occurs by the ordered interaction of the spliceosomal snRNPs and numerous other splicing factors (Will and Luhrmann, 2011). In general, splice sites that are more adjusted to the consensus sequence ("strong" splicing site) lead to constitutive splicing and full usage of the site. However, those splice sites that diverge from the consensus sequence ("weak" splice sites) are less efficiently recognized and used leading to alternative splicing (Kornblihtt et al., 2013). Alternative splicing is predominant in higher eukaryotes and permits the production of multiple mRNA variants from a single pre-mRNA contributing to transcriptomic and proteomic diversity. The selection of the final splice site is regulated by diverse factors including members of the SR and hnRNP protein families which bind to enhancers and silencers sequences, respectively. Moreover, alternative splicing when occurred co-transcriptionally is also regulated by a more complex process that involves the transcription machinery (Kornblihtt et al., 2013). Indeed, RNAPII pausing at the 3' end on yeast genes have been shown to favor the co-transcriptional excision of introns, and *vice versa*, splicing could also promote pausing in transcription (Alexander et al., 2010).

Other relevant step in gene expression is the 3' end cleavage of transcripts generated by RNAPII. This is an universal step that consists on the cleavage of the nascent transcript and the acquisition of the poly(A) tail in the majority of genes. This poly(A) tail is essential for stability, translocation to the cytoplasm and translation of the transcripts. Thus, the 3' end processing reaction requires multiple protein factors that recognizes poly(A) signals on the nascent transcripts and produce the endonucleolytic 3' end cleavage and the addition of a polyadenylated tail. This step serves as a bridge in the network connecting different transcription and other steps in the mRNP biogenesis. Indeed, 3' end polyadenylation factors and sequence elements of the poly(A) signal modulate transcription termination and, in turn, transcription factors/activators affect processing at the poly(A) signal. Indeed, a connection between 3' end processing and mRNP export are also linked. Accordingly, mRNPs that are not processed at the 3' end will be degraded or not transported efficiently to the cytoplasm (Millevoi and Vagner, 2010).

Additionally, others RBPs binds co-transcriptionally to the nascent mRNA and help in the formation of the mRNP, preventing its degradation and promoting its export. The export competent mRNP are drive to the cytoplasm by the export receptor factor NXF1/TAP-p15 (yeast Mex67-Mtr2) that interacts with nucleoporins that form the nuclear pore complex (NPC). This complex operates together with different adaptor proteins including proteins from the THO/TREX and THSC/TREX2 complexes (Figure I2) (Kohler and Hurt, 2007)

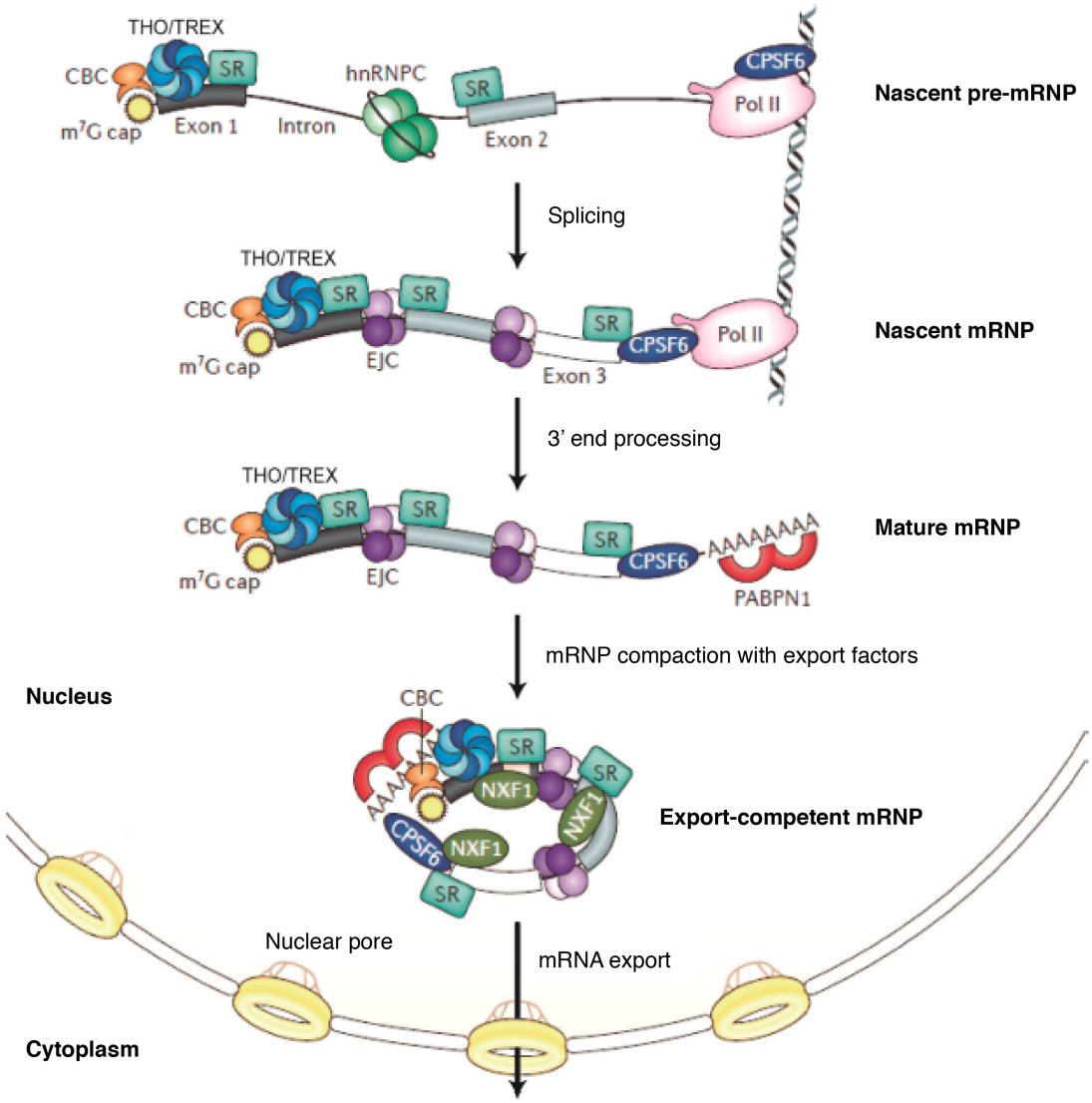


Figure I2. Co-transcriptional assembly of the mRNPs.
Illustration representing the co-transcriptional assembly of messenger ribonucleoprotein particles (mRNPs) in mammalian cells. During transcription, the cap-binding complex (CBC) binds to the 7-methylguanosine (m'G) cap and facilitates splicing. The splicing and the binding of numerous RNA-binding proteins occur co-transcriptionally. The recruitment of the THO/TREX complex is mediated by the interaction with proteins of the CBC. The exon junction complex (ECJ) and serine/arginine rich (SR) proteins contribute to mRNP biogenesis and export. After cleavage and polyadenylation, THO/TREX and other export adaptor factors recruit the RNA export receptor NXF1 to permit efficient nuclear export through the nuclear pore. nuclear export through the nuclear pore. hnRNPC, heterogeneous nuclear ribonucleoprotein C; PABPC1, cytoplasmic poly(A) binding protein 1; PABPN1, nuclear poly(A)-binding protein 1; Pol II, RNA polymerase II; pre-mRNP, precursor mRNP; CPSF6, cleavage factor and polyadenylation factor 6; NXF1, RNA export factor 1. Figure adapted from (Muller-McNicoll and Neugebauer, 2013).

2.2. THO/TREX as a paradigm of the connection between RNA metabolism and genome integrity

After the diverse mRNA processing steps, the mRNP is mature and ready for nuclear export. Initial studies in *S. cerevisiae* about the mRNA export pathway

allowed the identification of the key components of this machinery (Aguilera, 2005a; Reed and Cheng, 2005; Reed and Hurt, 2002; Vinciguerra and Stutz, 2004). These include THO/TREX, a conserved eukaryotic complex involved in transcription elongation and RNA export that contributes to the mRNP biogenesis. In this interface, the THO complex associates with a number of RBPs including the DEAD-box RNA helicase UAP56/DDX39B (yeast Sub2) or the mRNA export adaptor protein ALY/REF (yeast Yra1) in a supramolecular structure termed TREX (Transcription-Export) coupling transcription and export (Cheng et al., 2006; Jimeno et al., 2002; Masuda et al., 2005; Strasser et al., 2002).

Yeast THO/TREX mutants present mRNA export defects, transcription-elongation impairment and transcription-dependent-hyper-recombination phenotypes (Aguilera and Klein, 1990; Chavez et al., 2000; Jimeno et al., 2002; Piruat and Aguilera, 1998). This is explained by the formation of a suboptimal mRNP, which can impair transcription elongation and cause genome instability, all of which may be explained by an increase in the R loop accumulation (Huertas and Aguilera, 2003). Since THO contributes to the co-transcriptional formation of export-competent mRNP (Rondon et al., 2010), THO mutants increase the possibility to hybridize back to the DNA template during transcription due to an improper mRNP assembly (Aguilera, 2005b). Accordingly, THO-mutant phenotypes in yeast can be suppressed by the overexpression of specific RNA-binding proteins such as Sub2 (human UAP56) or Tho1. Thereby, yeast and human cells deprived of THO exhibit strong transcription- and R loop dependent genome instability phenotypes that are accompanied by altered RF progression (Dominguez-Sanchez et al., 2011; Gomez-Gonzalez et al., 2011; Wellinger et al., 2006).

Moreover, genome-wide studies in yeast pointed that THO/TREX complex is recruited to the active transcribed chromatin (Gomez-Gonzalez et al., 2011). In this line, recent work in our laboratory has shown a physical and functional connection between THO complex and the histone deacetylase mSin3A co-

repressor complex to suppress co-transcriptional R loops, DNA damage and replication impairment. Thus, THO may also play a key role in promoting local and transient chromatin closing via mSin3A to prevent harmful RNA-DNA hybrids (Salas-Armenteros et al., 2017).

This connection between RNA metabolism and genome integrity has been extended to other factors. These involved different RNA nuclear processes which are also associated to genome instability. Among them, there have been found proteins with a role in splicing as is the case of SRSF1 (Li and Manley, 2005), mRNA 3' end processing and degradation factors such as Trf4, Rrp6 and FIP1L1 (Gavalda et al., 2013; Luna et al., 2005; Stirling et al., 2012) or helicases such as SETX/Sen1, DDX19 and DDX23 (Hodroj et al., 2017b; Skourti-Stathaki et al., 2011; Sridhara et al., 2017).

2.3. DEAD-box RNA helicases and genome integrity

In the last years, there has been an emerging interest in the role of different helicases in the maintenance of genome integrity. Helicases are nucleic acid–dependent ATPases capable of unwinding DNA or RNA duplex substrates. They play roles in almost every process involving nucleic acids, including DNA replication and repair or transcription (Singleton et al., 2007). Among them, the DEAD/H box family is the largest family of superfamily 2 helicases, which in turn comprehends two groups: the DEAD-box group with 44 members and the DEAH-box family with 15 proteins. The majority of these helicases use ATP to bind or remodel RNA and RNA-protein complexes and for this reason, they are involved in nearly all aspects of RNA metabolism, from transcription and translation to mRNA decay. These proteins are characterized by the presence of an Motif/Motif II, D-E-A-D (asp-glu-ala-asp) or D-E-A-H (asp-glu-ala-his), which inspired the name of the family (Linder and Jankowsky, 2011). Hence, DEAD-box proteins can act as RNA chaperones (unwinding and refolding of RNA (Jarmoskaite and Russell, 2014; Putnam and Jankowsky, 2013). Importantly, emerging evidence suggest that these helicases are involved in genome instability and cancer onset since mutations in many of them lead to the appearance of such phenotypes (Cai

et al., 2017; Fuller-Pace, 2013; Sarkar and Ghosh, 2016). Interestingly, these features are linked to R loop imbalance in the cases of DHX9 (Cristini et al., 2018), DDX1 (Li et al., 2016) DDX23 (Sridhara et al., 2017), DDX21 (Song et al., 2017) or DDX19 helicases (Hodroj et al., 2017a), among others.

2.3.1. The helicase DDX5/DBP2

Human DDX5/p68 (yeast Dbp2) is one of the prototypic members of the DEAD box family of RNA helicases (Ford et al., 1988). It was initially identified due to its cross-reactivity with an antibody generated against the simian virus (SV40) large T antigen (Lane and Hoeffler, 1980). DDX5 and the highly related protein DDX17/p72 (Lamm et al., 1996) are RNA-dependent ATPases and ATP-dependent RNA helicases. Both proteins share 90% identical central core but with different N- and C-termini. DDX5 presents the characteristic “helicase core” divided into two flexibly linked RecA-like domains, which are critical for RNA binding, ATP binding and hydrolysis, and intermolecular interactions. (Dai et al., 2014). Moreover, human DDX5 possesses 10-fold higher unwinding activity than Dbp2, partially due to the presence of a mammalian specific C-terminal extension (Xing et al., 2017). DDX5 and DDX17 have been shown to function in multiple cellular processes, in most cases as RNP chaperones (Figure I3). However, despite these proteins being predominantly nuclear, different reports have shown that DDX5 is a nucleocytoplasmic shuttling protein (Iggo et al., 1991; Wang et al., 2009).

DDX5/Dbp2 act in multiple steps of RNA metabolism. It is involved in the regulation of transcription through lncRNAs (Wongtrakoongate et al., 2015; Zhang et al., 2016); mRNP processing, being required for the spliceosome assembly and alternative splicing (Dardenne et al., 2014; Wang et al., 2016), and mRNP export. Curiously, DDX5 co-immunoprecipitates with the export factors ALY and TAP (Zonta et al., 2013) and likewise Dbp2 in *S. cerevisiae* genetically interacts with Yra1 and Mex67 (Ma et al., 2016). In addition to these roles, it has also been proposed to be involved in the regulation of the mRNA levels in the nonsense-mediated decay, an RNA surveillance pathway that targets a selection

of mRNAs for degradation (Geissler et al., 2013; Lykke-Andersen and Jensen, 2015), and of microRNA processing, which are non-coding RNAs that target mRNAs for silencing (Gregory et al., 2004; Ha and Kim, 2014), and they also are involved in ribosome biogenesis (Saporita et al., 2011). DDX5 also acts as a transcriptional co-factor to activate transcription, as is the case of MyoD (Caretti et al., 2006), Snail1 (Carter et al., 2010) or p53 (Bates et al., 2005). (Figure I3). Finally, recent reports have also uncovered broader effects of DDX5 in glucose metabolism and carcinogenesis (Mazurek et al., 2014; Xing et al., 2017) through the regulation of gene expression. The importance of DDX5 in tumor development is also provided by the fact that its overexpression is well established in breast cancer (Wortham et al., 2009). Therefore, it has been proposed that the multiple roles in which DDX5/Dbp2 is found to intervene are dictated by the wide range of RNA targets acted on by DDX5/Dbp2 (Xing et al., 2019).

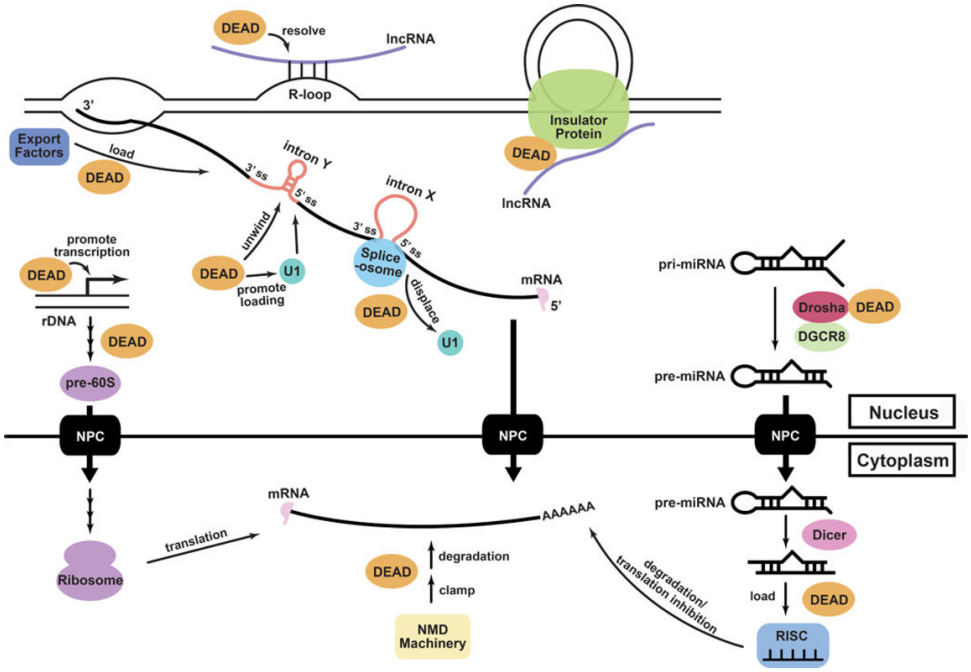


Figure I3. DDX5/Dbp2 functions as a RNP chaperone.
DDX5/Dbp2 (orange circle, "DEAD") participates as a RNP chaperone in multiple processes such as alternative splicing, the regulation of lncRNA activities, mRNA export, miRNA processing, nonsense mediated decay (NMD) and ribosome biogenesis. Adapted from (Xing et al., 2019).

2.3.2. The helicase UAP56/DDX39B: a key player in mRNP biogenesis

The DEAD-box helicase UAP56/Sub2 is a conserved RNA-dependent ATPase and ATP-dependent RNA helicase discovered through its interaction with the human splicing factor U2AF65 (Fleckner et al., 1997). Human UAP56 comprises the minimal helicase core with two canonical RecA-like helicase domains connected with a flexible liner (Figure I4). The N-terminal domain contains the conserved helicase motifs IV, motifs Q, I, Ia, GG, Ib, II, and III, whereas the C-terminal domain contains helicase motifs IV, QXXR, V, and VI. (Shi et al., 2004; Zhao et al., 2004). Structures of several DEAD/H-box proteins and other helicases have shown that Lys in motif I (GKT) interacts with the phosphate group of ATP and is important for ATP binding (Sengoku et al., 2006; Shi et al., 2004). The Glu in motif II (DEAD) has been postulated to be the key catalytic residue that activates a water molecule to hydrolyze ATP in DEXD/H-box proteins and other helicases (Cordin et al., 2006; Sengoku et al., 2006).

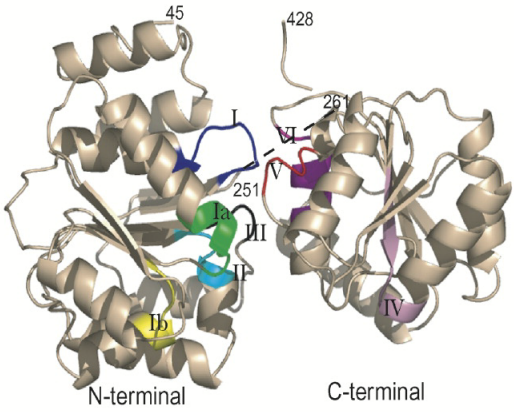


Figure I4. Human UAP56 structure.
Picture represents both N-terminal and C-terminal domain joined by the flexible linker (black line). Conserved helicase motifs are mapped on the structure in different colors. Adapted from (Zhao et al., 2004)

Specifically, UAP56 plays a role in splicing by facilitating the unwinding of U4/U6 snRNA duplex and it is also required for the first ATP-dependent spliceosome assembly step (Fleckner et al., 1997; Shen et al., 2008). UAP56 is also essential for the export of the majority of mRNAs from the nucleus to cytoplasm in *D. melanogaster*, *S. cerevisiae*, *C. elegans* and human (Herold et al., 2003; Kapadia et al., 2006; Luo et al., 2001; MacMorris et al., 2003). It recruits the mRNA export factor ALY to the intron-containing and intron-less pre-mRNAs.

In the first case, UAP56 remains as part of the exon junction complex (EJC) and couples splicing and mRNA export. In the second scenario, UAP56 and ALY together with the THO complex participate in the export of bulk mRNAs through the interaction with the NXF1 complex linking the nascent mRNA to its export factors, which allow translocation to the NPC (Strasser et al., 2002). UAP56/Sub2 is not an integral part of THO (Chavez et al., 2000; Pena et al., 2012) (Figure I5). However, it displays an important role in the assembly of the TREX complex by mediating the ATP-dependent interaction of CIP29, Aly, PDIP3 and ZC11A with THO complex (Dufu et al., 2010; Folco et al., 2012; Sugiura et al., 2007). The TREX components Aly and Chtop stimulate the UAP56 helicase activity and are recruited to mRNAs by UAP56 (Chang et al., 2013). UAP56/Sub2 is the most abundant protein of the THO/TREX complex (Heath et al., 2016). Interestingly, an additional UAP56-related helicase called URH49/DDX39A is also present in humans (Pryor et al., 2004). This protein shares 90% amino acid sequence homology with UAP56. Both proteins, UAP56 and URH49, are able to complement Sub2 deletion and interact with the Yra1 adaptor in yeast. However, URH49 has been found to bind preferentially to CIP29 (yeast Tho1) in what it is believed to be a novel complex called AREX (alternative mRNA export complex) (Yamazaki et al., 2010).

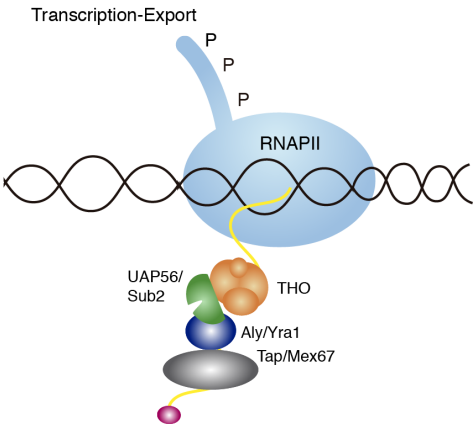


Figure I5. Role of UAP56 in transcription and nuclear export.
During transcription, the THO complex associates with the RNAPII travelling along the entire transcribed region. UAP56 and Aly associate with the 3'-end of the gene forming the active TREX complex. Thanks to the UAP56 helicase activity, different proteins could be transferred to the mRNAs. Afterwards, the Tap-p15 heterodimer recognizes the mRNAs through Aly facilitating the formation of an export competent mRNP. Finally during or after release from genes locus, mRNPs can be subject to more rearrangements prior its translocation through the nucleopores and its following translation in the cytoplasm.

In yeast, Sub2 is involved in multiple stages of mRNA maturation and its inactivation leads to non-productive spliceosome assembly (Kistler and Guthrie, 2001; Libri et al., 2001), decreased polyadenylation efficiency and mRNA instability (Rougemaille et al., 2008; Saguez et al., 2008). Therefore, this RNA helicase seems to play a chaperone role in the process of assembly of the mRNP (Saguez et al., 2013). In addition, Sub2 also plays a role in the maintenance of genome instability since mutations of *SUB2* lead to hyper-recombination phenotype, whereas Sub2 overexpression partially suppresses the growth-defect and hyper-recombination phenotypes associated to the *hpr1Δ* mutants of the THO complex (Jimeno et al., 2002). In human cells, UAP56 depletion leads to mRNA export defects and a strong genomic instability phenotype (Dominguez-Sanchez et al., 2011). Altogether, these findings suggest a double role of UAP56/DDX39B in mRNP biogenesis/export and the maintenance of genome, as it seems to be also the case of THO.

3. R LOOPS AS A SOURCE OF GENOME INSTABILITY

Co-transcriptional R loops represent one of the elements that contributes to exacerbate the negative impact of transcription and transcription-replication conflicts on genome integrity. These three-stranded-structures are the result of the hybridization of the nascent RNA with the transcribed DNA strand, leading the complementary strand as ssDNA ("thread-back" model). (Figure I6).

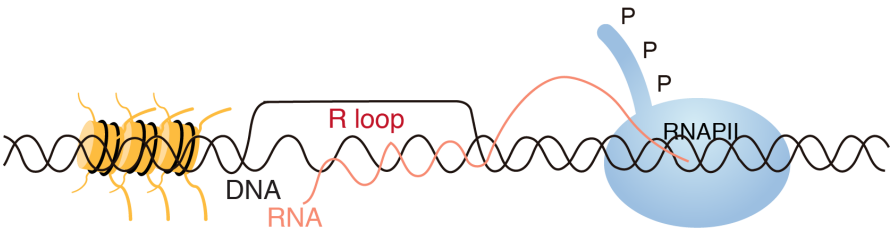


Figure I6. R loop structure.
R loops are the resulting structures formed by a RNA-DNA hybrid and the displaced ssDNA.

3.1. R loops as a cellular regulators

RNA-DNA hybrids occur naturally during transcription and replication. Despite their well-known localization at the active site of the RNA polymerase at the transcription bubble or during lagging-strand synthesis (Aguilera and Garcia-Muse, 2012), they have also found around DSB sites regulating the strand resection process (Ohle et al., 2016). Similarly, reports have widely described transient programmed R-loops with a role in different physiological processes such as *E. coli* plasmid replication (Kogoma, 1997), mitochondrial DNA replication (Xu and Clayton, 1996), or immunoglobulin (Ig) class-switch recombination in the B-cells (Yu et al., 2003). Strikingly, recent reports have also shown the implications of R loops in suppression of chromosome instability through R loop-driven ATR pathway that acts at centromeres and promotes faithful chromosome segregation (Kabeche et al., 2018).

R loops are formed in multiple regions of the eukaryotic genomes, such as the ones transcribed by RNA polymerases I, II and III (El Hage et al., 2010; Tran et al., 2017). Furthermore, at RNAPII-transcribed genes, genome-wide approaches have provided evidence showing that the R loops are preferentially found at promoters of genes enriched in CpG island showing a strong GC skew (Chen et al., 2017; Dumelie and Jaffrey, 2017; Ginno et al., 2012; Nadel et al., 2015; Sanz et al., 2016). At promoters, it has been shown that R loops may assist transcription by suppressing DNA methylation (Ginno et al., 2012; Grunseich et al., 2018) and by influencing the binding of chromatin remodelers. On one hand, in mouse embryonic stem cells (mESCs) R loops inhibit repressive chromatin-modifying enzymes and recruit activating chromatin-remodeling complexes such as Tip60-p400 to favor differentiation genes (Chen et al., 2015). On the other hand, R-loops can impede transcription-factor binding as occurred in the case of the human vimentin gene (Boque-Sastre et al., 2015). This dual situation may reflect the capacity of R loops in regulating gene expression through multiple context-dependent mechanism (Crossley et al., 2019).

Importantly, R loops have also been found to be enriched at the 3' end of some mammalian genes where they are supposed to mediate efficient transcription termination (Skourti-Stathaki et al., 2011). R loops can hamper transcription and it could be used 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). Thus, these findings suggest that R loops might play regulatory roles such as the control of the gene expression through epigenetic mechanisms or by transcriptional interference (Chan et al., 2014; Ginno et al., 2012; Wahba et al., 2016). Importantly, despite their important regulatory roles, they can also hinder RF progression and may also occur when R loops interfere with productive transcription.

3.2. R loops as genomic threats

3.2.1. Structural R loop-mediated genome instability

Given its structure and effects, R loops constitute a putative threat to genome stability. As said above, the formation of a RNA-DNA hybrid and the concomitant displacement of the ssDNA drive to a major exposure the DNA strand, which is more mutagenic than dsDNA. This could lead to SSBs, that if remaining unrepaired might block RF progression and could lead to DSB formation (Figure I7A). The ssDNA can be object of the action of different mutagenic agents or DNA-modifying enzymes, as it is the case of the human activation-induced cytidine deaminase (AID) (Figure I7A), which participates in immunoglobulin class switching recombination and somatic hypermutation in mammalian activated B cells (Chaudhuri and Alt, 2004). However, the action of AID is not restricted to such genes and it could act over other transcribed genes, particularly when R loops accumulate at high levels, as it occurs in THO yeast mutants (Gomez-Gonzalez and Aguilera, 2007). In addition, R loops can also promote hypermutation by non-canonical replication that arises when the R loop is used as a primer, as described in *E. coli* (Kogoma, 1997) or in the rDNA region of yeast deprived of Top1 and RNase H enzymes (Stuckey et al., 2015). Furthermore, the ssDNA is also more predisposed to generate non-B DNA structures which could

also be an impediment for the replisome (Aguilera, 2002). Due to this fact, an increase in R loops would lead to a hypermutation phenotype.

R loops can also alter transcription. Excessive pausing, arrest or backtracking of RNP can lead to transcription stress (Saponaro et al., 2014).Hybridization between the nascent RNA and the DNA template strand within an R loop could destabilize the transcription complex, that could render the RNAP prone to blockage at randomly-occurring weak pausing/termination mode. Indeed, R loops have been shown to become obstacles for RNA polymerase progression *in vitro* (Tous and Aguilera, 2007) and *in vivo* (Bonnet et al., 2017; Lang et al., 2017) Although it is not clear 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.

3.2.2. R loops and replication stress

Since the most prevalent source of R loop-dependent DNA damage seems to take place during S phase (Gan et al., 2011; Wellinger et al., 2006), R loops are supposed to exacerbate transcription-replication conflicts. There is a large body of work that has provided a mechanistic connection between recombination and replication-stress induced by R loops. The DSB repair factor and tumor suppressor BRCA2 and other members of the Fanconi Anemia pathway, which act at RF, have been shown to suppress R loops to prevent damage arising from these conflicts (Bhatia et al., 2014; Garcia-Rubio et al., 2015; Hatchi et al., 2015; Schwab et al., 2015). Indeed, reports indicate that head-on conflicts are more deleterious than codirectional encounters (Prado and Aguilera, 2005). Accordingly, recent studies show that R loop levels are also affected by the orientation of the conflict (Hamperl et al., 2017; Lang et al., 2017). This suggests that head-on formed R loops blocked replication and can result in DNA breaks and in the end, genome instability and chromosome fragility (Figure I7B) (Aguilera and Garcia-Muse, 2012; Zeman and Cimprich, 2014). Evidence from yeast to human cells corroborate this idea. RF progression defects are detected in several R loop-accumulating cells by different techniques as 2D-gel

electrophoresis, DNA combing or the recruitment of the Rrm3 DNA helicase (Gomez-Gonzalez et al., 2011; Herrera-Moyano et al., 2014; Salas-Armenteros et al., 2017; Tuduri et al., 2009; Wellinger et al., 2006).

From the mechanistic point of view, these differences between co-directional and head-on R loop might be based on the fact that RNAP can remain attached to the template, impeding the access of the replicative helicases to the R loop at head-on collisions. Conversely, in co-directional conflicts the replisome itself may unwind hybrids. However, the mechanism by which R loops block RF is still unclear. On one hand, RNA-DNA hybrids can retain the RNAP or recruit proteins that constitute a barrier. On the other hand, R loops can promote positive superhelicity accumulation between head-on transcription-replication conflicts or even it could be possible that the impediment is not the RNA-DNA hybrid by itself but the compacted chromatin structure induced by it (Castellano-Pozo et al., 2013; Garcia-Pichardo et al., 2017; Rondon and Aguilera, 2019) .

3.2.3. Chromatin modifications

Recent finding suggests that epigenetic mechanisms can act to regulate R loop homeostasis. Chromatin acetylation has been proposed to influence the R loop formation. A hyper-acetylated state of the chromatin could lead to a more opened and relaxed structure that would theoretically facilitate the probability of R loop generation. In human cells, it has been shown that depletion of the histone deacetylase mSin3a complex or the acetyltransferase MOF leads to an accumulation of R loops and genome instability (Salas-Armenteros et al., 2017; Singh et al., 2018). In addition, the treatment with histone deacetylase inhibitors such as trichostatin A (TSA) boosted a major acetylated state of the chromatin and R loop accumulation (Salas-Armenteros et al., 2017).

However, the accumulation of aberrant R loops is linked to heterochromatin and chromatin condensation marks in yeast, *C. elegans* and human cells. Specifically, depletion of the THO complex or the helicase SETX/Sen1 triggers an accumulation of histone H3 serine 10 phosphorylation

(H3S10-P) in yeast and human cells, a mark of chromatin condensation (Castellano-Pozo et al., 2013). Given that replisomes have difficulties to advance through a condensed chromatin (Castellano-Pozo et al., 2012b; Castellano-Pozo et al., 2013; El Achkar et al., 2005), it is logical to speculate about the role of chromatin condensation as a cause of RF stalling and genome instability mediated by R loops (Figure I7C) (Castellano-Pozo et al., 2013). Another piece of evidence linking chromatin with R loops is the FACT complex, necessary for the proper DNA replication and whose deprivation leads to an increase in R loops and genomic instability in yeast and human cells (Herrera-Moyano et al., 2014). Interestingly, recent studies in yeast have identified specific histone mutants that accumulate R loops without compromising genome integrity. This fact suggests that R loops may not generate DNA damage by themselves. Instead, it would require a subsequent chromatin-remodeling step connected to chromatin compaction that includes H3S10-P to cause genome instability (Garcia-Pichardo et al., 2017). Subsequent studies have also shown the H3S10-P accumulation at centromeres together with R loops and ATR, supporting the link between R loops and H3S10-P in these regions (Kabeche et al., 2018).

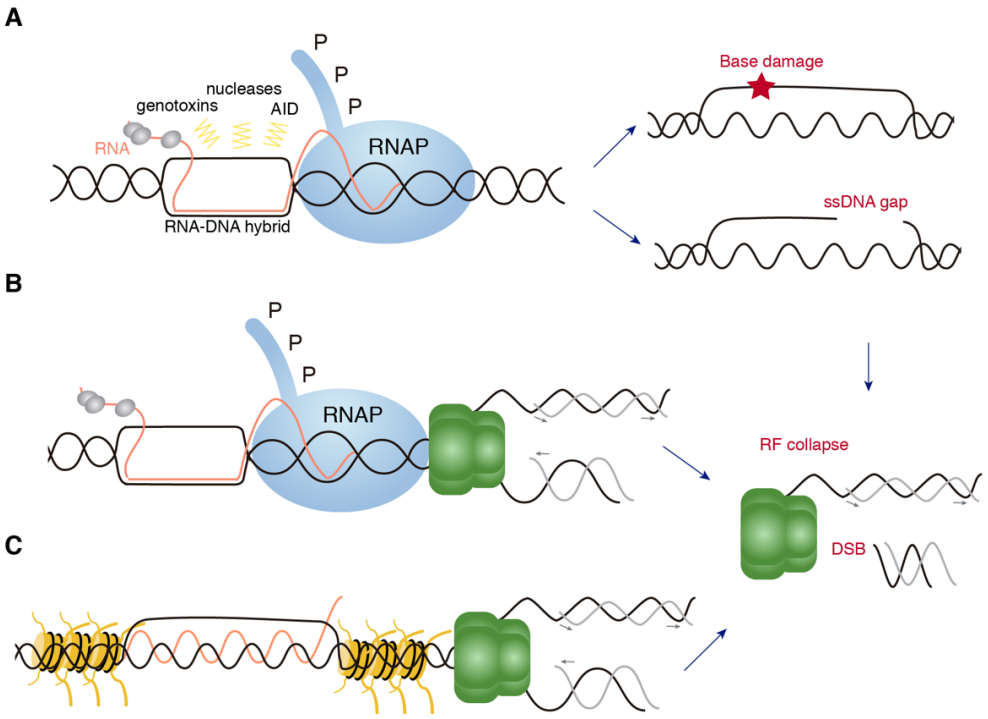


Figure I7. R loop-mediated genome instability.
R loops by themselves constitute a major source of genome instability. (A) R loop formation implies an increase in the exposition of the ssDNA, which is more sensible to suffer assaults from genotoxins or enzymatic activities such as AID. This could lead to DNA lesions such as base damage (red star) or ssDNA gaps, which can hamper RF progression leading to genome instability. (B) R loops by themselves could constitute an impediment for replication facilitating transcription-replication collisions. (C) Chromatin compaction after R loop formation could also hamper replication. Definitely, R loops could constitute a major barrier for RF progression and lead to fork stalling, subsequently fork collapse and breakage, leading to genome instability. Adapted from (Gaillard and Aguilera, 2016).

3.3. Mechanisms and elements involved in R loop prevention

As we previously indicated, R loops can occur naturally in cells. They are regularly generated over a large part of the genome (Chedin, 2016). Consequently, R loop homeostasis need to be regulated. For this purpose, cells have evolved different protection mechanisms and factors to overcome the R-loop formation and mitigate their harmful effect if deregulated (Santos-Pereira and Aguilera, 2015; Skourti-Stathaki and Proudfoot, 2014). In general, these factors are classified in two categories: those that prevent R loops formation (topoisomerases and mRNP biogenesis) and those which remove R loops once formed (ribonucleases, helicases and others).

3.3.1. Topoisomerases

During transcription, local topological changes need to be controlled since the accumulation of negative supercoils behind the RNAP is thought to facilitate R loop formation. Topoisomerases can relax negative supercoiling and prevent R loop accumulation. This has been observed in bacteria where the growth defect of *topA* mutants, defective in Top1, are rescued by RNase H1 overexpression (Drolet et al., 1995), and in yeast, in which *top1* and *top2* mutants present R loop accumulation at the rDNA locus (El Hage et al., 2010), and human cells, where TOP1-deficient cells show DNA breaks at active genes, replication defects and R loop accumulation that are suppressed by RNase H1 overexpression (Manzo et al., 2018; Tuduri et al., 2009).

3.3.2. mRNP biogenesis

The nascent mRNA is considered other important element that is target of the R loop prevention mechanisms, since deficiencies in the assembly of the messenger ribonucleoparticle (mRNP) can prompt to R loop formation. Eukaryotic gene expression is a tightly coupled process where transcription and mRNA processing need to be coordinated for the export of a functional mRNP. During this process, RNA-binding proteins (RBPs) associate with the nascent mRNA, contributing to the conformation of a correctly packaged and coated RNA molecule and thus, preventing its hybridization with the DNA template and R loop formation (Rondon et al., 2010). Evidence considering co-transcriptional R-loops as a source of genome instability were first shown in yeast cells where the absence of the yeast THO/TREX complex lead to hyperrecombination and instability phenotypes (Huertas and Aguilera, 2003). This complex binds to the nascent mRNA and ensure its assembly into a proper mRNP to be efficiently exported. Remarkably, the role of this complex seems to be conserved since deficient cells of yeast, *C. elegans* and human exhibit analogous transcription and mRNA export effects, as well as R loop-mediated genome instability (Castellano-Pozo et al., 2012a; Chavez and Aguilera, 1997; Dominguez-Sanchez et al., 2011; Gomez-Gonzalez et al., 2011; Huertas and Aguilera, 2003).

Later on, many other transcription factors involved in R loop prevention have come to light. For instance, the human serine/arginine-rich splicing factor 1 (SRSF1) gene involved in splicing and mRNA export also prevents R loop formation (Li and Manley, 2005). In addition, the mRNAs half-life is another target to prevent R loop formation, as is the case of the Trf4 (polyadenilation polymerase) component of the TRAMP complex (Gavalda et al., 2013) or the human exosome components EXOSC3 and EXOSC10 (Pefanis et al., 2015) (Figure I8A). In fact, many global and specific studies have uncovered more elements related with RNA metabolism that have a role in R loop prevention (Chan et al., 2014; Paulsen et al., 2009; Stirling et al., 2012; Wahba et al., 2011). Consequently, these findings have evidenced that almost every defective RNA processing function can lead to R-loop accumulation, strengthening the importance of the RNA protection to guarantee genome integrity.

3.3.3. Ribonucleases, helicases and others

Since R loops are dynamic and reversible structures, when formed, they can be removed by different mechanisms that prevent their long-live and accumulation. R loops can be directly resolved by RNase H enzymes: RNase H1 (monomeric) and RNase H2 (three subunits in eukaryotes). Despite their different composition and specialized roles, both proteins are able to degrade the RNA moiety of an RNA-DNA hybrid (Cerritelli and Crouch, 2009; Skourti-Stathaki and Proudfoot, 2014). However, RNase H1 is considered to be the key player in eliminating co-transcriptional R loops (Chon et al., 2013) and its overexpression is continuously used to suppress R loop-dependent genome instability phenotypes.

Additionally, R loops can be removed by RNA-DNA helicases which unwind the RNA-DNA hybrids. One well-studied example is the helicase senataxin SETX (Sen1 in yeast). This protein and others such as DHX9 are proposed to unwind R loops, specially at transcription termination pause sites (Cristini et al., 2018; Mischo et al., 2011; Skourti-Stathaki et al., 2011). Thus, after the unwinding of the RNA-DNA hybrid, the nascent RNA is exposed for its

degradation by exonucleases such as XRN2, leading to termination. Subsequently, absence of either of these factors results in R loop accumulation and altered termination (Cristini et al., 2018; Skourti-Stathaki et al., 2011). In this context, the human RNA helicase aquiarius (AQR) protein, which belongs to the same family as SETX, has been also suggested to have a role in R loop resolution since its depletion lead to an R loop accumulation phenotype (Sollier et al., 2014) (Figure I8B).

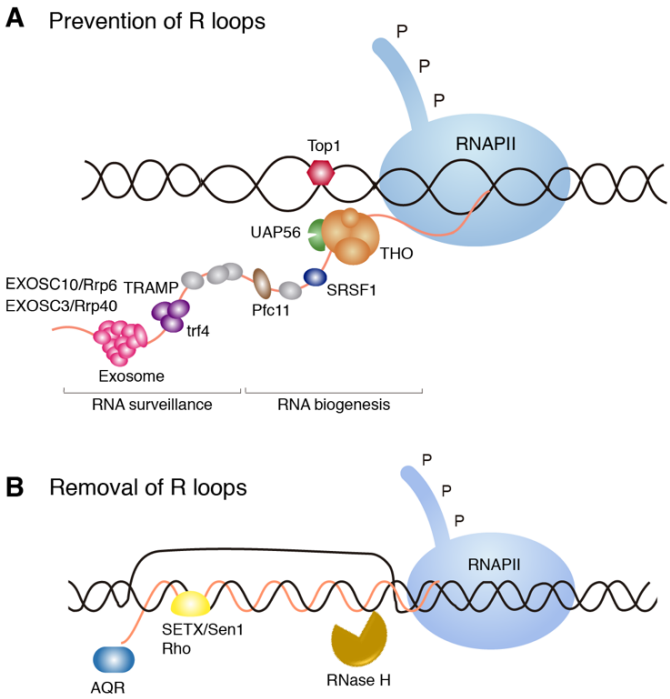


Figure I8. Action to prevent or resolve R loop accumulation.
(A) R loop control can occur at different steps during transcription. On one hand, topoisomerases (TOP1) avoid supercoiling accumulation that could facilitate R loop formation. On the other hand, specific RNA-binding proteins acting at different steps of RNA metabolism from RNA biogenesis (THO complex, UAP56, SRSF1 and Pcf11) to RNA surveillance (including the exoribonucleases exosome component 3 (EXOSC3) and EXOSC10 (Rrp40 and Rrp6 in yeast, respectively) and the TRAMP complex prevent the RNA to hybridize back with DNA. **(B)** R loop removal could be achieved by different mechanisms. RNase H1 and RNase H2 both recognize RNA-DNA hybrids and degrade the RNA moiety. Moreover, helicases such as SETX/Sen1 in human and yeast, Rho in bacteria and putative ones such human aquarius (AQR) could unwind R loops. Adapted from (Santos-Pereira and Aguilera, 2015).

There are other factors that normally resolve other forms of stresses that might indirectly help resolve R loops. For instance, the transcription-coupled nucleotide excision repair (TC-NER) factors can help process R loops (Shivji et al., 2018; Sollier et al., 2014; Yasuhara et al., 2018). Although the mechanism

involved is not completely understood, it is proposed that TC-NER nucleases XPG and XPF could excise R loops, leaving a ssDNA gap that could progress to a DSB (Sollier et al., 2014). Importantly, the breast cancer susceptibility factors BRCA1 and BRCA2 have also been found to help resolve R loops (Bhatia et al., 2014; Hatchi et al., 2015; Shivji et al., 2018). Generally, as both factors associates with RNAPII, they are able to suppress R loop-mediated transcriptional stress by promoting elongation and resolving R loops (Shivji et al., 2018; Zhang et al., 2017). It is known that BRCA1 recruits the helicase SETX to remove R loops at termination sites. Fanconi Anemia factors or the FACT chromatin reorganizing complex involved in RF progression also help prevent R-loop accumulation and R loop-mediated genome instability at transcription-replication conflicts (Figure I9) (Garcia-Rubio et al., 2015; Herrera-Moyano et al., 2014).

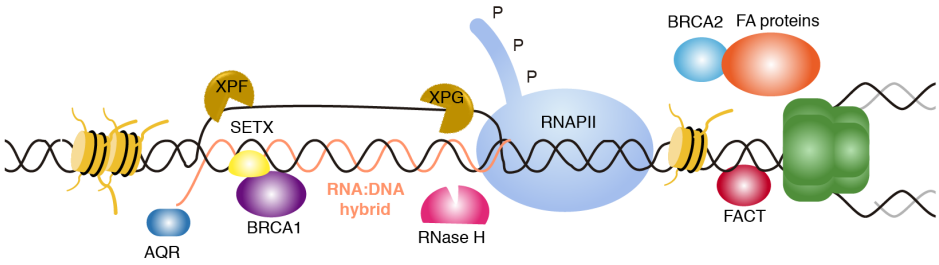


Figure I9. Multiple factors involved in R loop-mediated T-R conflicts.
R loops and R loop-mediated chromatin compaction can hamper RF progression. In addition to the already mentioned RNase H and SETX, which can be recruited by BRCA1 to R loop sites, there are other factors implied in R loop resolution under these circumstances. For instance, BRAC2, TC-NER factors and potentially other Fanconi anemia (FA) proteins or the FACT chromatin-reorganizing complex can help to counteract this situation.

3.4. Techniques to detect R loop accumulation

Direct detection of R loop have relied on the use of the S9.6 monoclonal antibody, which recognizes RNA-DNA hybrids. However, the suspicion of the lack of specificity of this antibody has always been a concern. Indeed, recent reports have demonstrated its ability to bind also to dsRNA (Hartono et al., 2018; Silva et al., 2018). As a consequence, other indirect strategies have been performed to demonstrate the existence of R loop accumulation, including the suppression of R loop-dependent phenotypes by RNase H overexpression. In particular, RNase H1 presents a hybrid-binding domain (HBD) in the N-terminus (Nowotny

et al., 2008), whereas the RNase H2 main activity is to cleave a single ribonucleotide embedded in the DNA duplex (Eder et al., 1993). The combination of both strategies using the S9.6 and the subsequent overexpression of RNase H to suppress the R loop-dependent phenotype has served to overcome the S9.6 limitations. In addition to S9.6 immunofluorescence (IF) or DRIP (DNA-RNA hybrid immunoprecipitation), numerous techniques have been applied for genome-wide detection of R loops using the S9.6 antibody such as DRIP followed by sequencing (DRIP-seq) (Ginno et al., 2012) or the most accurate DRIP followed by cDNA conversion coupled to high-throughput sequencing (DRIPc-seq) (Sanz et al., 2016). In addition, bis-DRIP-seq (Dumelie and Jaffrey, 2017) combines *in situ* ssDNA bisulfite footprinting with S9.6 hybrid pull-down to theoretically improve the specificity by targeting both, the hybrids and the ssDNA. However, there have been increasing efforts in the field to abandon the use of the S9.6 antibody. Due to these facts, new alternatives beyond the use of the S9.6 have been proposed such as R-ChIP, where the expression of the exogenous catalytically inactive form of the RNase H1 is followed by chromatin immunoprecipitation (ChIP) of the tagged RNase H1 (Chen et al., 2017) or the use of the fusion protein HB-GFP formed by the 52-residue DNA-RNA hybrid-binding (HB) domain of the RNase H1 and the enhanced green fluorescent protein (eGFP), which is able to detect RNA-DNA hybrids *in vivo* (Bhatia et al., 2014).

4. EPITHELIAL-MESENCHYMAL TRANSITION AND THE TRANSCRIPTION FACTOR SNAIL1

4.1. Epithelial-Mesenchymal transition

The epithelial-mesenchymal transition (EMT) is a crucial and reversible cell plasticity program occurring during embryonic development and in adult tissue homeostasis. When completed, it promotes the transition from an immotile epithelial to a motile mesenchymal cell type. When aberrantly activated, it could lead to pathogenic effects, particularly cancer. Phenotypically, cells undergoing EMT lose their apical-basal polarity, tight intercellular contacts and interaction

with the basal membrane. They acquire a spindle-like appearance, gain motility and invasiveness. These profound cellular attribute changes require an extensive transcriptional reprogramming, that relies on down- and upregulation of epithelial and mesenchymal gene expression programs, respectively. Under a wide range of pleiotropic signaling factors, the EMT is triggered in association with the activation of the expression of specific transcription factors such as Snail1 (SNAI1), Slug (SNAI2), Twist-related protein 1 (TWIST1) and zinc-finger E-box-binding homeobox 1 (ZEB1) and ZEB2; and miRNAs and epigenetic and post-translational regulators. These transcription factors are responsible of the control of the expression of two broad functional groups of genes: epithelial genes, whose expression is repressed and, mesenchymal genes, whose expression is induced. From a molecular point of view, the hallmark of EMT is the down-regulation of E-cadherin. Other epithelial markers such as claudins and occludins are also down-regulated. While Fibronectin 1 (FN1) or Vimentin, typical mesenchymal markers are up-regulated during EMT (Figure I10) (Nieto et al., 2016; Stemmler et al., 2019).

As mentioned before, EMT is also important under pathological circumstances, where a reactivation of developmental programs take place, such as cancer progression. EMT is thought to enhance stemness of cancer cell during the process of tumor metastasis, facilitating the execution of most of the steps of the invasion-metastasis cascade. Invasion permits the translocation of tumor cells from the initial neoplastic focus into neighbouring host tissues, allowing them to penetrate vessel endothelium and enter the circulation to form distant metastasis. (Guarino et al., 2007; Mani et al., 2008; Thiery, 2003; Thiery et al., 2009).

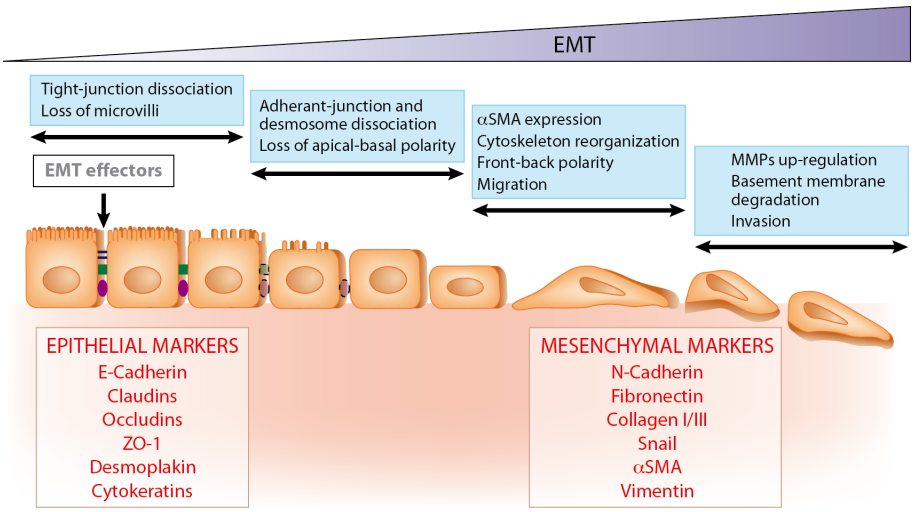


Figure I10. Principal traits during epithelial-mesenchymal transition.
During this process cells undertake different events to accomplish the entire EMT course. Image also shows the main epithelial and mesenchymal markers. Adapted from (Aroeira et al., 2007).

4.2. SNAIL1

Snail1 is a conserved zinc-finger transcription factor. Its C-terminal domain contains four zinc fingers of the C_2H_2 type that bind to the E-box motif 5′-CANNTG-3′ or 5'-CAGGTG-3' in target gene promoters. The divergent N-terminal region is involved in the repressive activity and is essential for the interaction with several co-repressor complexes such as mSin3A, Ajuba LIM proteins or the Polycomb repressive complex 2 (Wang et al., 2013). Lastly, the central part of the protein comprises a nuclear export signal (NES), a destruction box domain and a serine-proline rich region, involved in the stability and subcellular location of the protein (Figure I11).

Importantly, Snail1 has been purposed as a prototype inductor of EMT. It is fast induced by citokynes or stress conditions activating EMT and it is able to bind and repress the expression of E-cadherin and other epithelial genes and activate the expression of other mesenchymal genes. Given its relevance in this process, Snail1 expression and function are regulated at multiple levels from gene transcription to protein modifications, affecting its interaction with specific cofactors including TGFβ, NOTCH, WT or NF-κB (Thiery and Sleeman, 2006).

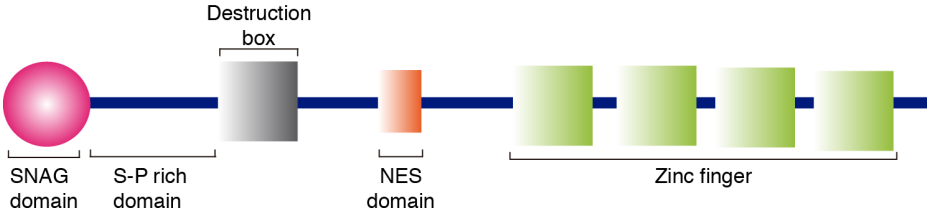


Figure I11. Structure of Snail1.
Snail1 contains an N-terminal SNAG domain, which interacts with different co-repressors and epigenetic remodeling complexes and the C-terminal zinc finger domain for DNA binding. The serine-rich domain, the destruction box and the nuclear export sequence (NES) are involved in the stability and subcellular location of the protein. Adapted from (Peinado et al., 2007)

Research in the last decade has revealed a connection between Snail1 and the co-repressor interactor LOXL2 in regulating major satellite transcription and heterochromatin reorganization during EMT. Upon the binding of Snail1 and LOXL2 to pericentromeric regions, HP1α is released from heterochromatin with a down-regulation of major satellite transcription, enabling chromatin reorganization and acquisition of mesenchymal traits. Notably, the histone-modifying enzyme LOXL2 replaces the active mark from histone H3 (H3K4me3) by a deaminated lysine generating the variant H3K4ox. Given that a particular chromatin state can influence the DDR response (Burgess and Misteli, 2015; Goodarzi et al., 2008), recent findings have suggested a connection between LOXL2 and the activation of the DDR pathway in the absence of any detectable DNA lesions in specific breast cancer cell lines. It has been proposed that the lack of LOXL2 and the reduction of H2K4ox levels modify the regulation of chromatin condensation (leading to decondensation) and trigger DDR (Cebria-Costa et al., 2019).

OBJECTIVES

The main goal of this thesis is to determine the implication of UAP56 and Snail1 in genome integrity and their mechanisms of action. For this purpose, we addressed the following specific objectives:

1. To define the role of UAP56 in the maintenance of genome integrity, with the aim of obtaining new insights into the molecular mechanism of this factor to prevent genome instability.
2. To assay the contribution of UAP56 to genome stability all over the genome and evaluate its significance in comparison with the helicase DDX5.
3. To explore the possible implication of Snail1 as a novel factor involved in R loop-mediated genome instability.

MATERIALS AND METHODS

1. GROWTH MEDIA AND CONDITIONS

1.1. Bacteria cell culture

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

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

1.2. Human cell culture

HeLa (ECACC, 93021013) and HEK-293T 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). K562 (ATCC, CCL-243) cells were cultured in Iscove's Modified Dulbecco's medium (IMDM; GIBCO) supplemented with 10% heat-inactivated fetal bovine serum (Sigma Aldrich) and 1% antibiotic-antimycotic (Biowest). Cells were maintained at 37ºC and 5% CO_2.

2. ANTIBIOTICS, DRUGS, INHIBITORS, ENZYMES AND ANTIBODIES

2.1 Antibiotics

- *Ampicillin* (SIGMA): β-lactam antibiotic that inhibits cell wall synthesis in *Echerichia coli*. Used for plasmid selection in *E. coli* (Use: 100 μg/ml).
- *Kanamycin* (SIGMA): aminoglycoside antibiotic that inhibits cell growth by inducing mistranslation and inhibiting 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 inhibits bacterial cell wall synthesis (Use: 60 μg/ml). Streptomycin inhibits prokaryote protein synthesis 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 since it interferes with fungal membrane permeability (Use: 0.25 μg/ml).

2.2 Drugs and inhibitors

- *Complete Protease Inhibitor Cocktail* (Roche): Mixture of several protease inhibitors including serine, cysteine and metalloproteases. It was used according to manufacturer's recommendations.
- *Phenylmethanesulfonyl fluride,* (PMSF) (SIGMA): Inhibitor of serine and cysteine proteases. (Use: 1 mM).
- *Dethyl pirocarbonate* (DEPC) (SIGMA): inhibitor of RNAses.
- *Cordycepin* (SIGMA): Adenosine antagonist 3' deoxyadenosine, inhibitor of chain elongation. (Use: 50 μM).

2.3. Enzymes and antibodies

- *iTaq universal SYBR Green supermix* (Bio-rad): 2x concentrated, ready-to-use reaction master mix optimized for dye-based quantitative PCR (qPCR). It contains antibody-mediated hot-start iTaq DNA polymerase, dNTPs, MgCl2, SYBR Green I dye, enhancers, stabilizers, and a blend of passive reference dyes (including ROX and fluorescein).
- *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 domain to support robust DNA amplification. Used as a part of the Q5 Site-Directed Mutagenesis Kit.
- *Proteinase K* (Roche): very efficient serine protease from Pichia pastoris with no pronounced cleavage specificity.
- *Restriction enzymes* (New England and Takara): DNA endonucleases with specific DNA targets.
- *RNase A* (Roche): endoribonuclease that degrades single-stranded RNA.
- *Zymolyase 20T* (USB): mixture of enzymes from *Arthrobacter luteus* used for digestion of the cell wall of *S. cerevisiae*. (Use: 2 mg/ml).
- *Lysozyme* (SIGMA): enzyme purified from chicken egg white that hydrolyzes peptidoglycans.
- *Spermidine* (SIGMA): polyamine involved in cell metabolism. It binds and precipitates DNA and protein-bound DNA. (Use: 0.5 mM).

- *Dynabeads protein A/G* (Invitrogen): it binds specifically to the Fc portion of IgG. Used for immunoprecipitation experiments (co-IP, 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	Mouse	Synthetic peptide conjugated to KLH derived from within residues 1-100 human beta-actin	ab8227 (Abcam)	WB (1:1000) TBS-T 5% milk
BrdU (clone B44)	Mouse	5-Bromo-2-deoxyuridine	347580(Becton Dickinson)	DNA combing (1/20)
BrdU (clone BU1/75)	Rat	5-Bromo-2-deoxyuridine	ABC1177513 (AbCys)	DNA combing (1/20)
ssDNA (poly dT)	Mouse	Poly dT	DSHB	DNA combing (1/50)
FANCD2	Mouse	Monoclonal antibody raised against the N-terminus of human FANCD2 fusion protein	sc-20022 (Santa Cruz Biotechnology)	IF (1:100)
FLAG	Rabbit	Polyclonal antibody recognizes the FLAG epitope located on FLAG fusion proteins	F7425 (Sigma)	IF (1:1000)
mSin3A	Mouse	Synthetic peptide corresponding to amino acids 1-19 of Mouse mSin3A	sc-5299 (Santa Cruz Biotechnology)	PLA (1:50) WB (1:200) odyssey buffer
Phospho-Histone H2AX.X (Ser 139), Clone JBW301	Mouse	Synthetic peptide Corresponding to amino acids 134-142 of human Histone H2A.X	05-636 (Milipore)	IF (1:1000)
Phospho-Histone H2AX.X (Ser 139), Clone 2F3	Mouse	Synthetic peptide Corresponding to amino acids 134-142 of human Histone H2A.X	613402 (Biolegend)	IF (1:1000)
Phospho-Histone H3 (Ser10) (H3S10p), Mitosis Marker	Rabbit	Linear peptide corresponding to human Histone H3 at Ser10. It recognizes Histone H3 when phosphorylated at Ser10	06-570 (Millipore)	IF (1:200)
S9.6	Mouse	Antibody that detects RNA-DNA hybrids	Hybridoma cell Line HB-8730	DRIP (20 μg) IF (1:500)

Antibody	Source	Epitope	Reference	Use
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)
RNase H1	Rabbit	Amino acids 1-286 RNASEH1 fusion protein	15606-1-AP (Proteintech)	IF (1:1000)
UAP56	Rabbit	Synthetic peptide corresponding to KLH derived from within residues 300-400 of human UAP56	ab47955 (Abcam)	IF (1:200) PLA (1:200) WB (1:1000) BB or odyssey buffer
UAP56	Rabbit	DDX39B fusion protein Ag6512	14798-1-AP (Proteintech)	ChIP (5 μg) IP (5 μg)
Vinculin	Mouse	Mouse monoclonal antibody derived from the hVIN-1 hybridoma	V9264 (Sigma)	WB (1:5000)

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-0.1 % Tween-20; BB: Blocking reagent (Roche).

Table M2. Secondary antibodies used in this study

Specificity	Conjugation	Reference	Use
Mouse	Horseradish peroxidase	SIGMA (A6154)	WB (1:5000)
Rabbit	Horseradish peroxidase	SIGMA (A6154)	WB (1:4000)
Mouse	IRDye 680RD	LI-COR (925-68074)	WB Odyssey (1:5000)/(1:15000)
Rabbit	IRDye 800CW	LI-COR (925-32211)	WB Odyssey (1:5000)/(1:15000)
Goat	IRDye 680RD	LI-COR (925-68070)	WB Odyssey (1:5000)/(1:15000)
Mouse	PLA probe (MINUS oligonucleotide)	Olink Bioscience	PLA (1:5)
Rabbit	PLA probe (PLUS oligonucleotide)	Olink Bioscience	PLA (1:5)
Mouse	Alexa fluor 488	Molecular Probes	IF (1:500)
Mouse	Alexa fluor 546	Molecular Probes	IF (1:500) DNA combing (1/50)
Mouse	Alexa fluor 594	Molecular Probes	IF (1:500)
Mouse	Alexa fluor 647	Molecular Probes	DNA combing (1/50)
Rabbit	Alexa fluor 488	Molecular Probes	IF (1:500)

Specificity	Conjugation	Reference	Use
Rabbit	Alexa fluor 488	Molecular Probes	IF (1:500)
Rat	Alexa fluor 488	Molecular Probes	DNA combing (1/50)

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

3. HUMAN CELL LINES

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

Table M3. Cell lines used in this study

Cell line	Description	Medium	Source
HeLa	Human cervical adenocarcinoma epithelial cells	DMEM	ECACC
HEK293T	Human Embryonic Kidney 293 cells. It contains the SV40 T-antigen	DMEM	PhD. Amelia Nieto CNB-CSIC, Madrid (España)
K562	Human Caucasian chronic myelogenous leukemia lymphoblast cells	IMDM	ATCC

ECACC: European Collection of Authenticated Cell Cultures;

ATCC: American Type Culture Collection.

All experiments were performed using HeLa cells except co-immunoprecipitation experiments, which were performed using HEK293T cells and genome-wide studies, which were performed using K562 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 the P_{CMV} for expression in mammalian cells	Ampicillin	(ten Asbroek et al. 2002)
pCDNA3-RNaseH1	pcDNA3 containing the human RNase H1 gene under P_{CMV}	Ampicillin	ten Asbroek et al. 2002)
pFLAG-CMV-6a	Expression vector derivative of pCMV5 used to establish transient intracellular expression of N-terminal Met-FLAG fusion proteins in mammalian cells	Ampicillin	SIGMA

Plasmid	Description	Resistance	Reference/Source
pFLAG-UAP56	pFLAG containing the human UAP56 gene fussed to FLAG	Ampicillin	PhD. Irene Salas Armenteros
pFLAG-UAP56-K95A	pFLAG containing the K95A mutation in human UAP56 gene fussed to FLAG	Ampicillin	This study
pFLAG-UAP56-E197A	pFLAG containing the E197A mutation in human UAP56 gene fussed to FLAG	Ampicillin	This study
pGEX-KG-UAP56	Expression vector derivative of pGEX-2T used to establish transient intracellular expression of gutathione S-transferase (GST) UAP56 fusion protein in mammalian cells	Ampicillin	This study Patrick Sung lab
pGEX-KG-UAP56-K95A	pGEX-KG vector containing the the K95A mutation in human UAP56 gene fussed to GST	Ampicillin	This study Patrick Sung lab
pGEX-KG-UAP56-K95A	pGEX-KG vector containing the the E197A mutation in human UAP56 gene fussed to GST	Ampicillin	This study Patrick Sung lab

P_{CMV}: Cytomegalovirus promoter.

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 hours after plasmid transfection.

5.2.1. siRNA transfection

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 UAP56 (L-003805-00)	72 h	Dharmacon

siRNA	Time	Source
ON-TARGETplus SMARTpool human THOC1 (L-016376-00)	72 h	Dharmacon
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 AQR (L-022214-01)	72 h	Dharmacon
ON-TARGETplus SMARTpool human FANCD2(L-016376-00)	72 h	Dharmacon
DDX5 siRNA: 5'-GCU CUU UAU AUU GUG UGU UAU dT-3'	72h	Sigma

Cells were transfected with siRNA using DharmaFECT 1 (Dharmacon) at 30-50% confluence. Transfection in a well of 6-well plate was performed using the following protocol:

- *Mixture A* (final volume 100 μl): 95 μl culture serum-free medium (medium without antibiotics or FBS) and 5 μl siRNA 20 μM (100 nM).
- *Mixture B* (final volume 100 μl): 95 μl culture serum-free medium (medium without antibiotics or FBS) and 5 μl of 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 medium. Transfection solution was added carefully drop by drop to the cell culture and incubated for 2 hours. Afterwards, 2 ml 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. 24 hours before transfection cell were cultured in antibiotic-free 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 100 μl): 2 μg DNA in Opti-MEM (Gibco).
- *Mixture B* (final volume 100 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. 24 hours before transfection cell were cultured in antibiotic-free 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 100 μl): 2 μg DNA and 4 μl of Enhancer Reagent in Opti-MEM (Gibco)
- *Mixture B* (final volume 100 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.

6. *IN VITRO* ANALYSIS

These experiments were performed in collaboration with Patrick Sung lab in Yale University.

6.1. Purification of UAP56 wild-type and mutant proteins

The cDNAs that encode the wild type, K95A and E197A variants of UAP56 were introduced into the pGEX-KG vector to add an N-terminal GST tag to these proteins. The resulting UAP56 expression plasmids were introduced into *E. coli* BL21:DE3 Rosetta cells, which were grown at 37°C to OD600 = 0.8, and protein expression was induced by the addition of 0.2 mM IPTG and incubation at 16°C for 16 h. Cells were harvested by centrifugation and all the subsequent steps were carried out at 0-4°C. 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. UAP56 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 GST 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 UAP56, 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 UAP56 (1 mg protein) were pooled, concentrated to 1 mg/ml, and stored in small aliquots at -80°C. The UAP56 K95A and E197A mutants were purified using the same procedure with a similar yield.

K buffer: 20 mM KH_2PO_4, pH 7.4, 10% glycerol, 0.5 mM EDTA, 0.01% Igepal, 1 mM DTT.

6.2. 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 and DNA-DNA duplex with a 5' overhang were prepared by annealing oligonucleotides (with one of the oligonucleotides being labeled with ^{32}P) listed in Table M6. In the unwinding reaction, UAP56 (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). UAP56 (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	RNA	30	5'GGGUGAACCUGCAGGUGGGCAAAGAUGUCC3'

7. PROTEIN-PROTEIN INTERACTION METHODS

7.1. Co-immunoprecipitation (Co-IP)

Whole-cell extracts from a 10 cm petri dish of HEK293T cells at 80% confluence were obtained by lysing cells in 200 µl lysis buffer during 30 min on ice with occasionally gently pipetting up and down. The lysate was centrifuged 10 min at 16,000 g. For each immunoprecipitation, 50 µl Dynabeads Protein A (Invitrogen) were washed twice in 1 ml PBS-0.5%BSA and 5 µg antibody was bound to the beads in 200 µl PBS-0.5% BSA, 4h at 4 ºC. 20 and 180 µl od the lysate were diluted with 180 µl of dilution buffer (to obtain a 0.25% (vol/vol) final concentration of NP-40) and incubated with beads-antibody complexes for 2h at 4 ºC. The same amount of lysate was incubated with beads (without antibody) and was used as a control. Then, beads were washed twice in PBS, four times with wash buffer and bound proteins were eluted by boiling the beads for 10 min 25 µl of 2X Laemmli loading buffer. Finally the result was visualized by Western blot.

Lysis buffer: 10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA pH 8, 0.5% (vol/vol) NP-40, 1 mM PMSF, and protease inhibitor cocktail.
Dilution buffer: 10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA pH 8, 1 mM PMSF, and protease inhibitor cocktail.
Wash buffer: 10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA pH 8, 0.2% (vol/vol) NP-40, 1 mM PMSF, and protease inhibitor cocktail.

7.2. 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 suitable conditions compatible with the two primary antibodies to be used (see Materials and methods 7). Coverslips were blocked with PBS-3% BSA for 1h at RT, incubated with primary antibodies diluted in PBS-3%BSA for 2h at RT (for antibodies dilution see Table M1), washed three times in 5 ml PBS for 5 min and incubated with PLA probes for 1 h at 37 °C. After two washed in 5 ml 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, incubated with the amplification reaction for 100 min at 37 °C in darkness, washed twice in 5 ml buffer B for 10 min and once in 5 ml 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 H2O.
- **Amplification reaction** (40 µl): 8 µl of 5X Buffer amplification, 0.5 µl Polymerase, 31.5 µl MQ H2O.

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

8. CELL CYCLE ANALYSIS IN HUMAN CELLS

8.1. FACS analysis

After siRNA transfection for 72h, HeLa cells were harvested with trypsin, washed twice with cold PBS and centrifuged (3000 rpm 5 min). Pellet was resuspended in 300 µl cold PBS and cells were fixed by adding 700 ul cold 96% ethanol drop by drop and stored at -20 ºC. Before cell cycle analysis and sorting, cells were incubated with Propidium Iodide (50 µg/ml) and treated for RNA degradation with RNase A (250 µg/ml). After that, cells were examined by flow cytometry (FACSCalibur; BD). Three population of cells were analyzed based on their DNA content (1n=haploid or 2n=diploid DNA content of the genome).

9. IMMUNOFLUORESCENCE

Cell were cultured on glass coverslips and, if needed, transfected as indicated in Matherial and Methods 5.2. Cells were fixed in formaldehyde or methanol and selected 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: cells were fixed in 2% formaldehyde in PBS for 20 min or 4% formaldehyde in PBS for 10 min and permeabilized with 70% ethanol for 5 min at -20 ºC, 5 min at 4 ºC and washed twice in PBS. This fixation was used for immunofluorescence with γH2AX, pFLAG and H3Ser10P antibodies.
- Triton-Formaldehyde fixation: Cells were pre-permeabilized with cold 0.1% triton in PBS on ice for 1 min and then fixed in formaldehyde as previously described. This fixation was used for immunofluorescence with UAP56, pFLAG, SIN3
- Methanol fixation: Cells were fixed in cold absolute methanol for 7 min at -20 °C and washed twice in PBS. This fixation was used for immunofluorescence with S9.6 and nucleolin antibodies.

Immunofluorescence: Blocking, incubation with primary and secondary antibodies, DAPI staining and mounting conditions.

Cells were blocking using 3% bovine serum albumin (BSA) in PBS. Then, primary antibodies were diluted in 3% BSA in PBS for 1h at RT. For antibody dilutions see Table M1. Afterwards, cells were washed three times in PBS for 5 min. Secondary antibodies conjugated with Alexa Fluor diluted (1:500) in 3% BSA in PBS were also incubated for 1h at RT. Finally, coverslips were washed twice in PBS before and after the staining of the DNA with 1 µg/ml DAPI (2-(4-Amidinophenyl)-6-indolecarbamidine dihydrochloride) for 5 min. At the end, coverslips were washed in water and a drop of ProLong Gold Antifade reagent (Thermo) was used for mounting. For S9.6 and nucleolin immunofluorescence see Materials and Methods 8.3.

Detection of cells in S phase by immunofluorescence:

When necessary, detection of cells in S phase by immunofluorescence was performed by EdU detection using a Click-iT EdU Imaging Kit (Invitrogen) following manufacturer's instructions.

To perform this assay, cells were culture on glass coverslips and then, the modified thymidine analogue EdU (5-ethynyl-2'-deoxyuridine) is added for 20 min to the cell culture in order to be incorporated into DNA during active synthesis before its fixation. Afterwards, cells were fixed, cells were washed twice with 1 ml of 3% BSA in PBS, permeabilized with 1 ml of 0.5% Triton X-100 in PBS for 20 min at RT and washed again twice in 1 ml of 3% BSA in PBS. Finally cells were incubated with 50 µl/coverslip of Click-iT reaction cocktail (43 µl of 1X Click-iT reaction buffer, 2 µl of 100 mM CuSO4, 0.12 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 in 3% BSA.

10. GENOME INSTABILITY ANALYSIS

10.1. Analysis of γH2AX foci

Cells were culture on glass coverslips and transfected with siRNA as indicated in Materials and Methods 5.2., fixed in 2% formaldehyde and immunofluorescence was performed with mouse monoclonal anti-γH2AX (JBW301, 05-636 Millipore) or mouse monoclonal anti- γH2AX (2F3, 613402 Biolegend) (see Materials and Methods 7 and Table M1). More than 100 cells from each experiment were analyzed (see Materials and Methods 10).

10.2. Single cell gel electrophoresis (Comet assay)

Comet assay was performed using a commercial kit (Trevigen, Gaithersburg, MD,USA) following the manufacturer's protocol, 48 h after siRNA transfection. When it is indicated, 50 µM cordycepin was added to the culture 4 hours before the experiment.

10.2.1. Alkaline comet assay

Cells were collected using accutase, washed and resuspended in ice cold 1X PBS, 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 30 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 dH2O for 5 min each, then in 70% ethanol for 5 min and dried at RT. DNA was stained with SYBR Green at 4 °C for 5 min.

Alkaline unwinding solution/Alkaline electrophoresis solution pH>13: 200 mM NaOH, 1 mM EDTA.

10.2.2. Neutral comet assay

Cells were collected and immobilized on CometSlides as previously described. Cells were lysed for 1h at 4 ºC and immersed in prechilled 1X neutral electrophoresis buffer at 21V for 45 min and then immersed in DNA precipitation solution for 30 min at RT. Finally, slides were immersed in 70% ethanol for 30 min at RT and dried. DNA was stained with SYBER Green at 4 ºC for 30 min.

10X Neutral electrophoresis buffer (500 ml): 60.57 g Tris Base and 204.12 g of sodium acetate dissolved in H2O. Adjust to pH=9.0 with glacial acetic acid. 1X stock was obtained by diluting the 10X stock in dH2O.

DNA precipitation solution: 1 M NH4Ac in 70% ethanol.

For γH2AX foci and comet assays analyses at least three independent experiments were performed. More than 100 cells were scored in each experiment (see Materials and Methods 10).

10.3. RNA-DNA hybrids detection

10.3.1. RNA-DNA 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 (Herrera-Moyano et al, 2014; Garcia-Rubio et al, 2015). After 72 h of siRNA transfection, pellet from a 6-cm plate of HeLa cells was collected using accutase, washed in PBS and resuspended in 800 µl of TE. Afterwards, 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 TE and digested overnight with 50 U of HindIII, EcoRI, BsrGI, XbaI and SspI and 2mM spermidine. For the negative control, half of the DNA was treated with 3 *µ*l RNase H overnight. 5 *µ*g of the digested DNA were bound to 10 *µ*l of S9.6 antibody (1mg/ml) in 500 *µ*l 1X binding buffer in TE, overnight at 4ºC. DNA-antibody complexes were immunoprecipitated using Dynabeads Protein A (Invitrogen) during 2h at 4ºC and washed 3 times with 1X binding buffer. DNA was eluted in 180 µl elution buffer, treated 45 min with 7 *µ*l proteinase K at 55ºC and cleaned with the NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, USA). Quantitative PCR of immunoprecipitated DNA fragments and input DNA was performed on a 7500 Fast & 7500 Real-Time PCR System SYBR qPCR Mix (Applied Biosystems, Thermo Fisher) with the primers listed in Table M7.

10X Binding buffer: 100 mM NaPO4 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 double-distilled H_2O. 4 µl of IP were used for qPCR. The relative abundance of RNA-DNA hybrid immunoprecipitated in each region was normalized to Input signal obtained.

10.3.2. S9.6 immunofluorescence

Cells were cultured on glass coverslips and fixed in methanol (see Materials and Methods 7). For S9.6 and nucleolin immunofluorescence, coverslips were blocked in 2% BSA in PBS overnight at 4°C and incubated with S9.6 mouse (1:500) and anti-nucleolin rabbit (1:1000, Abcam) or pFLAG (1:1000) primary antibodies diluted in 2% BSA. Then, coverslips were washed three times in PBS, and then incubated with secondary antibody Alexa Fluor 594 and 488 diluted in 2% BSA in PBS (1:1000) for 1 hour at RT. Finally, they were washed, DAPI staining and mounting as described above. More than 100 cells from each experiment were scored (see Materials and Methods 10).

11. REPLICATION ANALYSIS

11.1. Analysis of FANCD2 foci

Cells were cultured on glass coverslips, pre-permeabilized with 0.25% Triton X-100 in PBS for 1 minute on ice and then fixed with 2% formaldehyde in PBS. After blocking with 3% BSA in PBS, cells were incubated with primary antibodies FANCD2 (1:100 dilution) and RNase H1 (1:400) diluted in 3% BSA in PBS. Afterwards, coverslips were washed three times in PBS and incubated with secondary antibodies conjugated with Alexa 488 and Alexa 546 (1:1000). Finally, coverslips were washed, DAPI staining and mounting as described above. More than 100 cells from each experiment were scored (see Materials and Methods 10). In pre-permeabilized cells the overexpressed RNase H1 stained only nucleus and nucleoli because the rest of the protein had been washed out.

11.2. Single DNA fiber analysis in human cells (DNA combing)

DNA combing was performed as previously described (Michalet et al., 1997). Cells were transfected with siRNA and with the empty pcDNA3 plasmid or the pcDNA3 RNase H1 plasmid for 48 h, as indicated in Matherial and Methods 5.2. Iodo-deoxyuridine (IdU) and Cloro-deoxyuridine (CldU) labels were added for 20 min each. Subsequently, cells were harvested using accutase, resuspended in cold PBS and embedded in 1% low melting agarose plugs in PBS. Plugs were

incubated in proteinase K buffer at 50 ºC overnight, 6 hours more with a new preparation of the same buffer and washed 5 times with TE buffer for 10 min at RT at 300 rpm. DNA was stained with YOYO-1 (Molecular Probes) to check the integrity of the DNA fibers. Afterwards, each plug was melted in 3 ml of 50 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH 5.7 for 30-45 min at 67ºC. 3U of β-agarase (New England Biolabs) was added after the solution cooled down to 42ºC and it was incubated overnight. Next, DNA fibers were stretched on silanized coverslips by incubation for 15 min at RT and coverslips were removed from the reservoir at the speed of 300 µm/s. DNA was crosslinked to coverslips by baking at 60ºC for 2 hours.

For immunodetection, slides were dehydrated for 3 minutes in successive bath of 70%, 90% and 100% EtOH, incubated for 8 min in 0.5 M NaOH/ 1M NaCl, washed 5 times in PBS and blocked in PBS - 0.1% Triton X-100 - 1% BSA for 15 minutes. DNA molecules were counterstained with 18 µl of an anti-ssDNA antibody (DSHB, 1:500, 30 min) and an anti-mouse IgG coupled to Alexa 647 (Molecular Probes, 1:50, 30 min). CldU and IdU were detected with BU1/75 (AbCys, 1:20, 45 min) and BD44 (Becton Dickinson, 1:20, 45 min anti-BrdU antibodies, respectively. Antibodies were incubated in a humid chamber at 37 ºC and, between incubations, samples were washed 5 times for 2 minutes with PBS-0.1% Triton X-100. Secondary antibodies used were goat anti-mouse IgG Alexa 546 (1:50, 30 min) and chicken anti-rat Alexa 488 (1:50, 30 min). Finally, dried slides were mounted using 20 µl of Prolong Gold Antifade. Representative images of DNA fibers were assembled from different microscopic fields of view and were processed as described.

Proteinase K buffer: 10 mM Tris-HCl pH 7.5 - 50 mM EDTA - 1% Sarkosyl-0.5%.

TE buffer: 10 mM Tris-HCl pH 7.0 – 50 mM EDTA.

12. MICROSCOPY IMAGES ACQUISITION, DATA ANALYSIS AND STADISTICAL ANALYSIS

12.1. Fluorescence microscopy

DNA fibers were analyzed on a Leica DM4000 microscope equipped with a DFC365 FX camera (Leica). Data acquisition was performed with LAS AX (Leica). A 63x objective was used for immunofluorescence for immunofluorescence (γH2AX, FANCD2 and H3S10-P foci IF, PLA, S9.6 and nucleolin IF), a 40x objective was used for DNA combing experiments and a 10x objective was used for comet assays.

12.2. Data analysis

- γH2AX, FANCD2 and H3S10-P foci measurements were analyzed and processed with the MetaMorph v7.5.1.0. software using the granularity application.
- S9.6 signal intensity per nucleus were 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.
- Comet assays tail moments were analyzed using Comet-score (version 1.5) or TriTek CometScore Professional (version 1.0.1.36) softwares. 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).
- Combing measurements were analyzed and processed with the MetaMorph v7.5.1.0 software using measurements applications with the following setup conditions: 50 μM = 310 pixels and 1 μM = 2.2 Kb. Data analysis was performed as previously described (Salas-Armenteros et al., 2017; Tuduri et al., 2009). For fork velocity only CldU tracks that follow an IdU track were considered. Fork asymmetry was expressed as the percentage of distance that is longer the longest track than the shortest, for each pair of sister replication forks during the CldU pulse ((longest-shortest)/longest) x 100). Only sister replication forks with present more than a 25% of difference were considered asymmetric.

Fork velocity (Kb/min) D / t

Fork asymmetry (%) ((L-S)/L) x 100)

Figure M1. DNA combing measurements.
D/t: Distance/time. S: shortest track. L: longest track.

For all experiments, at least three biological repeats (n) were performed. More than 100 cells were scored in each repeat, when possible. Otherwise, a minimum of 50 cells were counted in each repeat.

12.3. Statistical analysis

- γH2AX, FANCD2 and H3S10-P foci: Graphs shows the mean of the percentage of cells with foci from at least three biological repeats. Data were analyzed with EXCEL program. For statistical analysis, Student's t-test was performed and a P value < 0.05 was considered as statistically significant.
- S9.6 signal intensity per nucleus: Graphs shows the median of the measurements from at least three biological repeats. Data were analyzed wit 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).
- Comet assay: Graphs shows the mean of the median of tail moment normalized with to the siC control from at least three biological repeats. Data were analyzed with EXCEL program and GraphPad Prism software. For statistical analysis, Mann-Whitney U-test was performed and P value < 0.05 was considered as statistically significant.
- DRIP: Graphs shows the signal values of RNA-DNA hybrids immunoprecipitated in each region as a function of input DNA 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, Mann-Whitney U-test was performed and P value < 0.05 was considered as statistically significant.

13. POLYMERASE CHAIN REACTION (PCR) ANALYSIS

13.1. 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 H2O, 2 µl primer mixture (each 10 µM), 2 µl DNA and 10 µl SYBR® green qPCR Mix (Bio-rad) were used. The program for PCR reaction used was the following: 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).

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

Relative qPCRs were used to determine the relative mRNA levels in human cells. cDNA was obtain from total RNA extracted using RNeasy Mini Kit (Qiagen) (1 µg) by reverse transcription using QuantiTect Reverse transcription (Qiagen) and random primers. mRNA expression values were normalized to mRNA expression of the Hypozanthine PhosphoRibosylTransferase (HPRT) housekeeping gene.

13.1.2. qPCR analysis for DRIP quantification

Absolute qPCRs were used for DRIP quantification.

Table M7. DNA primers used in this study

Primer	Sequence 5' to 3'	Use
APOE Fwd	CCGGTGAGAAGCGCAGTCGG	DRIP
APOE Rvs	CCCAAGCCCGACCCCGAGTA	DRIP

Primer	Sequence 5' to 3'	Use
RPL13A Fwd	GCTTCCAGCACAGGACAGGTAT	DRIP
RPL13A Rvs	CACCCACTACCCGAGTTCAAG	DRIP
EGR1 Fwd	GCCAAGTCCTCCCTCTCTACTG	DRIP
EGR1 Rvs	GGAAGTGGGCAGAAAGGATTG	DRIP
DDX23 Fwd	AGCCATTATCCCTGGAGGAG	Relative mRNA expression
DDX23 Rvs	CTTCAGCCTCTCGTTCTGCT	Relative mRNA expression
AQR Fwd	TGGGAGAATCTGAACCTAATCC	Relative mRNA expression
AQR Rvs	GCAGGGTAACCAAGTAAACACA	Relative mRNA expression
SETX Fwd	CACACTATGGAGAGGGAAGCA	Relative mRNA expression
SETX Rvs	TTAGATCCAAGGCGATCCAG	Relative mRNA expression
UAP56 Fwd	GACAGCAGCTGGGGGAGATG	Relative mRNA expression
UAP56 Rvs	CTCATGCTGGACTTCTGACG	Relative mRNA expression
DDX5 Fwd	GCAACCATTGACGCCATG	Relative mRNA expression
DDX5 Rvs	CCAAGTCCAAGCCGCAAA	Relative mRNA expression
APOE Fwd	AAGCTGGAGGAGCAGGCC	Relative mRNA expression
APOE Rvs	ACTGGCGCTGCATGTCTTC	Relative mRNA expression
RPL13A Fwd	GGGAGCAAGGAAAGGGTCTTA	Relative mRNA expression
RPL13A Rvs	ACAATTCTCCGAGTGCTTTCAAG	Relative mRNA expression

14. PROTEIN ANALYSIS

14.1. Human cells protein extraction

Pellet of HeLa, HEK293T or K562 cells was collected using trypsin or accutase, washed twice in cold PBS and gently resuspended in 2X Laemmli loading buffer (100 µl/1x10^6 cells). The lysate was boiled at 95 ºC for 5 min and sonicated for another 5 min on the maximum intensity setting, with fifteen pulses of 30s on and

30s off in Bioruptor (Diagenode). Prior to gel loading, samples we boiled for 2 min.

2X Lammeli buffer: 200 Mm Tris-HCl, 40% glycerol, 8% SDS, 0.4% Bromophenol Blue, 400 mM β-mercaptoethanol.

14.2. SDS-PAGE

Proteins were separated in 29:1 acrylamide:bis-acrylamide SDS-PAGE with appropriate concentrations to the molecular size of the proteins of interest or in 4-20% gradient SDS-PAGE CriterionTM TGXTM Precast Gels (BioRad) following the method described in (Laemmli, 1970). Electrophoreses were performed in a Mini-PROTEAN 3 Cell (BioRad) using running buffer at 100 V. Page RulerTM (Fermentas, CA) was used as a protein marker.

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

14.3. Western Blot analysis

For Western blot, proteins were wet-transferred using Trans-Blot system (Biorad) for 2 h at 400 mA in 1X Transfer Buffer with 20% methanol or o/n at 30 V in 1X Tris-glycine Buffer at 4 °C. After transfer, the membranes were strained with Ponceau S (0.1% w/v Ponceau (Sigma) in 5% Acetic acid) to check protein loading and correct transfer.

5X Transfer buffer: 6 g/L Tris base, 28.8 g/L glycine and 0.5% SDS.
10X Tris-glycine buffer: 30 g/L Tris base, 143.2 g/L glycine pH 8.3.

14.4. Non-fluorescence WB

Proteins were transferred to a nitrocellulose membrane (Hybond-ECL, GE Healthcare). Membranes were blocked with 1X TBS - 0.1% Tween 20 - 5% milk or Blocking Buffer solution (ROCHE) for 1 h. Primary antibodies were incubated for 2h at RT or o/n at 4 °C in 1X TBS - 0.1% Tween 20 - 5% milk, blocking buffer solution or BSA. After 3 washes of 10 min each one with 1X TBS- 0.1% Tween 20, membranes were were incubated with the corresponding secondary

antibodies conjugated with the horseradish peroxidase for 1h hour and washed again. Finally, SuperSignalR West Pico (Pierce) was used for chemiluminescence detection.

Blocking Buffer solution: 1% Blocking reagent (Roche), 0.05% Tween 20, 0.05 M Tris-HCl pH 7.5. A stock of 10% Blocking reagent was previously prepared dissolving 10 g of blocking reagent in 100 ml maleic buffer (0.1 M Maleic acid, 0.15 M NaCl, pH 7.5 M adjusted with NaOH and autoclaved) with heat (50-60 °C) and shake.
Wash solution: 1X TBS – 0.1% Tween 20.

14.5. Fluorescent WB

A PVDF membrane with low fluorescence background (Inmobilon-FL, Millipore) was used. This membrane was first activated in methanol for 15 s and equilibrated in transfer buffer before continuing. Commercial Odyssey Blocking Buffer (LI-COR Biosciences) was used to block the membrane for 1 hat RT or o/n at 4 °C. Primary antibody was prepared (for appropriate dilution see Table M1) in blocking buffer -0.1% Tween 20 and incubated for 2 h. Three washes of 10 min were performed with 1X TBS - 0.1% Tween 20 followed by incubation of 1 h with IRDye secondary antibodies. Finally, membranes where washed again 3 times, rinsed in 1X TBS and immediately scanned or left drying. Image acquisition was performed in an Odyssey CLx Imager (LI-COR Biosciences).

15. GENOME WIDE EXPERIMENTS

15.1. RNA-seq

Total RNA was isolated from K562 cells transfected with siC or siUAP56 siRNA for 72 h with an RNeasy Mini kit (Qiagen). Then, total RNA-seq was performed after ribosomal RNA depletion applying the TruSeq Stranded Total RNA library and sequenced on the platform NextSEq500 (Illumina).

15.2. Chromatin immunoprecipitation (ChIP) assay (ChIP-seq)

K562 cells were crosslinked for 10 min with formaldehyde at a final concentration of 1%, resuspended in 2.5 ml of cell lysis buffer, then centrifuged and 1 ml of nuclei lysis buffer was added. Chromatin was sonicated on the maximum intensity setting, with fifteen pulses of 30s on and 30s off in Bioruptor (Diagenode), to obtain approx. 400 bp fragments. Samples were diluted up to 1300 µl with IP buffer. 100 µl and 1200 µl of diluted chromatin were used for input and immunoprecipitation, respectively. 30 µl of Dynabeads Protein A (Invitrogen) per sample was incubated overnight at 4 ºC with 4 µg of UAP56 antibody. A negative control the corresponding IgG was used to calculate the background signal. Then, Dynabeads-antibody complexes were added to the samples used for immunoprecipitation for 2h 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 were then un-crosslinked in TE -1% SDS and treated with proteinase K. DNA was isolated using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel) and used to build the libraries using the ThruPLEX DNA-Seq 6S kit (Rubicon Genomics) according to manufacturer's instructions and then sequenced on the Illumina platform NextSEq500.

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

Nuclei lysis buffer: 1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8, 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, 167 mM NaCl.

Wash buffer 1: 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 150 mM NaCl.

Wash buffer 2: 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8, 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.

15.3. DNA-RNA immunoprecipitation followed by high-throughput DNA sequencing (DRIP-seq)

DRIP-seq was performed as previously described (Sanz & Chedin, 2019) with minor modifications of the DRIP-qPCR (see Materials and Methods 8.3). Basically, genomic DNA was digested overnight with 30 U of HindIII, EcoRI, BsrGI, XbaI and SspI and 2mM spermidine. For the negative control, 10 μg of the digested DNA were treated with 4 μl RNase H (New England Biolabs) for 6 h at 37 ºC. Five immunoprecipitation were performed, in each one 8 μg of the digested DNA were bound to 20 μl of S9.6 antibody (1mg/ml) in 500 μl 1X binding buffer in TE, overnight at 4ºC. DNA-antibody complexes were immunoprecipitated using Dynabeads Protein A (Invitrogen) during 2h at 4ºC and washed 3 times with 1X binding buffer. DNA was eluted in 300 µl elution buffer, treated 45 min with 7 μl proteinase K at 55ºC and phenol-chloroform purified. Finally, DNA was resuspended in 10 μl of RNase-free water, with a total volume of 50 μl. Finally, the DNA was sonicated and checked on a 2100 Agilent Bioanalyzer. Afterwards, this DNA was used to build the libraries using the ThruPLEX DNA-Seq 6S kit (Rubicon Genomics) according to manufacturer's instruction and sequenced on an Illumina NextSeq500 platform.

15.4. RNA-DNA immunoprecipitation followed by cDNA conversion couple to high throughput sequencing (DRIPc-seq)

DRIPc-seq was performed essentially as described (Sanz & Chedin, 2019) and described in Materials and Methods 11.3. Briefly, after DRIP, the eluted DNA from five immunoprecipitations of each sample was treated with 6 U of DNase I (New England BioLabs) for 45 min at 37°C to degrade all DNA. The resulting RNA was subjected to libraries construction using the TruSeq Stranded Total RNA protocol (Illumina) from the fragmentation step. The quality of the libraries was checked on a 2100 Agilent Bioanalyzer prior to sequencing on an Illumina NextSeq500 platform.

16. GENOME WIDE DATA ANALYSIS

16.1. RNA-seq

Sequenced paired-ends reads were subjected to quality control pipeline using the FASTQ Toolkit v.1.0.0 software (Illumina) and then mapped to the human reference genome hg38 canonical using HISAT2 (Kim et al., 2019). Reads mapping to mitochondrial chromosome and duplicated reads were discarded using SAMtools (Li et al., 2009) for our purposes. After the obtainment of BAM files, counts per peak were established using FeatureCounts and RPKM normalized. For genes in which expression changes are detected (Ilinear fold changel > 1.5), siC Log_2expression and Log_2(siUAP56/siC) was compared and represented for genes longer than 5 Kb.

16.2. ChIP-seq

Sequenced paired-ends reads were subjected to quality control pipeline using the FASTQ Toolkit v.1.0.0 software (Illumina) and then mapped to the human reference genome hg38 canonical using using BWA-MEM (Li et al., 2009). Reads mapping to mitochondrial chromosome and duplicated reads were discarded using SAMtools (Li et al., 2009) for our purposes. BAM files were further analyzed with deepTools Bamcompare from deepTool2 package (Ramirez et al., 2016) for track generation. The results were visualized with the Integrative Genomics Viewer (IGV), developed and maintained by The Broad Institute and the Regents of the University of California and UC San Diego.

Peak calling was performed using MACS2 package (Zhang et al., 2008) selecting those whose enrichment signal over the input had a p-value < 0.01. As well, selected peaks were annotated to genes using ChIPseeker (Yu et al., 2015) and genes retrieved from Ensembl release 94 2018 (Zerbino et al., 2018). For our purposes, only genes with more than 5 Kb were analyzed, considering promoter as -2/+5 Kb from TSS and downstream as -2/+5 Kb from TES.

16.3. DRIP-seq and DRIPc-seq

Sequenced paired-ends reads were subjected to quality control pipeline using the FASTQ Toolkit v.1.0.0 software (Illumina) and then mapped to the human reference genome hg38 canonical using using BWA-MEM (Li et al., 2009). Reads mapping to mitochondrial chromosome and duplicated reads were discarded using SAMtools (Li et al., 2009) for our purposes. Additionally, For DRIPc-seq, reads were separated into minus and plus strand using the same tool. BAM files were further analyzed with deepTool2 package (Ramirez et al., 2016) for track generation setting a window bien size of 10 nt and a smooth value of 40. The results were visualized with the Integrative Genomics Viewer (IGV), developed and maintained by The Broad Institute and the Regents of the University of California and UC San Diego.

Peak calling on DRIP-seq and DRIPc-seq was performed using MACS2 package (Zhang et al., 2008). Then, number of counts per peak was calculated using FeatureCounts and RPKM normalized. For analysis, significant DRIPc peaks (q-value < 0.01) in siUAP56 depleted cells were established selecting those peaks whose DRIPc signal fold change was higher than 1.5X respect to the siC control cells. Afterwards, peaks were annotated to genes using ChIPseeker (Yu et al., 2015) and genes retrieved from Ensembl release 94 2018 (Zerbino et al., 2018). For our purposes, only genes with more than 5 Kb were analyzed, considering promoter as -2/+5 Kb from TSS and downstream as -2/+5 Kb from TES.

Gene metaplots were obtained using deepTools computematrix and plotheatmap. Venn diagrams were created using Biovenn (Hulsen et al., 2008).

- **Statistical tests**

  Statistical tests (Mann-Whitney U-test and Pearson's correlation were calculated using GraphPad Prism software. Hypergeometric tests were calculated using R scripts. In general, a P-value < 0.05 was considered as statistically significant.

REFERENCES

Aguilera, A. (2002). The connection between transcription and genomic instability. EMBO J *21*, 195-201.

Aguilera, A. (2005a). Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. Curr Opin Cell Biol *17*, 242-250.

Aguilera, A. (2005b). mRNA processing and genomic instability. Nat Struct Mol Biol *12*, 737-738.

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

Aguilera, A., and Garcia-Muse, T. (2013). Causes of genome instability. Annu Rev Genet *47*, 1-32.

Aguilera, A., and Gomez-Gonzalez, B. (2008). Genome instability: a mechanistic view of its causes and consequences. Nat Rev Genet *9*, 204-217.

Aguilera, A., and Klein, H.L. (1990). HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the Saccharomyces cerevisiae TOP1 gene. Mol Cell Biol *10*, 1439-1451.

Alexander, R.D., Innocente, S.A., Barrass, J.D., and Beggs, J.D. (2010). Splicing-dependent RNA polymerase pausing in yeast. Mol Cell *40*, 582-593.

Alzu, A., Bermejo, R., Begnis, M., Lucca, C., Piccini, D., Carotenuto, W., Saponaro, M., Brambati, A., Cocito, A., Foiani, M.*, et al.* (2012). Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. Cell *151*, 835-846.

Aroeira, L.S., Aguilera, A., Sanchez-Tomero, J.A., Bajo, M.A., del Peso, G., Jimenez-Heffernan, J.A., Selgas, R., and Lopez-Cabrera, M. (2007). Epithelial to mesenchymal transition and peritoneal membrane failure in peritoneal dialysis patients: pathologic significance and potential therapeutic interventions. J Am Soc Nephrol *18*, 2004-2013.

Ayyanathan, K., Peng, H., Hou, Z., Fredericks, W.J., Goyal, R.K., Langer, E.M., Longmore, G.D., and Rauscher, F.J., 3rd (2007). The Ajuba LIM domain protein is a corepressor for SNAG domain mediated repression and participates in nucleocytoplasmic Shuttling. Cancer Res *67*, 9097-9106.

Barrallo-Gimeno, A., and Nieto, M.A. (2005). The Snail genes as inducers of cell movement and survival: implications in development and cancer. Development *132*, 3151-3161.

Bates, G.J., Nicol, S.M., Wilson, B.J., Jacobs, A.M., Bourdon, J.C., Wardrop, J., Gregory, D.J., Lane, D.P., Perkins, N.D., and Fuller-Pace, F.V. (2005). The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. EMBO J *24*, 543-553.

Batlle, E., Sancho, E., Franci, C., Dominguez, D., Monfar, M., Baulida, J., and Garcia De Herreros, A. (2000). The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nat Cell Biol *2*, 84-89.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat Rev Genet *15*, 163-175.

Bermejo, R., Capra, T., Gonzalez-Huici, V., Fachinetti, D., Cocito, A., Natoli, G., Katou, Y., Mori, H., Kurokawa, K., Shirahige, K.*, et al.* (2009). Genome-organizing factors Top2 and Hmo1 prevent chromosome fragility at sites of S phase transcription. Cell *138*, 870-884.

Bermejo, R., Lai, M.S., and Foiani, M. (2012). Preventing replication stress to maintain genome stability: resolving conflicts between replication and transcription. Mol Cell *45*, 710-718.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. Genes Dev *19*, 1635-1655.

Beyes, S., Andrieux, G., Schrempp, M., Aicher, D., Wenzel, J., Anton-Garcia, P., Boerries, M., and Hecht, A. (2019). Genome-wide mapping of DNA-binding sites identifies stemness-related genes as directly repressed targets of SNAIL1 in colorectal cancer cells. Oncogene.

Bhatia, V., Barroso, S.I., Garcia-Rubio, M.L., Tumini, E., Herrera-Moyano, E., and Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. Nature *511*, 362-365.

Bonnet, A., Grosso, A.R., Elkaoutari, A., Coleno, E., Presle, A., Sridhara, S.C., Janbon, G., Geli, V., de Almeida, S.F., and Palancade, B. (2017). Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability. Mol Cell *67*, 608-621 e606.

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., and Guil, S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. Proc Natl Acad Sci U S A *112*, 5785-5790.

Brewer, B.J. (1988). When polymerases collide: replication and the transcriptional organization of the E. coli chromosome. Cell *53*, 679-686.

Burgess, R.C., and Misteli, T. (2015). Not All DDRs Are Created Equal: Non-Canonical DNA Damage Responses. Cell *162*, 944-947.

Cai, W., Xiong Chen, Z., Rane, G., Satendra Singh, S., Choo, Z., Wang, C., Yuan, Y., Zea Tan, T., Arfuso, F., Yap, C.T.*, et al.* (2017). Wanted DEAD/H or Alive: Helicases Winding Up in Cancers. J Natl Cancer Inst *109*.

Caretti, G., Schiltz, R.L., Dilworth, F.J., Di Padova, M., Zhao, P., Ogryzko, V., Fuller-Pace, F.V., Hoffman, E.P., Tapscott, S.J., and Sartorelli, V. (2006). The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. Dev Cell *11*, 547-560.

Carter, C.L., Lin, C., Liu, C.Y., Yang, L., and Liu, Z.R. (2010). Phosphorylated p68 RNA helicase activates Snail1 transcription by promoting HDAC1 dissociation from the Snail1 promoter. Oncogene *29*, 5427-5436.

Castellano-Pozo, M., Garcia-Muse, T., and Aguilera, A. (2012a). The Caenorhabditis elegans THO complex is required for the mitotic cell cycle and development. PLoS One *7*, e52447.

Castellano-Pozo, M., Garcia-Muse, T., and Aguilera, A. (2012b). R-loops cause replication impairment and genome instability during meiosis. EMBO Rep *13*, 923-929.

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

Cebria-Costa, J.P., Pascual-Reguant, L., Gonzalez-Perez, A., Serra-Bardenys, G., Querol, J., Cosin, M., Verde, G., Cigliano, R.A., Sanseverino, W., Segura-Bayona, S.*, et al.* (2019). LOXL2-mediated H3K4 oxidation reduces chromatin accessibility in triple-negative breast cancer cells. Oncogene.

Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in eukaryotes. FEBS J *276*, 1494-1505.

Chan, Y.A., Aristizabal, M.J., Lu, P.Y., Luo, Z., Hamza, A., Kobor, M.S., Stirling, P.C., and Hieter, P. (2014). Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. PLoS Genet *10*, e1004288.

Chang, C.T., Hautbergue, G.M., Walsh, M.J., Viphakone, N., van Dijk, T.B., Philipsen, S., and Wilson, S.A. (2013). Chtop is a component of the dynamic TREX mRNA export complex. EMBO J *32*, 473-486.

Chaudhuri, J., and Alt, F.W. (2004). Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol *4*, 541-552.

Chavez, S., and Aguilera, A. (1997). The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. Genes Dev *11*, 3459-3470.

Chavez, S., Beilharz, T., Rondon, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera, A. (2000). A protein complex

containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. EMBO J *19*, 5824-5834.

Chedin, F. (2016). Nascent Connections: R-Loops and Chromatin Patterning. Trends Genet *32*, 828-838.

Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H.*, et al.* (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol Cell *68*, 745-757 e745.

Chen, P.B., Chen, H.V., Acharya, D., Rando, O.J., and Fazzio, T.G. (2015). R loops regulate promoter-proximal chromatin architecture and cellular differentiation. Nat Struct Mol Biol *22*, 999-1007.

Cheng, H., Dufu, K., Lee, C.S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. Cell *127*, 1389-1400.

Chon, H., Sparks, J.L., Rychlik, M., Nowotny, M., Burgers, P.M., Crouch, R.J., and Cerritelli, S.M. (2013). RNase H2 roles in genome integrity revealed by unlinking its activities. Nucleic Acids Res *41*, 3130-3143.

Cordin, O., Banroques, J., Tanner, N.K., and Linder, P. (2006). The DEAD-box protein family of RNA helicases. Gene *367*, 17-37.

Cristini, A., Groh, M., Kristiansen, M.S., and Gromak, N. (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. Cell Rep *23*, 1891-1905.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019). R-Loops as Cellular Regulators and Genomic Threats. Mol Cell *73*, 398-411.

Dai, T.Y., Cao, L., Yang, Z.C., Li, Y.S., Tan, L., Ran, X.Z., and Shi, C.M. (2014). P68 RNA helicase as a molecular target for cancer therapy. J Exp Clin Cancer Res *33*, 64.

Dardenne, E., Polay Espinoza, M., Fattet, L., Germann, S., Lambert, M.P., Neil, H., Zonta, E., Mortada, H., Gratadou, L., Deygas, M.*, et al.* (2014). RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. Cell Rep *7*, 1900-1913.

Deshpande, A.M., and Newlon, C.S. (1996). DNA replication fork pause sites dependent on transcription. Science *272*, 1030-1033.

Dominguez-Sanchez, M.S., Barroso, S., Gomez-Gonzalez, B., Luna, R., and Aguilera, A. (2011). Genome instability and transcription elongation

impairment in human cells depleted of THO/TREX. PLoS Genet *7*, e1002386.

Drolet, M., Phoenix, P., Menzel, R., Masse, E., Liu, L.F., and Crouch, R.J. (1995). Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. Proc Natl Acad Sci U S A *92*, 3526-3530.

Dufu, K., Livingstone, M.J., Seebacher, J., Gygi, S.P., Wilson, S.A., and Reed, R. (2010). ATP is required for interactions between UAP56 and two conserved mRNA export proteins, Aly and CIP29, to assemble the TREX complex. Genes Dev *24*, 2043-2053.

Dumelie, J.G., and Jaffrey, S.R. (2017). Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. Elife *6*.

Eder, P.S., Walder, R.Y., and Walder, J.A. (1993). Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. Biochimie *75*, 123-126.

El Achkar, E., Gerbault-Seureau, M., Muleris, M., Dutrillaux, B., and Debatisse, M. (2005). Premature condensation induces breaks at the interface of early and late replicating chromosome bands bearing common fragile sites. Proc Natl Acad Sci U S A *102*, 18069-18074.

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

Fleckner, J., Zhang, M., Valcarcel, J., and Green, M.R. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. Genes Dev *11*, 1864-1872.

Folco, E.G., Lee, C.S., Dufu, K., Yamazaki, T., and Reed, R. (2012). The proteins PDIP3 and ZC11A associate with the human TREX complex in an ATP-dependent manner and function in mRNA export. PLoS One *7*, e43804.

Ford, M.J., Anton, I.A., and Lane, D.P. (1988). Nuclear protein with sequence homology to translation initiation factor eIF-4A. Nature *332*, 736-738.

Fuller-Pace, F.V. (2013). DEAD box RNA helicase functions in cancer. RNA Biol *10*, 121-132.

Gaillard, H., and Aguilera, A. (2013). Transcription coupled repair at the interface between transcription elongation and mRNP biogenesis. Biochim Biophys Acta *1829*, 141-150.

Gaillard, H., and Aguilera, A. (2016). Transcription as a Threat to Genome Integrity. Annu Rev Biochem *85*, 291-317.

Gaillard, H., Garcia-Muse, T., and Aguilera, A. (2015). Replication stress and cancer. Nat Rev Cancer *15*, 276-289.

Gaillard, H., Herrera-Moyano, E., and Aguilera, A. (2013). Transcription-associated genome instability. Chem Rev *113*, 8638-8661.

Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J.L., and Li, X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. Genes Dev *25*, 2041-2056.

Garcia-Muse, T., and Aguilera, A. (2016). Transcription-replication conflicts: how they occur and how they are resolved. Nat Rev Mol Cell Biol *17*, 553-563.

Garcia-Pichardo, D., Canas, J.C., Garcia-Rubio, M.L., Gomez-Gonzalez, B., Rondon, A.G., and Aguilera, A. (2017). Histone Mutants Separate R Loop Formation from Genome Instability Induction. Mol Cell *66*, 597-609 e595.

Garcia-Rubio, M., Aguilera, P., Lafuente-Barquero, J., Ruiz, J.F., Simon, M.N., Geli, V., Rondon, A.G., and Aguilera, A. (2018). Yra1-bound RNA-DNA hybrids cause orientation-independent transcription-replication collisions and telomere instability. Genes Dev *32*, 965-977.

Garcia-Rubio, M., Huertas, P., Gonzalez-Barrera, S., and Aguilera, A. (2003). Recombinogenic effects of DNA-damaging agents are synergistically increased by transcription in Saccharomyces cerevisiae. New insights into transcription-associated recombination. Genetics *165*, 457-466.

Garcia-Rubio, M.L., Perez-Calero, C., Barroso, S.I., Tumini, E., Herrera-Moyano, E., Rosado, I.V., and Aguilera, A. (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. PLoS Genet *11*, e1005674.

Gatfield, D., Le Hir, H., Schmitt, C., Braun, I.C., Kocher, T., Wilm, M., and Izaurralde, E. (2001). The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. Curr Biol *11*, 1716-1721.

Gavalda, S., Gallardo, M., Luna, R., and Aguilera, A. (2013). R-loop mediated transcription-associated recombination in trf4Delta mutants reveals new links between RNA surveillance and genome integrity. PLoS One *8*, e65541.

Geissler, V., Altmeyer, S., Stein, B., Uhlmann-Schiffler, H., and Stahl, H. (2013). The RNA helicase Ddx5/p68 binds to hUpf3 and enhances NMD of Ddx17/p72 and Smg5 mRNA. Nucleic Acids Res *41*, 7875-7888.

Gersappe, A., Burger, L., and Pintel, D.J. (1999). A premature termination codon in either exon of minute virus of mice P4 promoter-generated pre-mRNA

can inhibit nuclear splicing of the intervening intron in an open reading frame-dependent manner. J Biol Chem *274*, 22452-22458.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. Mol Cell *45*, 814-825.

Gomez-Gonzalez, B., and Aguilera, A. (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. Proc Natl Acad Sci U S A *104*, 8409-8414.

Gomez-Gonzalez, B., and Aguilera, A. (2019). Transcription-mediated replication hindrance: a major driver of genome instability. Genes Dev *33*, 1008-1026.

Gomez-Gonzalez, B., Garcia-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marin, A., Foiani, M., and Aguilera, A. (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. EMBO J *30*, 3106-3119.

Gonatopoulos-Pournatzis, T., and Cowling, V.H. (2014). Cap-binding complex (CBC). Biochem J *457*, 231-242.

Goodarzi, A.A., Noon, A.T., Deckbar, D., Ziv, Y., Shiloh, Y., Lobrich, M., and Jeggo, P.A. (2008). ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. Mol Cell *31*, 167-177.

Gottipati, P., Cassel, T.N., Savolainen, L., and Helleday, T. (2008). Transcription-associated recombination is dependent on replication in Mammalian cells. Mol Cell Biol *28*, 154-164.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235-240.

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.*, et al.* (2018). Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. Mol Cell *69*, 426-437 e427.

Guarino, M., Rubino, B., and Ballabio, G. (2007). The role of epithelial-mesenchymal transition in cancer pathology. Pathology *39*, 305-318.

Ha, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol *15*, 509-524.

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell *170*, 774-786 e719.

Hartono, S.R., Malapert, A., Legros, P., Bernard, P., Chedin, F., and Vanoosthuyse, V. (2018). The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. J Mol Biol *430*, 272-284.

Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., McKinney, K.M., Eaton, M.L.*, et al.* (2015). BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. Mol Cell *57*, 636-647.

Hautbergue, G.M., Hung, M.L., Walsh, M.J., Snijders, A.P., Chang, C.T., Jones, R., Ponting, C.P., Dickman, M.J., and Wilson, S.A. (2009). UIF, a New mRNA export adaptor that works together with REF/ALY, requires FACT for recruitment to mRNA. Curr Biol *19*, 1918-1924.

Heath, C.G., Viphakone, N., and Wilson, S.A. (2016). The role of TREX in gene expression and disease. Biochem J *473*, 2911-2935.

Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. Nat Struct Mol Biol *20*, 412-418.

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

Herold, A., Teixeira, L., and Izaurralde, E. (2003). Genome-wide analysis of nuclear mRNA export pathways in Drosophila. EMBO J *22*, 2472-2483.

Herranz, N., Pasini, D., Diaz, V.M., Franci, C., Gutierrez, A., Dave, N., Escriva, M., Hernandez-Munoz, I., Di Croce, L., Helin, K.*, et al.* (2008). Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. Mol Cell Biol *28*, 4772-4781.

Herrera-Moyano, E., Mergui, X., Garcia-Rubio, M.L., Barroso, S., and Aguilera, A. (2014). The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. Genes Dev *28*, 735-748.

Hodroj, D., Recolin, B., Serhal, K., Martinez, S., Tsanov, N., Abou Merhi, R., and Maiorano, D. (2017a). An ATR-dependent function for the Ddx19 RNA helicase in nuclear R-loop metabolism. EMBO J *36*, 1182-1198.

Hodroj, D., Serhal, K., and Maiorano, D. (2017b). Ddx19 links mRNA nuclear export with progression of transcription and replication and suppresses genomic instability upon DNA damage in proliferating cells. Nucleus *8*, 489-495.

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

Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics *9*, 488.

Iacovoni, J.S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D., and Legube, G. (2010). High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. EMBO J *29*, 1446-1457.

Iggo, R.D., Jamieson, D.J., MacNeill, S.A., Southgate, J., McPheat, J., and Lane, D.P. (1991). p68 RNA helicase: identification of a nucleolar form and cloning of related genes containing a conserved intron in yeasts. Mol Cell Biol *11*, 1326-1333.

Izquierdo, J.M., and Valcarcel, J. (2006). A simple principle to explain the evolution of pre-mRNA splicing. Genes Dev *20*, 1679-1684.

Jarmoskaite, I., and Russell, R. (2014). RNA helicase proteins as chaperones and remodelers. Annu Rev Biochem *83*, 697-725.

Jensen, T.H., Boulay, J., Rosbash, M., and Libri, D. (2001). The DECD box putative ATPase Sub2p is an early mRNA export factor. Curr Biol *11*, 1711-1715.

Jimeno, S., Rondon, A.G., Luna, R., and Aguilera, A. (2002). The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. EMBO J *21*, 3526-3535.

Kabeche, L., Nguyen, H.D., Buisson, R., and Zou, L. (2018). A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. Science *359*, 108-114.

Kapadia, F., Pryor, A., Chang, T.H., and Johnson, L.F. (2006). Nuclear localization of poly(A)+ mRNA following siRNA reduction of expression of the mammalian RNA helicases UAP56 and URH49. Gene *384*, 37-44.

Katahira, J., Ishikawa, H., Tsujimura, K., Kurono, S., and Hieda, M. (2019). Human THO coordinates transcription termination and subsequent transcript release from the HSP70 locus. Genes Cells *24*, 272-283.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol *37*, 907-915.

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

Kistler, A.L., and Guthrie, C. (2001). Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. Genes Dev *15*, 42-49.

Kogoma, T. (1997). Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. Microbiol Mol Biol Rev *61*, 212-238.

Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nat Rev Mol Cell Biol *8*, 761-773.

Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J., and Nogues, G. (2004). Multiple links between transcription and splicing. RNA *10*, 1489-1498.

Kornblihtt, A.R., Schor, I.E., Allo, M., Dujardin, G., Petrillo, E., and Munoz, M.J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat Rev Mol Cell Biol *14*, 153-165.

Kota, K.P., Wagner, S.R., Huerta, E., Underwood, J.M., and Nickerson, J.A. (2008). Binding of ATP to UAP56 is necessary for mRNA export. J Cell Sci *121*, 1526-1537.

Kothari, A.N., Mi, Z., Zapf, M., and Kuo, P.C. (2014). Novel clinical therapeutics targeting the epithelial to mesenchymal transition. Clin Transl Med *3*, 35.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Lamm, G.M., Nicol, S.M., Fuller-Pace, F.V., and Lamond, A.I. (1996). p72: a human nuclear DEAD box protein highly related to p68. Nucleic Acids Res *24*, 3739-3747.

Lane, D.P., and Hoeffler, W.K. (1980). SV40 large T shares an antigenic determinant with a cellular protein of molecular weight 68,000. Nature *288*, 167-170.

Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E., and Merrikh, H. (2017). Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell *170*, 787-799 e718.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, L., Germain, D.R., Poon, H.Y., Hildebrandt, M.R., Monckton, E.A., McDonald, D., Hendzel, M.J., and Godbout, R. (2016). DEAD Box 1 Facilitates

Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks. Mol Cell Biol *36*, 2794-2810.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell *122*, 365-378.

Libri, D., Dower, K., Boulay, J., Thomsen, R., Rosbash, M., and Jensen, T.H. (2002). Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. Mol Cell Biol *22*, 8254-8266.

Libri, D., Graziani, N., Saguez, C., and Boulay, J. (2001). Multiple roles for the yeast SUB2/yUAP56 gene in splicing. Genes Dev *15*, 36-41.

Linder, P., and Jankowsky, E. (2011). From unwinding to clamping - the DEAD box RNA helicase family. Nat Rev Mol Cell Biol *12*, 505-516.

Linskens, M.H., and Huberman, J.A. (1988). Organization of replication of ribosomal DNA in Saccharomyces cerevisiae. Mol Cell Biol *8*, 4927-4935.

Liu, B., and Alberts, B.M. (1995). Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. Science *267*, 1131-1137.

Luna, R., Jimeno, S., Marin, M., Huertas, P., Garcia-Rubio, M., and Aguilera, A. (2005). Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. Mol Cell *18*, 711-722.

Luna, R., Rondon, A.G., and Aguilera, A. (2012). New clues to understand the role of THO and other functionally related factors in mRNP biogenesis. Biochim Biophys Acta *1819*, 514-520.

Luo, M.L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., and Reed, R. (2001). Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. Nature *413*, 644-647.

Lykke-Andersen, S., and Jensen, T.H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Biol *16*, 665-677.

Ma, W.K., Cloutier, S.C., and Tran, E.J. (2013). The DEAD-box protein Dbp2 functions with the RNA-binding protein Yra1 to promote mRNP assembly. J Mol Biol *425*, 3824-3838.

Ma, W.K., Paudel, B.P., Xing, Z., Sabath, I.G., Rueda, D., and Tran, E.J. (2016). Recruitment, Duplex Unwinding and Protein-Mediated Inhibition of the Dead-Box RNA Helicase Dbp2 at Actively Transcribed Chromatin. J Mol Biol *428*, 1091-1106.

MacMorris, M., Brocker, C., and Blumenthal, T. (2003). UAP56 levels affect viability and mRNA export in Caenorhabditis elegans. RNA *9*, 847-857.

Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M.*, et al.* (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704-715.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499-506.

Manzo, S.G., Hartono, S.R., Sanz, L.A., Marinello, J., De Biasi, S., Cossarizza, A., Capranico, G., and Chedin, F. (2018). DNA Topoisomerase I differentially modulates R-loops across the human genome. Genome Biol *19*, 100.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes Dev *19*, 1512-1517.

Maturi, V., Moren, A., Enroth, S., Heldin, C.H., and Moustakas, A. (2018). Genomewide binding of transcription factor Snail1 in triple-negative breast cancer cells. Mol Oncol *12*, 1153-1174.

Mazurek, A., Park, Y., Miething, C., Wilkinson, J.E., Gillis, J., Lowe, S.W., Vakoc, C.R., and Stillman, B. (2014). Acquired dependence of acute myeloid leukemia on the DEAD-box RNA helicase DDX5. Cell Rep *7*, 1887-1899.

McDonald, O.G., Wu, H., Timp, W., Doi, A., and Feinberg, A.P. (2011). Genome-scale epigenetic reprogramming during epithelial-to-mesenchymal transition. Nat Struct Mol Biol *18*, 867-874.

Merrikh, H., Zhang, Y., Grossman, A.D., and Wang, J.D. (2012). Replication-transcription conflicts in bacteria. Nat Rev Microbiol *10*, 449-458.

Mersaoui, S.Y., Yu, Z., Coulombe, Y., Karam, M., Busatto, F.F., Masson, J.Y., and Richard, S. (2019). Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids. EMBO J *38*, e100986.

Michalet, X., Ekong, R., Fougerousse, F., Rousseaux, S., Schurra, C., Hornigold, N., van Slegtenhorst, M., Wolfe, J., Povey, S., Beckmann, J.S.*, et al.* (1997). Dynamic molecular combing: stretching the whole human genome for high-resolution studies. Science *277*, 1518-1523.

Millanes-Romero, A., Herranz, N., Perrera, V., Iturbide, A., Loubat-Casanovas, J., Gil, J., Jenuwein, T., Garcia de Herreros, A., and Peiro, S. (2013). Regulation of heterochromatin transcription by Snail1/LOXL2 during epithelial-to-mesenchymal transition. Mol Cell *52*, 746-757.

Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic Acids Res *38*, 2757-2774.

Mischo, H.E., Chun, Y., Harlen, K.M., Smalec, B.M., Dhir, S., Churchman, L.S., and Buratowski, S. (2018). Cell-Cycle Modulation of Transcription Termination Factor Sen1. Mol Cell *70*, 312-326 e317.

Mischo, H.E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Mol Cell *41*, 21-32.

Muller-McNicoll, M., and Neugebauer, K.M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. Nat Rev Genet *14*, 275-287.

Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N.A., P, O.B., Sato, H., Zhang, Z., Jeddeloh, J., Montagna, C., Golden, A.*, et al.* (2015). RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. Epigenetics Chromatin *8*, 46.

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

Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability--an evolving hallmark of cancer. Nat Rev Mol Cell Biol *11*, 220-228.

Nieto, M.A., Huang, R.Y., Jackson, R.A., and Thiery, J.P. (2016). Emt: 2016. Cell *166*, 21-45.

Nowotny, M., Cerritelli, S.M., Ghirlando, R., Gaidamakov, S.A., Crouch, R.J., and Yang, W. (2008). Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD. EMBO J *27*, 1172-1181.

Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I., and Fischer, T. (2016). Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. Cell *167*, 1001-1013 e1007.

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

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.P.*, et al.* (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell *161*, 774-789.

Peinado, H., Ballestar, E., Esteller, M., and Cano, A. (2004). Snail mediates E-cadherin repression by the recruitment of the Sin3A/histone deacetylase 1 (HDAC1)/HDAC2 complex. Mol Cell Biol *24*, 306-319.

Peinado, H., Del Carmen Iglesias-de la Cruz, M., Olmeda, D., Csiszar, K., Fong, K.S., Vega, S., Nieto, M.A., Cano, A., and Portillo, F. (2005). A molecular role for lysyl oxidase-like 2 enzyme in snail regulation and tumor progression. EMBO J *24*, 3446-3458.

Peinado, H., Olmeda, D., and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nat Rev Cancer *7*, 415-428.

Pena, A., Gewartowski, K., Mroczek, S., Cuellar, J., Szykowska, A., Prokop, A., Czarnocki-Cieciura, M., Piwowarski, J., Tous, C., Aguilera, A.*, et al.* (2012). Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. EMBO J *31*, 1605-1616.

Petryk, N., Kahli, M., d'Aubenton-Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., Thermes, C., Chen, C.L., and Hyrien, O. (2016). Replication landscape of the human genome. Nat Commun *7*, 10208.

Phillips, D.D., Garboczi, D.N., Singh, K., Hu, Z., Leppla, S.H., and Leysath, C.E. (2013). The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. J Mol Recognit *26*, 376-381.

Piruat, J.I., and Aguilera, A. (1998). A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. EMBO J *17*, 4859-4872.

Powell, W.T., Coulson, R.L., Gonzales, M.L., Crary, F.K., Wong, S.S., Adams, S., Ach, R.A., Tsang, P., Yamada, N.A., Yasui, D.H.*, et al.* (2013). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proc Natl Acad Sci U S A *110*, 13938-13943.

Prado, F., and Aguilera, A. (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. EMBO J *24*, 1267-1276.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501-512.

Pryor, A., Tung, L., Yang, Z., Kapadia, F., Chang, T.H., and Johnson, L.F. (2004). Growth-regulated expression and G0-specific turnover of the mRNA that encodes URH49, a mammalian DExH/D box protein that is highly related to the mRNA export protein UAP56. Nucleic Acids Res *32*, 1857-1865.

Putnam, A.A., and Jankowsky, E. (2013). DEAD-box helicases as integrators of RNA, nucleotide and protein binding. Biochim Biophys Acta *1829*, 884-893.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.

Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. Curr Opin Cell Biol *17*, 269-273.

Reed, R., and Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. Cell *108*, 523-531.

Rondon, A.G., and Aguilera, A. (2019). What causes an RNA-DNA hybrid to compromise genome integrity? DNA Repair (Amst) *81*, 102660.

Rondon, A.G., Jimeno, S., and Aguilera, A. (2010). The interface between transcription and mRNP export: from THO to THSC/TREX-2. Biochim Biophys Acta *1799*, 533-538.

Rougemaille, M., Dieppois, G., Kisseleva-Romanova, E., Gudipati, R.K., Lemoine, S., Blugeon, C., Boulay, J., Jensen, T.H., Stutz, F., Devaux, F.*, et al.* (2008). THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association. Cell *135*, 308-321.

Saguez, C., Gonzales, F.A., Schmid, M., Boggild, A., Latrick, C.M., Malagon, F., Putnam, A., Sanderson, L., Jankowsky, E., Brodersen, D.E.*, et al.* (2013). Mutational analysis of the yeast RNA helicase Sub2p reveals conserved domains required for growth, mRNA export, and genomic stability. RNA *19*, 1363-1371.

Saguez, C., Schmid, M., Olesen, J.R., Ghazy, M.A., Qu, X., Poulsen, M.B., Nasser, T., Moore, C., and Jensen, T.H. (2008). Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. Mol Cell *31*, 91-103.

Salas-Armenteros, I., Perez-Calero, C., Bayona-Feliu, A., Tumini, E., Luna, R., and Aguilera, A. (2017). Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. EMBO J *36*, 3532-3547.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning : a laboratory manual, 2nd edn (New York: Cold Spring Harbor Laboratory).

Santos-Pereira, J.M., and Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. Nat Rev Genet *16*, 583-597.

Sanz, L.A., and Chedin, F. (2019). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat Protoc *14*, 1734-1755.

Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chedin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol Cell *63*, 167-178.

Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G.P., Heron, M., Williams, H., Soding, J., Stewart, A., and Svejstrup, J.Q. (2014). RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. Cell *157*, 1037-1049.

Saporita, A.J., Chang, H.C., Winkeler, C.L., Apicelli, A.J., Kladney, R.D., Wang, J., Townsend, R.R., Michel, L.S., and Weber, J.D. (2011). RNA helicase DDX5 is a p53-independent target of ARF that participates in ribosome biogenesis. Cancer Res *71*, 6708-6717.

Sarkar, M., and Ghosh, M.K. (2016). DEAD box RNA helicases: crucial regulators of gene expression and oncogenesis. Front Biosci (Landmark Ed) *21*, 225-250.

Schwab, R.A., Nieminuszczy, J., Shah, F., Langton, J., Lopez Martinez, D., Liang, C.C., Cohn, M.A., Gibbons, R.J., Deans, A.J., and Niedzwiedz, W. (2015). The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. Mol Cell *60*, 351-361.

Selth, L.A., Sigurdsson, S., and Svejstrup, J.Q. (2010). Transcript Elongation by RNA Polymerase II. Annu Rev Biochem *79*, 271-293.

Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. Cell *125*, 287-300.

Shen, H., Zheng, X., Shen, J., Zhang, L., Zhao, R., and Green, M.R. (2008). Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. Genes Dev *22*, 1796-1803.

Shen, J., Zhang, L., and Zhao, R. (2007). Biochemical characterization of the ATPase and helicase activity of UAP56, an essential pre-mRNA splicing and mRNA export factor. J Biol Chem *282*, 22544-22550.

Shi, H., Cordin, O., Minder, C.M., Linder, P., and Xu, R.M. (2004). Crystal structure of the human ATP-dependent splicing and export factor UAP56. Proc Natl Acad Sci U S A *101*, 17628-17633.

Shivji, M.K.K., Renaudin, X., Williams, C.H., and Venkitaraman, A.R. (2018). BRCA2 Regulates Transcription Elongation by RNA Polymerase II to Prevent R-Loop Accumulation. Cell Rep *22*, 1031-1039.

Shkreta, L., and Chabot, B. (2015). The RNA Splicing Response to DNA Damage. Biomolecules *5*, 2935-2977.

Silva, S., Camino, L.P., and Aguilera, A. (2018). Human mitochondrial degradosome prevents harmful mitochondrial R loops and mitochondrial genome instability. Proc Natl Acad Sci U S A *115*, 11024-11029.

Singh, D.K., Pandita, R.K., Singh, M., Chakraborty, S., Hambarde, S., Ramnarain, D., Charaka, V., Ahmed, K.M., Hunt, C.R., and Pandita, T.K. (2018). MOF Suppresses Replication Stress and Contributes to Resolution of Stalled Replication Forks. Mol Cell Biol *38*.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem *76*, 23-50.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436-439.

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

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell *42*, 794-805.

Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C., Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T.*, et al.* (2016). ENCODE data at the ENCODE portal. Nucleic Acids Res *44*, D726-732.

Soderberg, O., Gullberg, M., Jarvius, M., Ridderstrale, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G.*, et al.* (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. Nat Methods *3*, 995-1000.

Sollier, J., Stork, C.T., Garcia-Rubio, M.L., Paulsen, R.D., Aguilera, A., and Cimprich, K.A. (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. Mol Cell *56*, 777-785.

Song, C., Hotz-Wagenblatt, A., Voit, R., and Grummt, I. (2017). SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. Genes Dev *31*, 1370-1381.

Sridhara, S.C., Carvalho, S., Grosso, A.R., Gallego-Paez, L.M., Carmo-Fonseca, M., and de Almeida, S.F. (2017). Transcription Dynamics Prevent RNA-Mediated Genomic Instability through SRPK2-Dependent DDX23 Phosphorylation. Cell Rep *18*, 334-343.

Stemmler, M.P., Eccles, R.L., Brabletz, S., and Brabletz, T. (2019). Non-redundant functions of EMT transcription factors. Nat Cell Biol *21*, 102-112.

Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes Dev *26*, 163-175.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R.*, et al.* (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304-308.

Stuckey, R., Garcia-Rodriguez, N., Aguilera, A., and Wellinger, R.E. (2015). Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. Proc Natl Acad Sci U S A *112*, 5779-5784.

Sugiura, T., Sakurai, K., and Nagano, Y. (2007). Intracellular characterization of DDX39, a novel growth-associated RNA helicase. Exp Cell Res *313*, 782-790.

Sulli, G., Di Micco, R., and d'Adda di Fagagna, F. (2012). Crosstalk between chromatin state and DNA damage response in cellular senescence and cancer. Nat Rev Cancer *12*, 709-720.

Tedeschi, F.A., Cloutier, S.C., Tran, E.J., and Jankowsky, E. (2018). The DEAD-box protein Dbp2p is linked to noncoding RNAs, the helicase Sen1p, and R-loops. RNA *24*, 1693-1705.

Thiery, J.P. (2003). Epithelial-mesenchymal transitions in development and pathologies. Curr Opin Cell Biol *15*, 740-746.

Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871-890.

Thiery, J.P., and Sleeman, J.P. (2006). Complex networks orchestrate epithelial-mesenchymal transitions. Nat Rev Mol Cell Biol *7*, 131-142.

Tous, C., and Aguilera, A. (2007). Impairment of transcription elongation by R-loops in vitro. Biochem Biophys Res Commun *360*, 428-432.

Tran, P.L.T., Pohl, T.J., Chen, C.F., Chan, A., Pott, S., and Zakian, V.A. (2017). PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. Nat Commun *8*, 15025.

Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C.*, et al.* (2009). Topoisomerase I suppresses genomic instability by preventing

interference between replication and transcription. Nat Cell Biol *11*, 1315-1324.

Vinciguerra, P., and Stutz, F. (2004). mRNA export: an assembly line from genes to nuclear pores. Curr Opin Cell Biol *16*, 285-292.

Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Mol Cell *44*, 978-988.

Wahba, L., Costantino, L., Tan, F.J., Zimmer, A., and Koshland, D. (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. Genes Dev *30*, 1327-1338.

Wang, H., Gao, X., Huang, Y., Yang, J., and Liu, Z.R. (2009). P68 RNA helicase is a nucleocytoplasmic shuttling protein. Cell Res *19*, 1388-1400.

Wang, X., Liu, H., Zhao, C., Li, W., Xu, H., and Chen, Y. (2016). The DEAD-box RNA helicase 51 controls non-small cell lung cancer proliferation by regulating cell cycle progression via multiple pathways. Sci Rep *6*, 26108.

Wang, Y., Shi, J., Chai, K., Ying, X., and Zhou, B.P. (2013). The Role of Snail in EMT and Tumorigenesis. Curr Cancer Drug Targets *13*, 963-972.

Wei, X., Samarabandu, J., Devdhar, R.S., Siegel, A.J., Acharya, R., and Berezney, R. (1998). Segregation of transcription and replication sites into higher order domains. Science *281*, 1502-1506.

Wellinger, R.E., Prado, F., and Aguilera, A. (2006). Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. Mol Cell Biol *26*, 3327-3334.

Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*.

Wongtrakoongate, P., Riddick, G., Fucharoen, S., and Felsenfeld, G. (2015). Association of the Long Non-coding RNA Steroid Receptor RNA Activator (SRA) with TrxG and PRC2 Complexes. PLoS Genet *11*, e1005615.

Wortham, N.C., Ahamed, E., Nicol, S.M., Thomas, R.S., Periyasamy, M., Jiang, J., Ochocka, A.M., Shousha, S., Huson, L., Bray, S.E., *et al.* (2009). The DEAD-box protein p72 regulates ERalpha-/oestrogen-dependent transcription and cell growth, and is associated with improved survival in ERalpha-positive breast cancer. Oncogene *28*, 4053-4064.

Xing, Z., Ma, W.K., and Tran, E.J. (2019). The DDX5/Dbp2 subfamily of DEAD-box RNA helicases. Wiley Interdiscip Rev RNA *10*, e1519.

Xing, Z., Wang, S., and Tran, E.J. (2017). Characterization of the mammalian DEAD-box protein DDX5 reveals functional conservation with S.

cerevisiae ortholog Dbp2 in transcriptional control and glucose metabolism. RNA *23*, 1125-1138.

Xu, B., and 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 J *15*, 3135-3143.

Yamazaki, T., Fujiwara, N., Yukinaga, H., Ebisuya, M., Shiki, T., Kurihara, T., Kioka, N., Kambe, T., Nagao, M., Nishida, E.*, et al.* (2010). The closely related RNA helicases, UAP56 and URH49, preferentially form distinct mRNA export machineries and coordinately regulate mitotic progression. Mol Biol Cell *21*, 2953-2965.

Yang, Q., Del Campo, M., Lambowitz, A.M., and Jankowsky, E. (2007). DEAD-box proteins unwind duplexes by local strand separation. Mol Cell *28*, 253-263.

Yasuhara, T., Kato, R., Hagiwara, Y., Shiotani, B., Yamauchi, M., Nakada, S., Shibata, A., and Miyagawa, K. (2018). Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. Cell *175*, 558-570 e511.

Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382-2383.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol *4*, 442-451.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. Nat Cell Biol *16*, 2-9.

Zenklusen, D., Vinciguerra, P., Wyss, J.C., and Stutz, F. (2002). Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. Mol Cell Biol *22*, 8241-8253.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Giron, C.G.*, et al.* (2018). Ensembl 2018. Nucleic Acids Res *46*, D754-D761.

Zhang, H., Xing, Z., Mani, S.K., Bancel, B., Durantel, D., Zoulim, F., Tran, E.J., Merle, P., and Andrisani, O. (2016). RNA helicase DEAD box protein 5 regulates Polycomb repressive complex 2/Hox transcript antisense intergenic RNA function in hepatitis B virus infection and hepatocarcinogenesis. Hepatology *64*, 1033-1048.

Zhang, X., Chiang, H.C., Wang, Y., Zhang, C., Smith, S., Zhao, X., Nair, S.J., Michalek, J., Jatoi, I., Lautner, M.*, et al.* (2017). Attenuation of RNA

polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. Nat Commun *8*, 15908.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhao, R., Shen, J., Green, M.R., MacMorris, M., and Blumenthal, T. (2004). Crystal structure of UAP56, a DExD/H-box protein involved in pre-mRNA splicing and mRNA export. Structure *12*, 1373-1381.

Zhu, X., Li, K., and Salah, A. (2013). A data parallel strategy for aligning multiple biological sequences on multi-core computers. Comput Biol Med *43*, 350-361.

Zonta, E., Bittencourt, D., Samaan, S., Germann, S., Dutertre, M., and Auboeuf, D. (2013). The RNA helicase DDX5/p68 is a key factor promoting c-fos expression at different levels from transcription to mRNA export. Nucleic Acids Res *41*, 554-564.