Nuevos genes humanos asociados a la biogénesis de mRNPs necesarios para la integridad del genoma

Irene Salas Armenteros

Tesis Doctoral

Universidad de Sevilla

2017

Nuevos genes humanos asociados a la biogénesis de mRNPs necesarios para 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 Biología por la licenciada Irene Salas Armenteros

Sevilla, 2017

La doctoranda,   Los directores de tesis,

Irene Salas Armenteros   Andrés Aguilera López   Rosa Luna Varo

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

APPENDIX

ABBREVIATIONS

4-NQO	4-Nitroquinoline 1-Oxide
5-FOA	5-Fluorotic Acid
53BP1	p53 Binding Protein 1
AA	Anacardic Acid
AbA	Aureobasidin A
Ade	Adenine
AID	Activation-Induced cytidine Deaminase
AD	Activation Domain
AREX	Alternative mRNA Export complex
AS	Alternative Splicing
ASE	Alternative Splicing Event
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine triphosphate
ATR	Ataxia Telangiectasia and Rad3 Related
ATRIP	ATR Interacting Protein
A.U.	Arbitrary Units
BD	Binding Domain
bp	base pairs
BRCA	BReast CAncer
BRCT	BRCA1 C-terminal
BrdU	Bromodeoxyuridine
BRMS1	Breast Cancer Metastasis Suppressor 1
CBC	Cap-Binding Complex
CBP	Cap-Binding Protein
CC	Coiled-Coil
CDC25	Cell Division Cycle 25
cDNA	complementary DNA
ChIP	Chromatin Immunoprecipitation

ChIP-chip	ChIP combined with microarray
CHK	Checkpoint Kinase
CIN	Chromosome Instability
CNVs	Copy Number Variants
CPT	Camptothecin
CS	CHORD-containing proteins and SGT1
CTD	Carboxy-Terminal Domain
dCTP	Deoxycytidine triphosphate
DDR	DNA Damage Response
DDX	DEAD box
DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide triphosphate
DRIP	DNA-RNA immunoprecipitation
DSB	Double-Strand Break
EGFP	Enhanced Green Fluorescent Protein
EJC	Exon Junction Complex
ESR1	Estrogen Receptor 1
FA	Fanconi Anemia
FACS	Fluorescence-Activated Cell Sorting
FACT	FAcilitates Chromatin Transcription
FAD	Flavin Adenine Dinucleotide;
fSAP	Splicing-Associated Factor
G+C	Guanine and Cytosine
Gal	Galactose
GCRs	Gross Chromosomal Rearrangements
GFP	Green Fluorescent Protein
GO	Gene Ontology
H3K9me2	Histone H3 dimethylation at lysine 9
H3S10-P	Histone H3 phosphorylation at serine 10

HAT	Histone Acetyltransferase
HCR	Highly Conserved Region
HDAC	Histone Deacetylase
HID	Histone Deacetylase Interaction Domain
hnRNP	heterogeneous nuclear Ribonucleoprotein
HR	Homologous Recombination
HU	Hydroxyurea
Ig	Immunoglobulin
ING	Inhibitor of growth
Kb	kilobases
kDa	kilodalton
Leu	Leucine
LOH	Loss Of Heterozygosity
MCM	Mini chromosome maintenance
MFAP1	Microfibrillar Associated Protein 1
MIT	Microtubule Interacting and Trafficking;
MMS	Methyl Methanesulfonate
MRN	Mre11-Rad50-Nbs1
mRNA	messenger RNA
mRNP	messenger Ribonucleoprotein Particle
NHEJ	Non-homologous end joining
NPC	Nuclear Pore Complex
NTC	NineTeen complex
nt	Nucleotides
NTS	Non-transcribed strand
NXF1	RNA export factor 1
ORF	Open reading frame
PAH	Paired Amphipathic Helices
PCR	Polymerase Chain Reaction

pGAL	*GAL1-10* promoter (yeast)
PIKK	Phosphatidylinositol 3 like kinase
Pol	Polymerase
pTET	*Tetracycline promoter* (yeast)
pRB	Retinoblastoma protein
Prp19C	Prp19 complex
Ptet	tet-on promoter (human)
Q	Glutamine
qPCR	quantitative Polymerase Chain Reaction
RBBP	Retinoblastoma-binding protein
RBP	RNA-binding protein
rDNA	Ribosomal DNA
RF	Replication fork
RING	Really Interesting New Gene
RNAi	RNA interference
RNAP	RNA polymerase
RNase	Ribonuclease
RPA	Replication protein A
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SAHA	Suberoylanilide Hydroxamic Acid
SAP	Sin3-associated protein / Scaffold-associated protein
SAP130	Sin3A Associated Protein 130kDa
SH2	Src homology2
SI	Splicing Index
siC	Non-targeting siRNA control
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoprotein

SR	Serine/arginine-rich protein
STUB1	STIP1 homology and U-box containing
SSB	Single-Strand Break
ssDNA	single-stranded DNA
SUDS	Suppressor of defective silencing 3
TAM	Transcription-associated mutation
TAR	Transcription-associated recombination
TCR	Transcription-Coupled Repair
Top	Topoisomerase
Topcc	Topoisomerase Cleavage Complex
TREX	TRanscription-Export complex
Trp	Tryptophan
TRP	Tetratricopeptide
TS	Transcribed Strand
TSA	Trichostatin A
U	Uridine
UBA	Ubiquitin-associated
Ura	Uracil
UV	Ultraviolet
U2AF	U2 Small Nuclear RNA Auxiliary Factor
WT	Wild type
YFP	Yellow Fluorescent Protein

RESUMEN

La integridad del genoma es una condición necesaria para la transmisión fidedigna de la información genética. Numerosos procesos altamente regulados trabajan de forma coordinada para evitar o solucionar problemas que pueden comprometer la estabilidad del genoma. Su inestabilidad es una patología celular que se manifiesta generalmente en forma de mutaciones y reordenaciones cromosómicas y se encuentra asociada a la predisposición a cáncer y envejecimiento. El origen de la inestabilidad genética es variado y no sólo es consecuencia de la acción de agentes genotóxicos externos, sino resultado del propio metabolismo celular, estando ligada a procesos básicos como la replicación, transcripción y recombinación. En esta tesis nos hemos centrado en la transcripción y en el metabolismo del ARN mensajero (ARNm) como fuentes endógenas de inestabilidad genética.

La transcripción puede suponer una amenaza para la integridad del genoma debido a que durante la misma se facilita la aparición de ADN de cadena sencilla, que es más susceptible a daños que la doble cadena. Paralelamente, la transcripción puede suponer un obstáculo para la replicación que puede derivar en un incremento de roturas en el ADN y recombinación, responsable de reordenaciones cromosómicas. Estos fenómenos se pueden agravar cuando además se forman unas estructuras denominadas bucles R (*R loops*), estructuras compuestas por un híbrido de ARN-ADN y la cadena sencilla de ADN desplazada. La formación de *R loops* se produce cuando el ARN naciente resultante de la transcripción, hibrida con la hebra molde de ADN, desplazando así a la hebra no transcrita que queda como cadena sencilla. Aunque los *R loops* pueden desempeñar papeles positivos, se ha demostrado que también pueden amenazar la integridad del genoma.

Muchos factores implicados en las diferentes etapas del procesamiento del ARNm contribuyen a proteger el genoma de la formación de *R loops*. En esta tesis nos hemos centrado en factores con un papel en la biogénesis de las ribonucleoproteínas mensajeras (mRNPs), y en concreto en el complejo THO/TREX. Durante la transcripción, el ARNm necesita ser correctamente empaquetado en mRNPs, permitiendo así la elongación de la transcripción, la integridad y el procesamiento del ARNm y su transporte al citoplasma. Para ello, numerosas proteínas de unión al ARN (RBPs) se asocian con el ARN naciente, de forma que este queda empaquetado y

protegido, reduciendo así la probabilidad de que el ARN hibride con el ADN molde y forme *R loops*. Uno de los factores claves en este proceso es THO/TREX, conservado de levaduras a humanos con un papel en el acoplamiento de la transcripción con la biogénesis y transporte de mRNPs. Los mutantes de este complejo acumulan *R loops* y muestran alta inestabilidad genética. Esto se explica en gran medida por el hecho de que en ausencia de este complejo la mRNP no se forma correctamente y por tanto el ARNm queda más desprotegido, facilitando así la formación de *R loops* e incrementando la inestabilidad genética.

En esta tesis hemos querido profundizar sobre los mecanismos por los cuales la correcta biogénesis de mRNPs contribuye a la integridad del genoma. Para ello, hemos realizado un escrutinio para identificar nuevas proteínas que interaccionen con el complejo THO/TREX humano. Como resultado, hemos identificado dos nuevas interacciones. Hemos mostrado que la subunidad THOC1 del complejo THO/TREX humano, interacciona con el complejo histona desacetilasa Sin3A, y con MFAP1, un factor asociado al madurosoma (*spliceosome*).

THOC1 interacciona físicamente con las subunidades SAP130 y SIN3 del complejo Sin3A. El silenciamiento de las subunidades del complejo *SAP130*, *SIN3*, *SAP30* y *SUDS3* causa inestabilidad genética, determinada por el incremento de roturas en el ADN. Esta inestabilidad, al igual que sucede en ausencia de THO, está mediada por la formación de *R loops*, puesto que el incremento de roturas en el ADN en ausencia de subunidades del complejo Sin3A tales como SAP130 y SIN3 se suprime mediante la sobreexpressión de RNasa H, la cual degrada los híbridos de ARN-ADN. Hemos demostrado que la inhibición de la desacetilación de histonas mediante compuestos químicos conduce a una acumulación de *R loops*, y aún más importante, que la inhibición de la acetilación de histonas mediante compuestos químicos suprime el daño en el ADN y la formación de *R loops* causados por el silenciamiento de *THOC1*. En general, la acetilación de las histonas da lugar a una cromatina más abierta y una activación de la transcripción, mientras que la desacetilación de histonas se asocia con una cromatina más cerrada o compactada y a una represión de la transcripción. Por tanto, estos resultados permiten proponer un modelo en el que la desacetilacion de histonas sería necesaria para prevenir la formación de *R loops* tras el paso de la ARN polimerasa. THO podría interaccionar con Sin3A para contribuir a la desacetilación de

histonas, con objeto de cerrar la cromatina transitoriamente y así prevenir que el ARN naciente hibride con el ADN molde.

El silenciamiento de *MFAP1*, en cambio, provoca roturas en el ADN que no dependen de la formación de *R loops*. Los análisis globales de expresión génica y maduración de intrones (*splicing*) en células silenciadas para *MFAP1* sugieren que el impacto de este factor en la estabilidad del genoma puede deberse principalmente a su papel en *splicing*. El hecho de que el silenciamiento de *MFAP1* afecte en gran medida al *splicing* de genes implicados en la reparación del ADN, el ciclo celular y la organización y modificación de la cromatina entre otros, sugiere que el papel de MFAP1 en la estabilidad del genoma es indirecto; es decir, mediado por los genes cuyo *splicing* regula. El *splicing* de algunos de estos genes también presenta cambios en ausencia de THOC1. Dado que THO sí desempeña un papel directo en el mantenimiento de la estabilidad del genoma, estos cambios se podrían explicar como una consecuencia indirecta de la inestabilidad genética dependiente de transcripción y *R loops* que causa la ausencia de THOC1. No obstante, no se puede descartar que THOC1 también regule la integridad del genoma a través de su interacción con MFAP1 y su efecto en el *splicing* de algunos genes. Estos datos son particularmente relevantes al sugerir que no todas las proteínas de unión a ARN tienen un papel directo en estabilidad del genoma, pudiendo ser su efecto en muchos casos indirecto, consecuencia de su efecto en la regulación de la expresión de otros genes.

INTRODUCTION

1. GENOME INSTABILITY

Genome comprises the complete genetic information of an organism, which is stored in the form of DNA (or deoxyribonucleic acid) molecule. Genome integrity must be preserved to ensure its intact transmission to its offspring. Numerous highly regulated processes work in a coordinated manner to avoid or solve problems that may affect the stability of the genome.

Genome changes occur naturally at low frequency induced by environmental agents or generated spontaneously due to the cellular metabolism itself. This generates genetic variability, which although it can have negative consequences, is also the basis of the evolution (Aguilera and Gomez-Gonzalez 2008). In addition, this variability can be also an advantage when it occurs in a regulated manner. This is the case of immunoglobulin (Ig) genes of B lymphocytes, in which altering its sequences is the key for the immune system to provide a large battery of different antigens to deal with a high variety of pathogens. However, the frequency of these genome alterations can increase under certain circumstances and therefore, the stability of the genome is seriously threatened. Exposure to external genotoxic agents or cellular pathologies, mainly those that affect a proper DNA repair and/or replication, are examples of situations that lead to increased genetic instability. The fact that genome instability is a hallmark of aging, predisposition to various types of cancer and different genetic diseases; highlights the importance of maintaining genome stability (Aguilera and Garcia-Muse 2013).

Many factors can lead to DNA damage and subsequent changes in the genome, since DNA is a dynamic molecule. These include both exogenous sources, such as ultraviolet (UV) and ionizing radiations and chemicals agents, and endogenous sources, including byproducts from cellular metabolism itself and those processes that act on the DNA (replication, repair, recombination and transcription) (Hoeijmakers 2009). Due to these factors, DNA can undergo different types of lesions that include abasic sites, bases mismatch, DNA adducts, inter- and intra-strand crosslinks, single-stranded DNA (ssDNA) gaps and double-strand breaks (DSBs) (Shiloh 2003; Helleday et al. 2014). As a consequence, genetic instability is generated in the form of mutations, such as point mutation and microsatellite contractions and expansions, and chromosomal rearrangements, such as chromosome instability (CIN), gross chromosomal

rearrangements (GCRs), copy number variants (CNVs), loss of heterozygosity (LOH) and hyper-recombination. CIN refers to variations in the number of chromosomes due to their gain or loss. GCRs include gene amplification, repeat expansions, translocations, inversions, insertions, deletions and duplications. CNVs refer to the variation in the number of copies of a particular DNA fragment in the genotype. LOH is the loss of one of the pair of alleles of a diploid, either due to deletion, gene conversion or chromosome loss (Aguilera and Garcia-Muse 2013).

Given all the factors and failures that can be harmful to genome integrity, cells have evolved a wide range of mechanisms focused on preserving genetic information. This set of mechanisms form a complex signal transduction pathway also known as DNA Damage Response (DDR). DDR comprises from the detection and signaling of DNA damage to recruitment, regulation, coordination and activation of the right factors for efficient DNA repair, thus avoiding pathological consequences (Jeggo et al. 2016).

1.1. The DNA damage response

DDR consists in a complex signaling cascade that senses DNA damage and replication stress and coordinates cell cycle progression and DNA repair with the aim of counteracting DNA damage (Ciccia and Elledge 2010). The activation of DDR starts with the recognition of specific DNA lesions by sensor proteins, which in turn recruit and activate the apical kinases in the cascade: the phosphatidylinositol 3 like kinases (PIKKs) ataxia telangiectasia mutated (Tel1/ATM), ataxia telangiectasia and Rad3 related (Mec1/ATR) (as named in yeast/humans) (Giglia-Mari et al. 2011). ATM is recruited to DSBs by the DSB-recognizing protein complex MRN (Mre11 Rad50 Nbs1). ATR is recruited to ssDNA generated during replication stress by replication protein A (RPA) and its association with ATRIP (ATR interacting protein). Then, ATM/ATR phosphorylate and thus activate various proteins that coordinate the arrest of cell cycle progression and DNA repair pathways to preserve genome integrity (Sulli et al. 2012). This includes phosphorylation of DNA damage mediators such as the histone variant H2AX or p53 binding protein 1 (53BP1) among others, and downstream kinases including the checkpoint kinases 1 and 2 (Chk1/CHK1 and Rad53/CHK2). Phosphorylation of the histone variant H2AX on Ser139 (known as γH2AX) by ATM/ATR spreads over long distances surrounding the DSB and serves as a signal for recruitment or assembly of other factors (including 53BP1, BRCA1 and chromatin

remodelers that facilitates the access of RNA repair factors) at sites of DNA damage, thus promoting DSB repair amplifying the DDR signal (van Attikum and Gasser 2009). 53BP1 is also recruited to DSB sites and enhances ATM activation (Sulli et al. 2012). CHK2 (which is mainly phosphorylated by ATM) and CHK1 (which is mainly phosphorylated by ATR) activate downstream effectors such as p53 and the cell division cycle 25 (CDC25) phosphatases (Sulli et al. 2012). These effectors coordinate the transient arrest of cell cycle progression to allow DNA repair before replication or mitosis ensues. Finally, once the DNA damage is repaired, the cell cycle is restored. Alternatively, if the lesion cannot be repaired or persists, DDR signaling triggers cell death by apoptosis or cellular senescence (Jackson and Bartek 2009; Sulli et al. 2012).

2. TRANSCRIPTION-ASSOCIATED GENOME INSTABILITY

DNA has to deal continuously with several processes that act on it and generate transitory changes in its structure. This implies a risk for genome integrity but is also necessary for the complete functioning of the cell and of an organism. However, such DNA metabolism processes influence genome stability differently. While DNA replication and repair actively contribute to avoid genome alterations and ensure faithfully genome transmission, transcription can contribute negatively to genome integrity (Gaillard and Aguilera 2016).

Transcription is the cellular process that allows the expression of genetic information into RNA. During transcription, the RNA polymerase (RNAP) uses one of the DNA strand, which is transiently separated from its complementary strand, as template (transcribed strand, TS) to generate a complementary RNA chain. While, the non-transcribed strand (NTS) remains unpaired. This occurs in the transcription bubble located inside the RNAP. Moreover, transcription is accompanied by the positive and negative DNA supercoiling and chromatin remodeling changes that are necessary to allow the progression of the RNAP through the template DNA (Selth et al. 2010). At the same time, since transcription is a central part of gene expression, it is coupled with translation in prokaryotes or with mRNA (messenger RNA) processing and export in eukaryotes (Gaillard et al. 2013).

In addition to carrying out its main function, transcription also contributes to a lesser extent in the maintenance of genome stability since it promotes a type of DNA repair

called transcription-coupled repair (TCR) or transcription-coupled nucleotide excision repair (TC-NER). From bacteria to mammalian cells, the arrest of transcription machinery due to an encounter with a DNA lesion serves as a signal to trigger DNA repair through this NER subpathway (Hanawalt and Spivak 2008; Vermeulen and Fousteri 2013).

However, due to the intrinsic characteristics of transcription itself and its relationship with those processes that are coordinated with it, transcription is also a source of genome instability. The first evidence of this were the increase in mutations and recombination rates of transcribed regions versus non-transcribed regions, known as transcription-associated mutation (TAM) and transcription-associated recombination (TAR) respectively. Both TAM and TAR are conserved from prokaryotes to eukaryotes (Aguilera 2002).

2.1. Transcription-associated mutagenesis (TAM)

In 1971, two independent laboratories were the first to report that transcription increases the rate of mutations. Both studies in *Escherichia coli* showed that reversion rates of a particular mutated gene in the presence of genotoxic agents are higher when such gene is transcribed (Brock 1971; Herman and Dworkin 1971). Since then, numerous assays and systems to sense mutations in different organisms, from bacteria to human cells, have continued to evidence this fact (Jinks-Robertson and Bhagwat 2014).

In addition to showing that transcription increases the susceptibility to undergo mutations, two more conclusions are deduced from all these studies. On one hand, the rate of mutation increases directly proportional to the level of transcription. This was revealed through the use of systems under de control of inducible and modulable promoters, such as *GAL1-10* promoter (*pGAL*) and tetracycline promoter (*pTET*) in yeast or tet-on promoter (Ptet) in human cells, thus allowing different levels of transcription (Datta and Jinks-Robertson 1995; Bachl et al. 2001; Kim et al. 2007). On the other hand, the analysis of mutations between the two DNA strand (NTS and TS) showed that mutations along transcribed regions occur predominantly in the NTS, besides the fact that TC-NER occurs preferentially in TS (Hanawalt and Spivak 2008; Gaillard et al. 2013).

The key to understanding how transcription leads to TAM resides in the higher susceptibility of the ssDNA. The unequal proportion of mutations in favor of NTS supports this fact. During transcription, while TS remains paired with nascent RNA, NTS remains transiently single stranded. ssDNA is more susceptible to spontaneous and induced chemical reactions. One example of this is the case of cytosine to uracil conversion by spontaneous deamination. When uracil is not removed and repaired, it is paired with adenine, resulting in C-to-T mutations after replication. Other typical transversions induced by oxidative DNA damage are also more frequent in NTS than in TS (Jinks-Robertson and Bhagwat 2014). Similarly, DNA-damaging drugs or enzymes that act on DNA such as nucleases or the activation-induced cytidine deaminase (AID), required for antibody diversification in B cells, can also access more easily unpaired DNA strand and cause TAM (Maizels 2005). However, only a few nucleotides (9-12) remain transiently unpaired in the transcription bubble, which is also located mostly within the RNAP. In addition it is also unclear whether DNA damage agents can access to them (Kim and Jinks-Robertson 2012; Gaillard et al. 2013).

The topological changes that accompanying the RNAP can lead to an increase in the proportion of unpaired DNA, thus facilitating TAM (Figure I1A) (Wu et al. 1988). The RNAP progression entails local positive supercoiling (overwinding) ahead and negative supercoiling (underwinding) behind. Topoisomerases are the proteins in charge of releasing the DNA from such torsion. Topoisomerase 1 (Top1) and topoisomerase 2 (Top2) resolve negative and positive supercoils respectively, thus allowing the progression of the polymerase. For that, topoisomerases introduce transient DNA breaks that enable to relax the supercoiled DNA and then reseals the nick (Pommier et al. 2016). The incorrect action or the absence of topoisomerases can result in TAM in a different way. It has been reported that failures in the reactions carried out by Top1 causes its irreversible attachment to the end of the nicked strand (Lippert et al. 2011; Takahashi et al. 2011). This intermediate structure termed the topoisomerase 1 cleavage complex (Top1cc) is presumably removed generating a gap, which in turn could lead to deletion after replication. Topological changes also can contribute to TAM by increasing the proportion of ssDNA. In the absence of Top1, negatively supercoiled DNA is accumulated behind the RNAP. These regions of DNA with hypernegative coiling (underwound DNA) favor the formation of ssDNA stretches. Furthermore, this scenario facilitates the formation of R loops, which are formed by co-transcriptional

RNA-DNA hybrids and a displaced single strand DNA (ssDNA) (Drolet et al. 1995; El Hage et al. 2010). Hypernegative supercoiled DNA enables the RNA to hybridize back to the TS more easily, leading the NTS single stranded and more vulnerable to genotoxic agents (Figure I1B) (Gaillard and Aguilera 2016).

In other cases it is the RNA instead of the DNA which promotes R-loop formation. When nascent mRNA is not correctly packed into a messenger ribonucleoprotein particle (mRNP), as occurs in mRNA biogenesis mutants, and invades the DNA duplex (Figure I1B). Many other factors can contribute to the accumulation R loops and therefore lead to increased genomic instability (Huertas and Aguilera 2003; Aguilera and Garcia-Muse 2012).
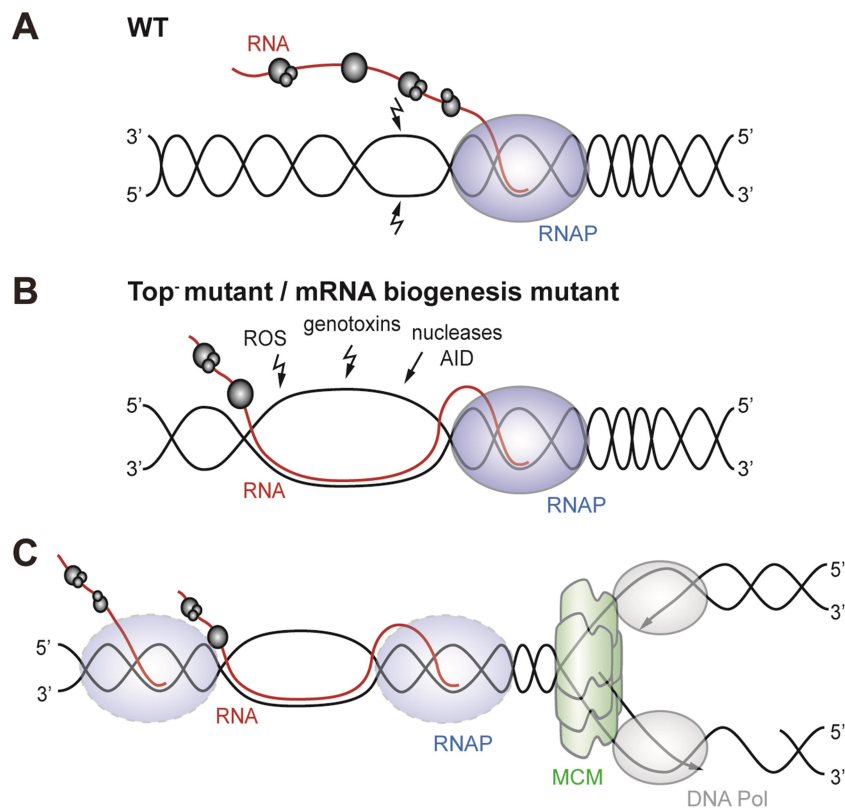
Figure I1. Transcription-associated genome instability.
(A) The progression of transcription entails the formation of local negative supercoiling behind the RNAP that makes the DNA more susceptible to be damaged in wild type (WT) cells. **(B)** Hypernegative supercoiling in topoisomerase (Top⁻) mutant and deficient packaging of the nascent mRNA in mRNA biogenesis mutant facilitate the formation of R loops. The displaced NTS within the R-loop remains ssDNA and therefore is more susceptible to be damaged by genotoxic agents, including reactive oxygen species (ROS), nucleases and other modifying enzymes such as AID. **(C)** Transcription by itself can hamper replication when the machinery of both processes collide. Head on collisions (transcription-replication machineries progress in opposite directions) lead to the accumulation of positive DNA supercoiling between both machineries and negative supercoiling behind the RNAP and therefore seem to be a stronger and more harmful impediment than co-directional collisions (transcription-replication machineries progress in the same direction). MCM, minichromosome maintenance complex; DNA Pol, DNA polymerase. Figure adapted from (Gaillard et al. 2013).

2.2. Transcription-associated recombination (TAR)

The phenomenon of TAR is observed from bacteria to human. From the first evidence, numerous work have reported this influence of transcription on recombination by using different systems that allow to sense ectopic recombination genetically, which would lead to a hyper-recombination phenotype, as well as chromosomal rearrangements (Gaillard et al. 2013). Many of these recombination systems are based on direct repeats that are transcribed at different levels depending on the promoter. Transcription stimulates deletions within these systems, normally detected by the gain or loss of a marker, which occur mostly as a result of DSBs repair by HR. Thus, TAR is mainly caused by DNA breaks within the actively transcribed region (Gaillard and Aguilera 2016). Moreover, TAR is necessarily linked to replication. In yeast, the use of recombination systems in which transcription is activated from an S or G_1-specific promoter showed that recombination is increased only when transcription occurs during S phase (Prado and Aguilera 2005). The DNA Pol2, the catalytic subunit of DNA polymerase epsilon, and the Rrm3 helicase, required for RF progression through obstacles in *S. cerevisiae*, both were found enriched along highly transcribed genes, thus supporting the fact that transcription interferes with replication (Azvolinsky et al. 2009). In mammalian cells, it has also been reported that TAR is dependent on replication (Gottipati et al. 2008). This connection between TAR and replication is related to how DNA damage is repaired during this process. Homologous recombination (HR) and non-homologous end joining (NHEJ) are the two main pathways to repair DSBs. In HR repair, the ends of the DSB are processed and prepared to invade the homologous sister chromatid, which will be used as a template to repair. In contrast, NHEJ pathway repairs by directly ligation of DSB ends, without using any template (Heyer et al. 2010; Chapman et al. 2012; Gaillard and Aguilera 2016). Thus, HR is promoted during S and G_2 phases, where sister chromatid is present, and NHEJ acts preferentially in G_1 and in the early S phases. Altogether evidences that TAR is promoted by DNA breaks generated as a consequence of transcription-replication conflicts.

2.3. Transcription-replication conflicts

Transcription is a major source of obstacles for replication, resulting in an increase of DNA breaks, DNA recombination and genome instability. Several other factors can also negatively influence the replication, including dysfunctions of the replication machinery itself as well as the encounter with an obstacle such as DNA bulky adducts, non-B DNA structures or other tightly-bound DNA-protein complex (Aguilera and Garcia-Muse 2013). Any of them can hamper the progression of the replication fork (RF) and therefore entails a threat to genome integrity (Zeman and Cimprich 2014). The encounter of the RF with an obstacle can cause transient pausing of the fork or RF stalling, if it is a longer pausing. At this point, replication checkpoint is activated and triggers the stabilization of the stalled RF, thus avoiding its dissociation from the DNA template. Once a stalled RF is stabilized, it can be restarted if the obstacle has been removed. There are also pathways involved in RF restart through bypassing the obstacle when it has not been removed. However, if these systems fail or the cause of the replicative stress persists DSBs may accumulate at stalled forks and/or DNA polymerases may disassemble leading to RF collapse (Aguilera and Garcia-Muse 2013; Zeman and Cimprich 2014; Gaillard and Aguilera 2016). Consequently, the repair of such breaks during replication involves an increase of recombination events. Due to this fact, also increases the probability of undergoing ectopic recombination between fragment that shares homologies and different chromosomal rearrangement depending on how these recombination events are resolved, leading to genome instability (Aguilera and Garcia-Muse 2013).

Transcription can hinder RF progression indirectly, through obstacles resulting from its activity, or directly, due to transcription-replication collisions. Transient formation of ssDNA during transcription, which is more susceptible to DNA damage, can lead to different DNA lesions that can interfere with replication if they are not properly repaired (Garcia-Muse and Aguilera 2016). Thus, the formation of ssDNA stretches as a consequence of negative supercoiled DNA accumulation and/or RNA-DNA hybrid favors the formation of these types of obstacles that may interfere with DNA synthesis. Moreover, ssDNA not only facilitates RNA-DNA but also other types of secondary structures including hairpins, G-quadruplexes, cruciforms and others. These non-B DNA structures lead to genome instability, derived in part from its ability to block the replisome (Aguilera and Gomez-Gonzalez 2008; Kim and Jinks-

Robertson 2012; Aguilera and Garcia-Muse 2013; Gaillard and Aguilera 2016). However, all this does not seem to be sufficient to explain TAR.

Accumulating evidence indicates that TAR is mainly due to transcription-replication collisions. Since both processes share the same substrate, the collisions between transcription and replication machineries are inevitable. Depending on the direction, transcription-RF collisions can be co-directional or head-on (Kim and Jinks-Robertson 2012). When transcription and replication of leading strand use the same DNA strand as template, both machineries move in the same direction (co-directional). In contrast, if transcribed strand coincides with the lagging strand, direction of transcription and replication converges (head-on) (Figure I1C). Both types of collisions hamper RF progression and increase TAR but head-on collisions seem to be more harmful (Prado and Aguilera 2005).

There are several evolutionary examples about the strategies developed by the cells to prevent or minimize such collisions. In *E. coli*, most of the genes are oriented co-directionally, with respect to the unique origin of replication of bacteria chromosome (Brewer 1988). In the human gene organization, transcription is also preferentially co-oriented with RFs (Huvet et al. 2007). However, additional strategies are necessary due to the higher complexity of eukaryotic genomes, with multiple chromosomes and numerous replication origins. In budding yeast, replications fork barriers downstream of rRNA loci physically block fork to avoid head-on collisions (Brewer and Fangman 1988). Transcription through gene looping can serve as barrier to avoid both types of collisions (Duch et al. 2013). Moreover, transcription and replication in eukaryotic cells have some spatial and temporal separation although some genes are transcribed during the S-phase (Wei et al. 1998). On the other hand, all those mechanism and factors that help avoid RNAPII stalling also contribute to minimize transcription-replication encounters. For example, the RECQ5 DNA helicase, which associates with RNAPII, prevents RNAPII from pausing or arrest by modulating its movement across genes (Li et al. 2011; Saponaro et al. 2014).

3. R LOOPS AS A SOURCE OF GENOME INSTABILITY

Co-transcriptional R loops are one of the major contributors to the potential negative impact of transcription on genome integrity. As mentioned above, an R-loop is a three-strand nucleic acid structure formed by an RNA-DNA hybrid and the resultant displaced ssDNA. During transcription, R loops can form when nascent RNA transcript hybridizes with its DNA template strand, leading the complementary strand as ssDNA (known as the "thread-back" model). Although R loops may have positive roles, evidence has accumulated that they are a major threat to genome integrity (Aguilera and Garcia-Muse 2012).

R loops occur naturally since they are intermediary structures formed during several processes such as *E. coli* plasmid replication, mitochondrial DNA replication or immunoglobulin (Ig) class switching (Aguilera and Garcia-Muse 2012). Moreover, over the last few years, accumulating evidence indicates that R loops can be also potential regulators of gene expression. Genome-wide analyses reported that R loops are enriched at promoter and terminator regions of numerous human genes (Ginno et al. 2013). In fact, R loops seem to be important intermediaries in transcription activation (Ginno et al. 2012), especially at CpG island promoters, as well as in transcription termination in numerous yeast and human genes (Ginno et al. 2013; Skourti-Stathaki et al. 2014). Nevertheless, when R loops form as aberrant byproducts of transcription and accumulate at a higher frequency they entail a major threat to gene expression and genome integrity.

3.1. R-loop-mediated genome instability

Due to its intrinsic characteristics and the consequences of its formation, R loops are a potential source of genome instability. The formation of RNA-DNA hybrids during transcription with the consequent displacement of the complementary DNA strand increases the proportion of ssDNA, which is more susceptible to mutagenic DNA damage than dsDNA (Figure I2A) (Aguilera 2002; Santos-Pereira and Aguilera 2015). Due to this fact, hypermutation is one of the phenotypes caused by R loops. These mutations may occur either spontaneously or enhanced by mutagenic agents or by the action of DNA-modifying enzymes such as AID (Figure I2A). During immunoglobulin diversification in B cells, AID activity initiates both processes; somatic hypermutation

34

(SHM) and class switch recombination (CSR) at the switch regions (S regions) of the Ig genes. Transcription is necessary for this process since AID acts on ssDNA (Chaudhuri and Alt 2004). The formation of R loops behind the RNAP contributes to provide this substrate to the AID enzyme (Yu et al. 2005). However, AID can erroneously act on off-target genes, especially when R loops accumulate at high levels. An evidence of this is the fact that off-target AID activity is enriched at transcribed enhancers and promoters, regions with high presence of hybrids (Ginno et al. 2013; Qian et al. 2014). Due to the relationship between AID and R loops, the increase on genome instability after ectopic expression of AID had been used as an indirect method to detect an increase in R loops. For example, AID expression in R-loop-accumulating yeast THO mutant increases mutation in transcribed genes (Gomez-Gonzalez and Aguilera 2007). Moreover, hypermutation mediates by R loops can occur as a consequence of non-canonical replication, when the R-loop is used as a primer (Kogoma 1997), since the fidelity of replication could be lower.

In addition to facilitating mutagenesis, R-loop-mediated genome instability is mainly attributed to increased recombination and chromosome rearrangements as a result of DNA breaks. Although the mechanisms by which R loops lead to the formation of ssDNA gap or DSBs are not fully understood yet, evidence suggest that replication is a key factor in R-loop-mediated genome instability (Wellinger et al. 2006; Madireddy et al. 2016). Many examples from bacteria to humans support this idea. Replication fork progression impairment had been detected in many R-loop-accumulating yeast and human cells, using 2D gel electrophoresis, DNA combing or the recruitment of the Rrm3 DNA helicase. Moreover, in these cases, the overexpression of ribonuclease H1 (RNase H1), which specifically degrades the RNA moiety of DNA-RNA hybrids, decreases or suppresses replication impairments as well as DNA breaks and recombination. Thus, the harmful potential of R loops relies in part in its ability to hamper RF progression (Aguilera and Garcia-Muse 2012; Santos-Pereira and Aguilera 2015).

There are several possible scenarios to explain how DNA breaks can occur as a consequence of R loops and R loops-replication conflicts. On one hand, the higher susceptibility of the ssDNA present in an R-loop to DNA lesions, especially when they are not properly repaired, can lead to single-strand breaks (SSBs). These unrepaired DNA lesions can block RF progression, which in turn can lead to DSBs formation

(Figure I2A). Moreover, the ssDNA is also more prone to form non-B DNA structures such as G-quadruplexes and hairpins, which may also be an obstacle to replication. On the other hand, the RNA-DNA hybrid structure is a major barrier for replication. It is able to stall or block RF progression directly by itself. Indirectly, R loops can hamper replication by promoting transcription-replication collisions (Figure I2B). This occurs when the RNAP remain trapped at the transcription site as a consequence of a stable R-loop (Aguilera and Garcia-Muse 2012). However, recent findings establish a connection between R loops and chromatin changes which in turn could be largely responsible for R-loop-associated genome instability (Chedin 2016). In particular, it had been reported that the accumulation of aberrant R loops triggers histone H3 serine 10 phosphorylation (H3S10-P), a mark of chromatin condensation (Castellano-Pozo et al. 2013). Since this compaction could be an important barrier to the RF progression, it is proposed that the responsible for RF staling and genome instability mediated by R loops it may not be just the R loops, but a more compacted chromatin triggered by them (Figure I2C) (Castellano-Pozo et al. 2013).
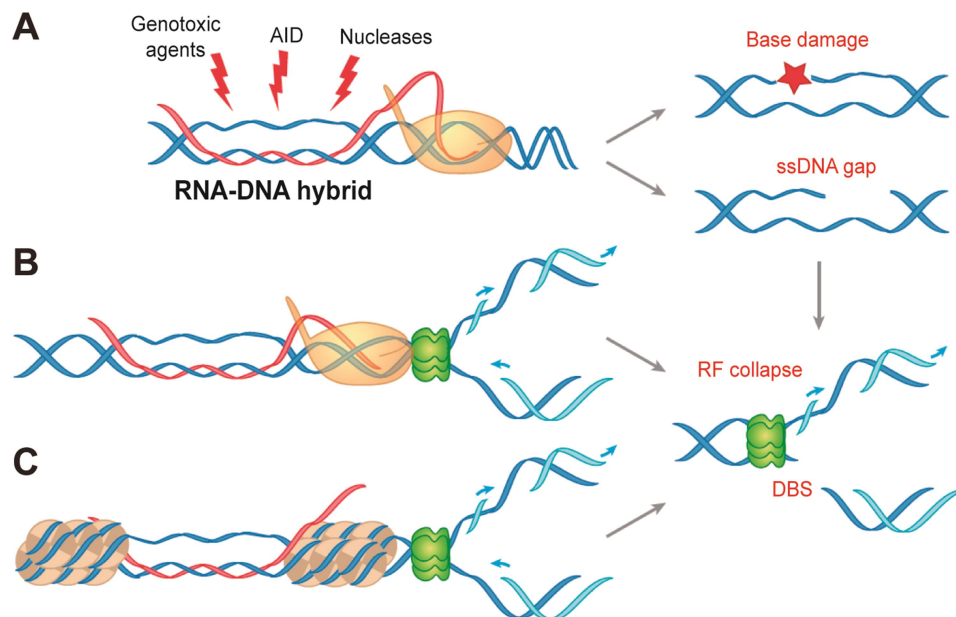
Figure I2. R-loop-mediated genome instability.
R loops are a major source of source of genome instability. **(A)** The formation of R loops involves an increase in the proportion of ssDNA, which is more vulnerable to genotoxic and enzymatic activities such as AID. This in turn can promote different DNA lesions such as base damage (red star), which can lead to mutagenesis, and ssDNA nicks or gaps, which can hamper RF progression and lead to genome instability. **(B)** R loops can also interfere with replication directly by acting as an obstacle or indirectly by promoting transcription-replication collisions. **(C)** Replication can be hinder by local chromatin compaction promoted by R-loop formation. In all this cases R loops impede RF progression and lead to fork stalling, potential collapse, and breakage, thus leading to genome instability. Figure adapted from (Gaillard and Aguilera 2016).

36

3.2. Mechanisms and factors involved in preventing R-loop accumulation

R loops are dynamic structures that are continuously formed and resolved over a large fraction of the genome (Chedin 2016). Cells possess various mechanisms and factors to prevent R-loop formation and accumulation. The coordinated action of these factors maintains R loops controlled *in vivo*. Dysfunction of any of them causes R-loop accumulation and genome instability (reviewed in (Skourti-Stathaki and Proudfoot 2014; Santos-Pereira and Aguilera 2015)).

In general, two types of mechanisms or factors to avoid R-loop accumulation are distinguished: those that help prevent R-loop formation (mRNP biogenesis and topoisomerases) and those that remove R loops once they have formed (ribonucleases, helicases and others).

3.2.1. Topoisomerases

During transcription, the advance of the polymerase is accompanied by topological changes including positive supercoils ahead and negative supercoils behind the RNAP. The accumulation of these supercoils could have detrimental effects on the DNA. In particular, the accumulation of negative supercoils behind the RNAP has been linked to R-loop formation and genome instability since underwinded DNA might facilitate DNA strand opening and subsequent R-loop formation. Topoisomerases are responsible for alleviating these supercoiling and for preventing its accumulation (Santos-Pereira and Aguilera 2015; Gaillard and Aguilera 2016). Therefore, they are considered important factors for R-loop prevention. In *E. coli*, the growth defect of *topA* mutants, which lack Top1, is suppressed by RNase H1 overexpression (Drolet et al. 1995). In yeast, *top1* and *top2* mutant show transcription defects and R-loop accumulation at the ribosomal DNA (rDNA) locus (El Hage et al. 2010). In mammalian, it had been shown that R loops are responsible for RF stall and chromosome breaks observed in TOP1-deficient cells (Tuduri et al. 2009). Therefore, topoisomerases are an important mechanism to prevent R-loop formation during transcription (Figure I3A).

3.2.2. mRNP biogenesis

Many factors involved in different steps of mRNA processing had been shown to be relevant to protect the genome from R loops. During transcription, nascent mRNA needs to be correctly packaged, which is crucial for pre-mRNA processing, mRNA export, mRNA integrity and transcription elongation efficiency. For this, numerous RNA-binding proteins (RBPs) associates with the nascent mRNA, thus leaving the RNA molecule completely packaged and protected, therewith also reducing the chance of the RNA to hybridize back with the DNA template (Figure I3A) (Rondon et al. 2010). The first evidence of this was observed in yeast mutant of THO/TREX, a conserved complex involved in the coupling of transcription with mRNP biogenesis and export (see below). Mutants of this complex accumulate co-transcriptional R loop and exhibit R-loop-dependent hyper-recombination phenotype (Huertas and Aguilera 2003). This is some extent explained by the fact that the absence of this complex leads to the formation of a suboptimal mRNP, which in turn facilitates R-loop formation and increases genome instability (Huertas and Aguilera 2003; Gomez-Gonzalez and Aguilera 2007). Importantly, the role of THO/TREX seems to be conserved since its absence lead to similar transcription and mRNA export defects as well as R-loop-mediated genome instability in yeast, *Caenorhabditis elegans* and human cells (Chavez and Aguilera 1997; Huertas and Aguilera 2003; Dominguez-Sanchez et al. 2011a; Gomez-Gonzalez et al. 2011; Castellano-Pozo et al. 2012).

From this first evidence, many other examples have supported the idea that factors with a role in RNA metabolism are crucial to protect nascent mRNA and to prevent R-loop formation and genome instability. The yeast THSC/TREX-2 complex, also acts at the interface of transcription and mRNA export since it provides an anchoring of transcribed genes to the nuclear pore complex (NPC) (Luna et al. 2012). Mutant of this complex show transcription and mRNA export defects and genome instability, similarly to those observed in THO/TREX mutant (Fischer et al. 2002; Gallardo et al. 2003; Gonzalez-Aguilera et al. 2008). Moreover, this is attributed to R-loop formation since these mutants are sensitive to AID and part of the genome instability can be suppress by RNase H1 overexpression (Gonzalez-Aguilera et al. 2008). Along the same lines, the yeast Npl3, a highly abundant RNA-binding heterogeneous nuclear ribonucleoprotein (hnRNP) involved in mRNP processing and export, has a key role in preventing R-loop-mediated transcription-replication conflicts

and genome instability (Santos-Pereira et al. 2013). In vertebrate cell, serine/arginine-rich (SR) splicing factor 1 (SRSF1), an SR protein involved in splicing but also in mRNA export, was one of the first proteins identified as an important factor in the prevention of R-loop formation (Li and Manley 2005). Chicken DT40 and human HeLa cells depleted of SRSF1 show hypermutation, DNA damage and higher sensitivity to bisulfite mutagenesis, which specifically acts on ssDNA. This is attributed to R-loop formation since genome instability caused by SRSF1 depletion can be suppressed by RNase H overexpression. Similarly, depletion of either BuGZ or Bub3, two mitotic regulators that in interact with the splicing machinery and are required for pre-mRNA splicing, cause splicing defects and increase R-loop formation in several cancer cell lines (Wan et al. 2015).
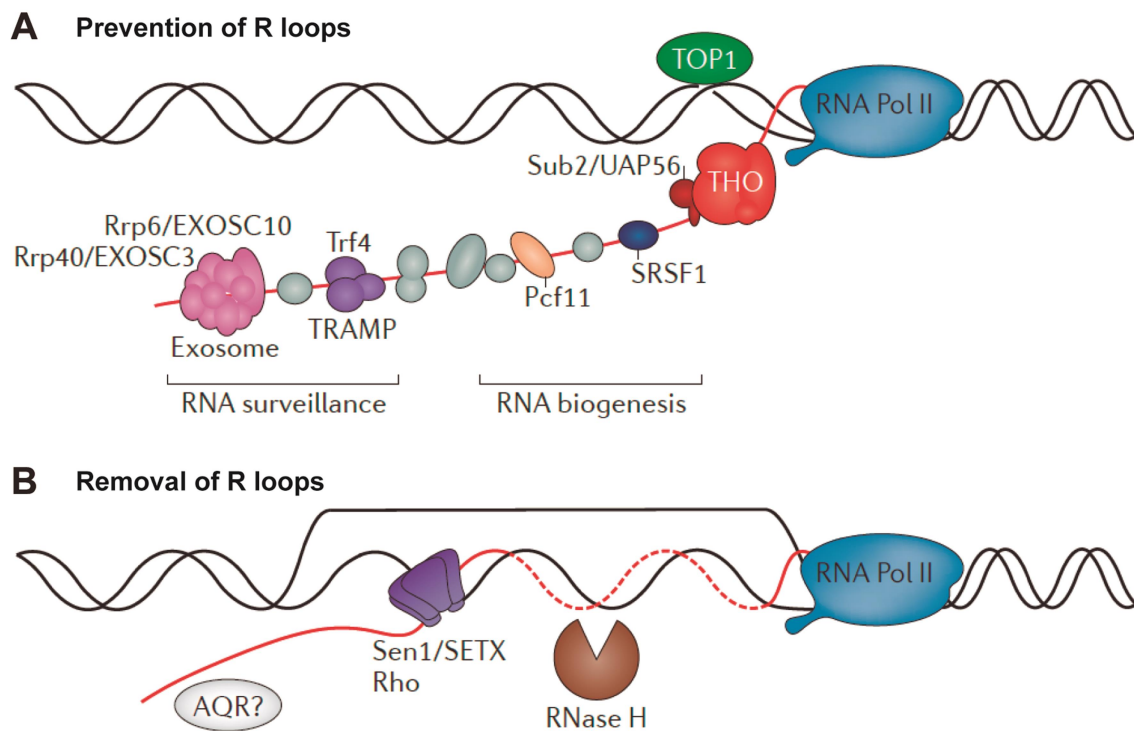
Figure I3. Mechanisms and factors that help prevent R-loop accumulation.
(A) Topoisomerase 1 (TOP1), which alleviates the local negative supercoiling behind the elongating RNAPII, and specific RNA-binding proteins that are involved in RNA biogenesis (including the THO complex, SRSF1 and Pcf11) or RNA surveillance (including Rrp40/EXOSC3 and Rrp6/EXOSC10) help prevent the formation of R loops. **(B)** R loops can be removed by RNase H enzymes, which degrade RNA moiety. Additionally, helicases such as Rho in bacteria, Sen1 in yeast, senataxin (SETX) in human and other putative helicases such as aquarious (AQR) can unwound the RNA-DNA hybrid within the R-loop. Figure adapted from (Santos-Pereira and Aguilera 2015).

In addition to RNA biogenesis, other aspects of RNA metabolism have been involved in the prevention of R-loop formation. This is the case of factors with a role in RNA surveillance such as yeast Trf4, a component of the TRAMP (Trf4-Air2-Mtr4p polyadenilation) complex, or mouse exoribonucleases exosome component 3 (EXOSC3) and EXOS10 (Rrp40 and Rrp6 in yeast, respectively) (Figure I3A) (Gavalda et al. 2013; Pefanis et al. 2015).

In general, several genome-wide screenings as well as more focused analyses of specific factors have continued identifying factors involved in RNA metabolism with an important role in R loops prevention (Paulsen et al. 2009; Wahba et al. 2011; Stirling et al. 2012; Chan et al. 2014), thus supporting the idea that protection of RNA contributes to maintain genome integrity.

3.2.3. Ribonucleases, helicases and others

Once R loops are formed, cells also possess mechanisms to remove them and prevent their accumulation. R loops can be removed by RNase H enzymes, which specifically degrade the RNA moiety of DNA-RNA hybrids (Figure I3B). RNase H1 and RNase H2 are the two main types of RNases H. They differ in composition and have different specialized roles, however both types have the ability to degrade the RNA strand of DNA-RNA hybrids (Cerritelli and Crouch 2009; Skourti-Stathaki and Proudfoot 2014). Despite this, RNase H1 seems to be the major player in removing transcription-associated R loops (Chon et al. 2013). In fact, overexpression of RNase H1 has been widely used to experimentally remove R loops and to suppress R-loop-dependent genome instability phenotypes.

Additionally, the unwinding of DNA-RNA hybrid by the action of DNA-RNA helicases also contributes resolve R loops. A clear example of this is the yeast Sen1 helicase and its human homologue senataxin (SETX). In *S. cerevisiae*, *sen1-1* mutants, which carry a mutation in the Sen1 helicase domain, shows hyper-recombination phenotype and R-loop accumulation (Mischo et al. 2011). In human cells, SETX depletion leads to R-loop accumulation at transcription termination pause sites (Skourti-Stathaki et al. 2011). This suggests that SETX is necessary to resolve R loops generated as intermediaries of the transcription termination process (Figure I3B). Interestingly, depletion of the human aquarius (AQR), an helicase of the same subfamily as

40

Sen1/SETX, also leads to R-loop accumulation, suggesting that AQR may have a role in R-loop resolution (Figure I3B) (Sollier et al. 2014).

Many other factors may indirectly contribute to the resolution of R loops, especially those with a role in DNA repair and/or in mediation of transcription-replication conflicts, since R-loop formation interferes with RF progression. The DSB repair factor and tumor suppressors BRCA1 and BRCA2 had been shown to have a role in preventing R-loop accumulation (Bhatia et al. 2014). BRCA1 is recruited to R loops formed at transcription termination regions and mediates the recruitment of SETX (Figure I4) (Hatchi et al. 2015). Moreover, BRCA1 has a Fanconi anemia (FA)-associated function and BRCA2 is the FANCD1 component of the FA DNA repair pathway and plays a role in stabilizing RFs (Moldovan and D'Andrea 2009). Importantly, it has been recently reported that FA pathway is relevant in protecting cells from naturally formed R loops, thus allowing RF progression through transcribed regions (Figure I4) (Garcia-Rubio et al. 2015). Similarly, the FACT (facilitates chromatin transcription) chromatin-reorganizing complex, that swaps nucleosomes around the RNA polymerase during transcription elongation, has a role in the resolution of R-loop-mediated transcription-replication conflicts (Figure I4). The accumulation of R loops in cells depleted of this factor suggests that chromatin reorganization-mediated by FACT is required for the RF to move through transcribed regions (Herrera-Moyano et al. 2014).
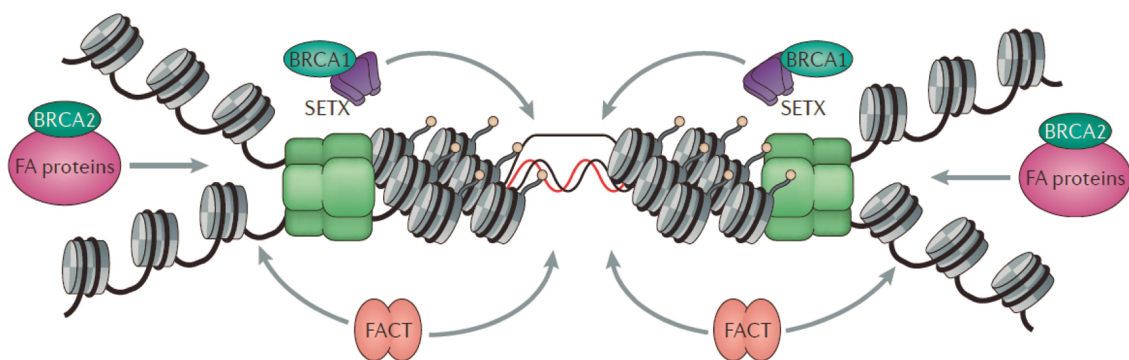
Figure I4. Factors that help resolve R-loop-mediated replication conflicts.
RF progression can be hampered by R loops and by local R-loop-mediated chromatin compaction. In this scenario, resolution of R loops can be facilitated by factors that protect stalled replication forks from collapse or that are required for RF progression through an obstacle such. This includes BRCA2 and potentially other Fanconi anemia (FA) proteins, the FACT (facilitates chromatin transcription) chromatin-reorganizing complex and probably senataxin (SETX) in cooperation with BRCA1. Figure adapted from (Santos-Pereira and Aguilera 2015).

4. THE THO/TREX COMPLEX: COUPLING CO-TRANSCRIPTIONAL mRNP BIOGENESIS WITH GENOME INTEGRITY

As part of the eukayotic gene expression process, nascent pre-mRNA needs to be processed and packaged into a stable and export-competent mRNP particle that is transported into the cytoplasm. For this, many RBPs are recruited co-transcriptionally to the nascent mRNA (Muller-McNicoll and Neugebauer 2013). This includes RBPs involved in processing (5′ end capping, splicing, 3′ end processing), nuclear export, subcellular localization, translation and stability of the mRNA. All these different processing steps and mRNP export are connected to each other and at the same time are coupled to transcription (Figure I5) (Bentley 2014). This makes the processing more efficient and allows its regulation at multiple steps, thus ensuring that only properly modified mRNPs are transported (Bjork and Wieslander 2017). The composition of the mRNP is dynamic, changing as processing and export takes place. The different steps of mRNP biogenesis are interconnected and frequently mediated by interaction with RNA polymerase II (RNAPII) (Figure I5). The carboxy-terminal domain (CTD) of the largest RNAPII subunit comprises a series of heptapeptide repeats with the consensus sequence YSPTSPS conserved from fungi to human. The CTD plays important roles in transcription and RNA processing, as a platform to recruit factors and modulate the efficiency of these processes (Buratowski 2009; Heidemann et al. 2013).
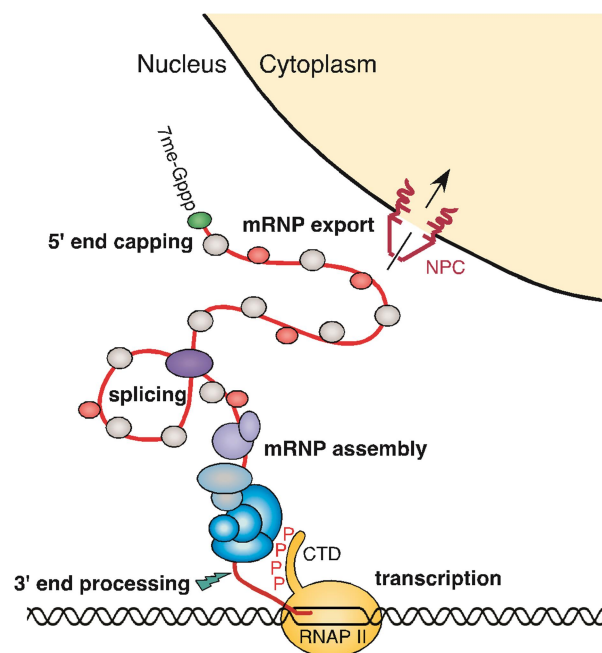
Figure I5. Nuclear steps of eukaryotic gene expression, from transcription to nuclear export.
NPC, Nuclear pore complex; CTD, C-terminal domain of Rpb1, RNAPII, RNA polymerase II. Figure adapted from (Luna et al. 2008).

4.1.   mRNP biogenesis: Coupling between transcription, mRNA processing and mRNA export

As soon as mRNA emerges from the RNAP, is capped by methyl-guanylation at its 5´ end. This is carried out by capping enzymes that binds to the phosphorylated carboxy-terminal domain (CTD) of the RNAPII largest subunit. Once the cap is added to the nascent mRNA, the nuclear cap-binding complex (CBC) binds co-transcriptionally to the monomethylated cap (Figure I5). The CBC binding to mRNA seems to be the first step in the assembly of the mRNP and required for the subsequent processes that occur on the mRNA molecule (splicing, transcription termination, export, nuclear mRNA decay, translation, non-sense-mediated decay and decapping) (Aguilera 2005; Luna et al. 2008; Gonatopoulos-Pournatzis and Cowling 2014).

Another step of mRNA processing is splicing. It consists in the elimination of introns from the pre mRNA and the junction of exons to produce a mature mRNA. Splicing factors are recruited rapidly to nascent mRNA and introns are removed co-transcriptionally, but also some of them can be eliminated after transcription (Kornblihtt et al. 2004). Splicing occurs in eukaryotes form yeast to human. However, while in mammals almost all RNAPII transcribed genes contain introns (except for histones and a few other genes), in *S. cerevisiae* introns are rather short and only present in a small subset of genes (for example in ribosomal protein genes) (Izquierdo and Valcarcel 2006; Kornblihtt et al. 2013; Shkreta and Chabot 2015).

Splicing is performed by the spliceosome, a complex ribonucleoprotein megaparticle that contains five uridine (U)-rich small nuclear ribonucleoproteins (snRNPs) (U1, U2, U4, U5 and U6) and more than 150 accessory proteins. The spliceosome recognizes and acts over very short consensus elements present in the nuclear pre-mRNA introns: The 5´ splice site (SS), the branch site (BS) and the 3´ SS; that, in metazoans, are also very poorly conserved (Figure I6A) (Wahl et al. 2009). The spliceosome, through its dynamic and sequential assembly, performs the two transesterification reactions that are necessary to excise introns and join together the selected exons (Figure I6B). Spliceosome assembly occurs by sequential interaction of the spliceosomal snRNPs with the numerous other splicing factors, and associated proteins as MFAP1, which is studied in more detail in this thesis (Figure I6C) (Will and Luhrmann 2011; Papasaikas and Valcarcel 2016).
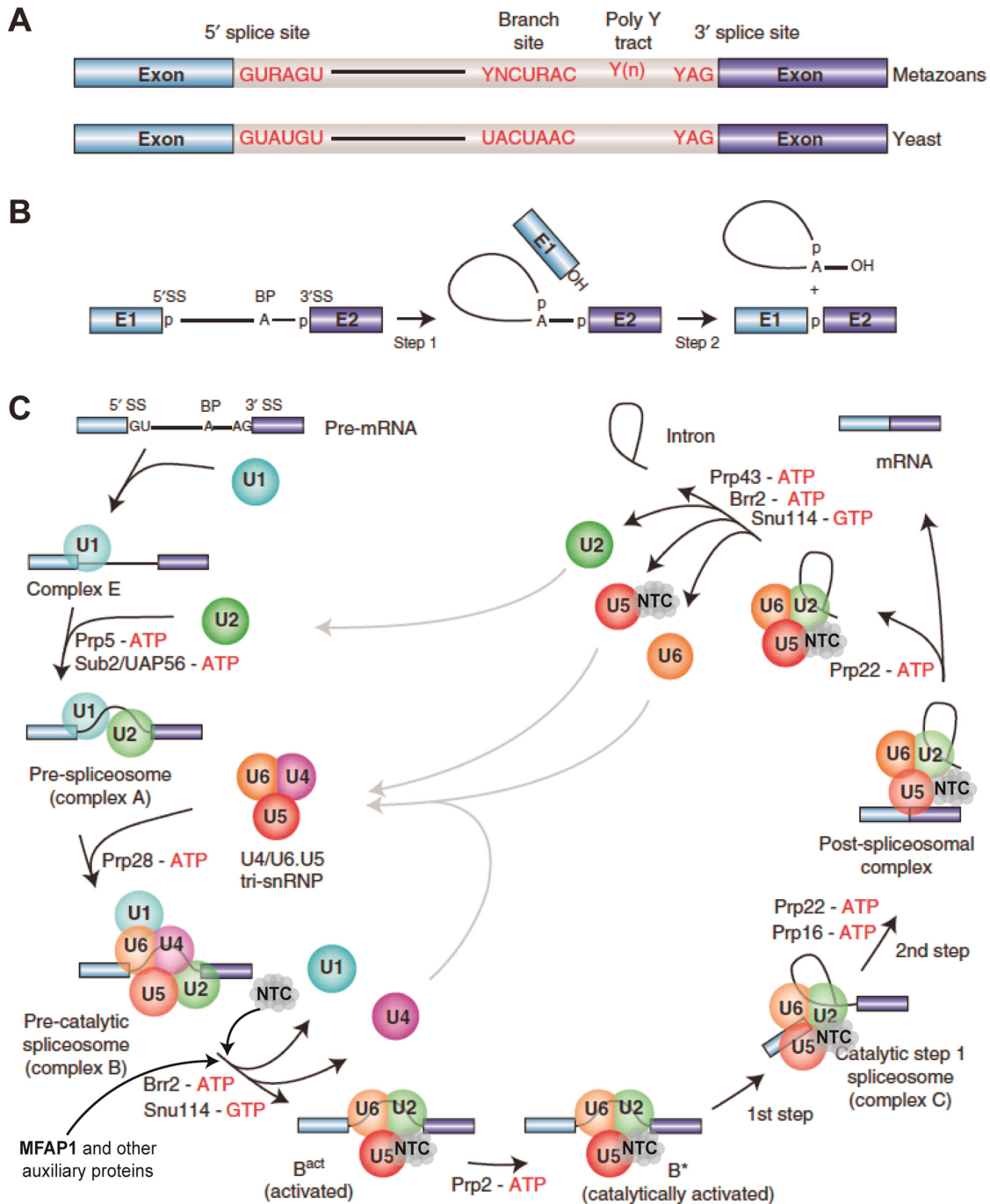
Figure I6. Pre-mRNA splicing and the spliceosome.
(A) Conserved sequences found at the 5′ and 3′ splice sites and branch site of pre-mRNA introns in metazoans and budding yeast (*S. cerevisiae*). The polypyrimidine tract is indicated by (Yn). **(B)** Splicing consists of two sequential transesterification reactions. In the first reaction, the 2′-hydroyl of an adenosine of the branch site (BS) in the intron attacks the phosphodiester bond at the 5′ splice site (SS). This generates a free 5′exon and an intron lariat-3′ exon. In the second reaction, the 3′-hydroxyl of the 5′ exon attacks the phosphodiester bond at the 3′ SS, leading to the ligation of the 5′and 3′ exons (forming the mRNA) and excision of the lariat intron (Wahl et al. 2009). **(C)** Scheme of the dynamic assembly of the spliceosome. The different U snRNPs and complexes (E, A, B and C) involved in the splicing process are shown. The participation of helicases and other associated proteins, such as MFAP1, is also indicated. NTC, NineTeen Complex. Figure adapted from (Will and Luhrmann 2011).

Depending on the usage of a splice site, two type of splicing can be distinguished, constitutive and alternative. In general, those splice sites that are more adjusted to the consensus sequence ('strong' splice site) are always efficiently recognized and spliced within a pre-mRNA (constitutive splicing). However, those slice sites whose sequence diverge far from the consensus sequence ('weak' splice sites) are only sometime recognized and used (alternative splicing) (Kornblihtt et al. 2013). Alternative splicing allows the production of multiple mRNA variants from a single pre-mRNA, which in turn is a major contributor to transcriptomic and proteomic diversity. This occurs especially in higher eukaryotes, where alternative splicing is predominant. The final splice site selection is influenced by regulatory factors, thus, members of the SR and hnRNP protein families often bind to enhancers and silencers sequences, respectively, leading to antagonistic effects on splice-site usage (Kornblihtt et al. 2013). Moreover, when splicing occurs co-transcriptionally transcription elongation rate is another factor that can be determinant for alternative splicing decisions (de la Mata et al. 2003). Changes in RNAPII elongation rates can affect the recruitment of splicing factors and the inclusion of exon with weak polypyrimidine tracts, and even the chromatin structure by histone modifications (Kornblihtt et al. 2013; Saldi et al. 2016). RNAPII pausing at the 3' end on yeast genes have been proposed to facilitate the co-transcriptional removal of introns, and vice versa, the splicing also could promote pausing in transcription (Alexander et al. 2010; Carrillo Oesterreich et al. 2010).

Another relevant step in gene expression is the mRNA 3′ end processing that consists on the cleavage of nascent transcript and acquisition of a poly(A) tail in most of them, which is essential for the stability, export to the cytoplasm and translation of the transcripts. This is carried out by a large cleavage/polyadenylation machinery that recognizes poly(A) signals on the nascent transcripts and produce the endonucleolytic 3′ end cleavage and the addition of a polyadenylated tail (Figure I7) (Millevoi and Vagner 2010). 3′ end formation occurs co-transcriptionally and influences reciprocally transcription and different steps of mRNA biogenesis. Indeed, 3′ end processing factors are recruited via CTD of RNAPII at the 3′ end of transcribed genes. Several transcription factors/activators affect processing at the poly(A) signal, in turn, 3′ end processing factors and sequence elements of the poly(A) signal modulate transcription termination. There is also a physical and functional interconnection with splicing. Several splicing-related factors interact with cleavage and polyadenylation factors.

Moreover, 3´ end processing contributes to recognize the last intron of a pre-mRNA. On the other hand, splicing factors bound to the last intron 3´ SS facilitates cleavage and polyadenylation (Millevoi and Vagner 2010; Bentley 2014). There is also a link between 3´ end processing and mRNA export (Rondon et al. 2010), thus mRNP particles that are not well processed are retained in the nucleus, and a surveillance mechanism controlled by the exosome and other factors (Schmid and Jensen 2008).
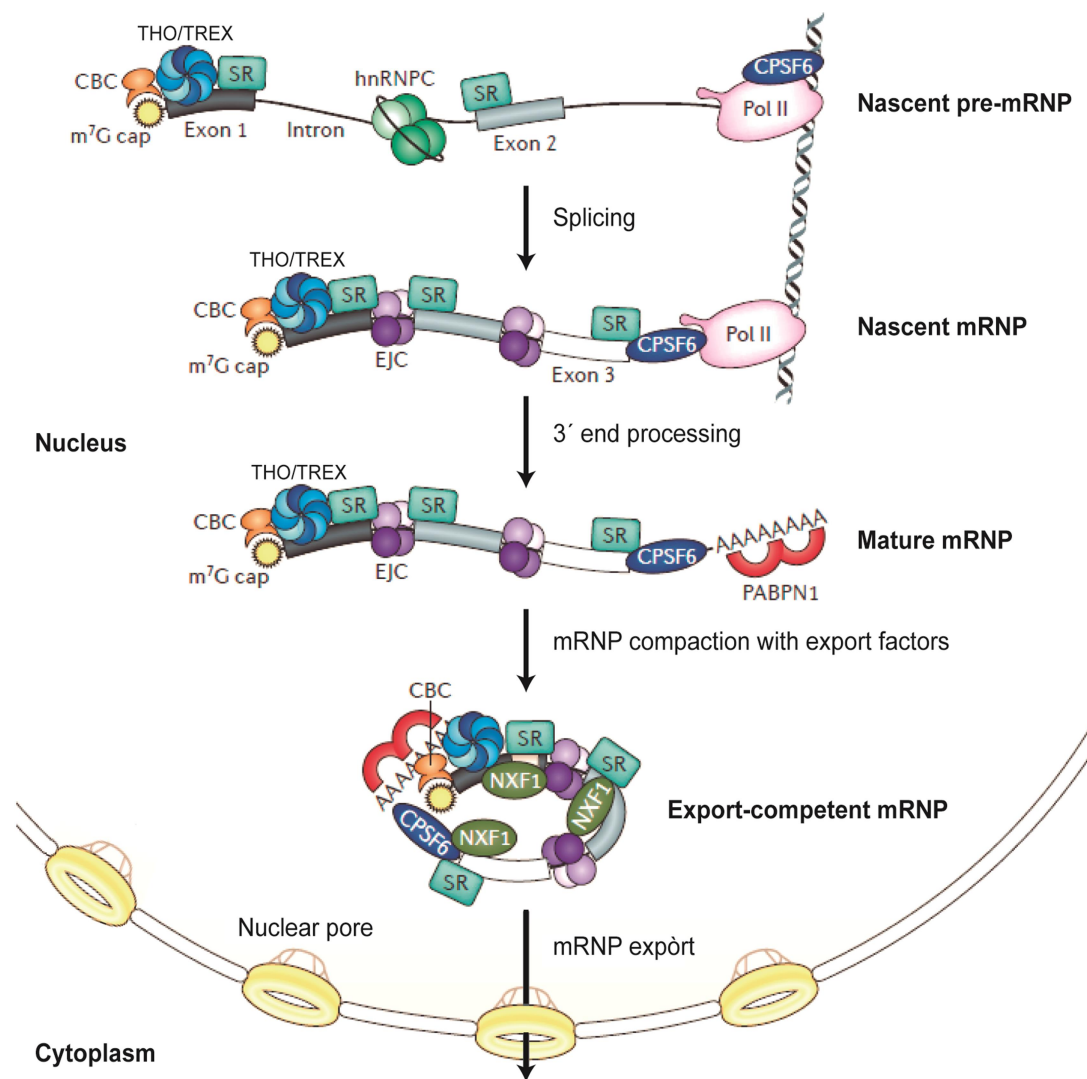
Figure I7. Co-transcriptional assembly of export-competent mRNPs.
Schematic representation of co-transcriptional assembly of messenger ribonucleoprotein particles (mRNPs) in mammalian cells. During transcription, the cap-binding complex (CBC) binds to the 7-methylguanosine (m^{7}G) cap of the 5´ end of nascent mRNA and participates in subsequent stages of mRNA biogenesis and export. The splicing and the binding of numerous RNA-binding proteins including THO/TREX also occurs co-transcriptionally. The recruitment of THO/TREX is mediated by interaction with proteins of the CBC, the exon junction complex (EJC) and serine/arginine serine/arginine-rich (SR) proteins and contributes to mRNP biogenesis and export. After cleavage and polyadenylation, THO/TREX and other export adaptor factors recruit the RNA export receptor NXF1 to allow efficient 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).

In addition to the mRNA processing proteins, many others RBPs bind co-transcriptionally to the nascent mRNA and contribute to the formation of the mRNP, protecting it from degradation and preparing it to be exported (Figure I7). Mature mRNPs are exported to the cytoplasm by the mRNA export receptor factor Mex67/NXF1 that interact with the nucleoporins that form the nuclear pore complex (Kohler and Hurt 2007). As this factor cannot bind directly the mRNA, the mRNP export take place through NXF1 interaction with mRNA adaptors, such SR proteins, THO/TREX and THSC/TREX2 complexes (Katahira 2012).

4.2. The THO/TREX complex as a key player of mRNP biogenesis

THO/TREX is a conserved eukaryotic complex that is required for mRNP biogenesis linking transcription to mRNA export, as well other steps of mRNP biogenesis. It was first isolated in yeast as a four subunit complex formed by Tho2, Hpr1, Mft1 and Thp2 (Figure I8) (Chavez et al. 2000). These four subunits appear in stoichiometric amounts and strongly interact between them forming a core. In fact, it can be purified under high salt conditions and it has been reported that THO is stable only when the four component are present (Chavez et al. 2000; Huertas et al. 2006). THO has also been purified in *Drosophila* and human cells and the complexes contain counterparts of the yeast subunits hHpr1/THOC1, hTho2/THOC2, as well as additional components such as THOC5, THOC6 and THOC7, which do not have apparent yeast homologs (Figure I8) (Rehwinkel et al. 2004; Masuda et al. 2005). THOC5-7 were originally identified as splicing associated factors and due to this were previously called fSAP79, fSAP35 and fSAP24, respectively. Moreover, THO associates with the mRNA export factors Sub2/UAP56 and Yra1/REF/ALY, and the Tex1 protein forming a larger complex termed TREX (TRanscription-EXport) (Figure I8) (Strasser et al. 2002). As in yeast, purification of the complex from *Drosophila* and human cells show that THO subunits form a stable core, while the interactions with UAP56 and ALY are weaker (Rehwinkel et al. 2004; Masuda et al. 2005). UAP56 is a DEAD-box helicase required for the recruitment of ALY to both spliced and intronless mRNAs, which in turn is necessary for subsequent steps of mRNA export (Luo et al. 2001; Taniguchi and Ohno 2008). ALY (hYra1 or REF) is an RNA binding protein that belong to the evolutionarily conserved REF (RNA and export factor binding proteins) family of hnRNP-like

proteins (Stutz et al. 2000). ALY stimulates nuclear mRNA export since acts as an adaptor for Mex67/NXF1 recruitment and binding to mRNPs (Zhou et al. 2000; Hautbergue et al. 2008; Taniguchi and Ohno 2008; Hung et al. 2010). The interaction of THO with UAP56, ALY and other mRNP factors facilitates the coupling between transcription and export. Human THO/TREX colocalizes with nuclear speckles which are domains enriched in pre-mRNA splicing factors located in the interchromatin regions of the nucleoplasm of mammalian cells (Zhou et al. 2000; Gatfield et al. 2001; Lamond and Spector 2003; Masuda et al. 2005). Recent evidence shows that splicing occurs in these nuclear speckle domains and that TREX functions in the release of spliced mRNA/polyA+ RNA from nuclear speckles for export to the cytoplasm (Dias et al. 2010).
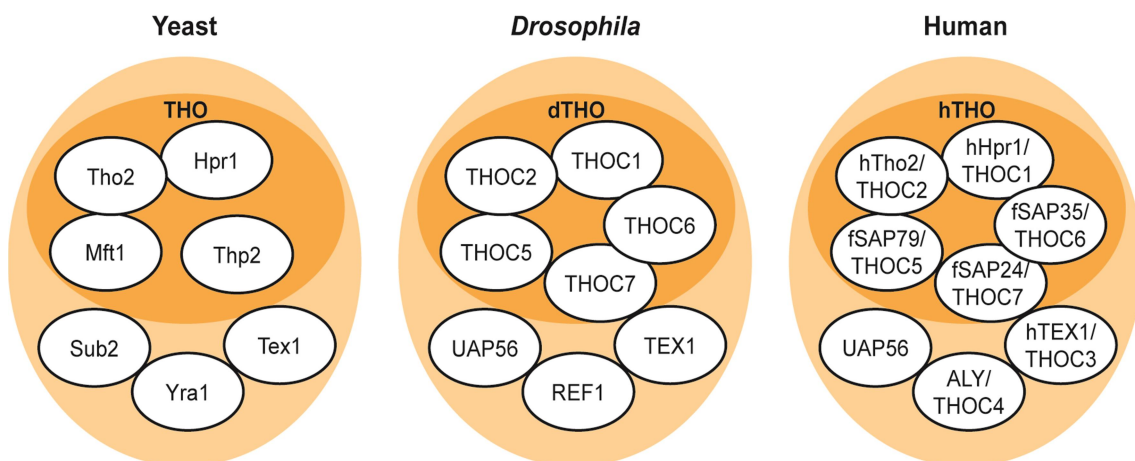
Figure I8. The THO/TREX complex is conserved from yeast to human.
The *Drosophila* and human THO complexes contain subunits homologous to Tho2 and Hpr1 (hTho2/THOC2 and hHpr1/THOC1, respectively) and three additional subunits THOC5/fSAP79, THOC6/ fSAP35 and THOC7/ fSAP24. THO also associates with substoichiometric amounts of the homologue to Tex1, Yra1 (REF/ALY) and Sub2 (UAP56), forming TREX. Figure adapted from (Reed and Cheng 2005).

The close connection between transcription and mRNP biogenesis can explain the pleiotropic phenotypes observed in THO mutants including, transcription impairment, hyper-recombination, mRNA export defects, increased levels of exosome-dependent mRNA instability, and defects in 3′ mRNA formation (Figure I9) (Chavez et al. 2000; Strasser et al. 2002; Rondon et al. 2003; Rougemaille et al. 2008; Dominguez-Sanchez et al. 2011a). *In vivo* and *in vitro* assays in yeast showed that THO/TREX mutants have transcription elongation and mRNA export defects (Figure

I9) (Chavez and Aguilera 1997; Piruat and Aguilera 1998; Chavez et al. 2000; Rondon et al. 2003). Genome-wide analysis in THO/TREX mutants shows that expression of long, G+C-rich and highly transcribed genes are affected, according the role of this complex in transcription elongation (Gomez-Gonzalez et al. 2011).

THO/TREX is recruited to the active transcribed chromatin. Genome-wide profile in yeast show that THO levels increases towards the 3′ end of genes in a gradient manner to disappear after the polyadenylation signal (Gomez-Gonzalez et al. 2011). This pattern is in agreement with the role of THO in transcription elongation and in 3′ end formation and the assembly of an export competent mRNP (Luna et al. 2012). THO can interact with the RNA molecule and with RNAPII (Jimeno et al. 2002; Meinel et al. 2013); but interaction with other mRNP components also seems to help recruit THO to the nascent mRNP (Figure I7). It has been suggested that both exon junction complex (EJC) and cap-binding complex (CBC), through the CBP80-ALY interaction, promote the binding of TREX to the 5′ end of mRNAs (Cheng et al. 2006). This connection between co transcriptional splicing and TREX is also established through the conserved Prp19 complex (Prp19C), also known as NineTeen Complex (NTC). This complex functions in transcription elongation and also is crucial for catalytic activation of the spliceosome and subsequent splicing reactions and rearrangement within the spliceosome, among other functions (Chanarat and Strasser 2013). In yeast, NTC seems to be necessary to stabilize the recruitment of TREX to transcribed genes (Chanarat et al. 2011). In addition, subunits of TREX have been shown to be associated with 3′ end processing factors in yeast and human cells (Johnson et al. 2011; Katahira et al. 2013). Beside the physical interaction, functional connection is also reported as depletion of THO leads to mRNA 3′ end formation defects and vice versa, mutation in 3′ end processing factors have transcription elongation defects (Luna et al. 2005; Rougemaille et al. 2008).

Despite the differences in composition and in the recruitment of the complex in yeast and human cells, THO/TREX seems to play a similar role in transcription, mRNP biogenesis/export and in the maintenance of genome integrity from yeast to human (Durfee et al. 1994; Guo et al. 2005; Li et al. 2005; Masuda et al. 2005; Dominguez-Sanchez et al. 2011a; Viphakone et al. 2012).

The current view in the field is that THO could participate in the co-transcriptional formation of export-competent mRNP during transcription elongation by controlling the assembly of heterogeneous nuclear ribonucleoproteins (hnRNPs) onto the mRNA and avoiding the formation of suboptimal mRNPs. The observation that overexpression of the RNA-dependent ATPase Sub2 or the novel RNA binding protein Tho1 suppresses both the transcription and RNA-export defects of THO mutants is consistent with this view (Fan et al. 2001; Jimeno et al. 2002; Jimeno et al. 2006).
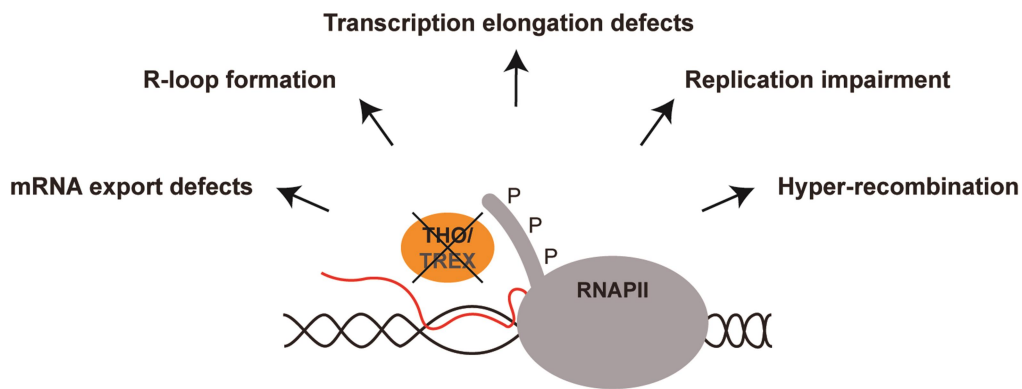
Figure I9. THO/TREX plays a role at the interface of transcription and genome instability.
Lack of THO/TREX leads to transcription elongation and mRNA export defects, R-loop formation, replication impairment and hyper-recombination.

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

The analyses of THO mutants in *S. cerevisiae* provided the first evidence of the relevant role of mRNP biogenesis in preventing co-transcriptional genome instability. In general, yeast THO/TREX mutants show mRNA export defects, transcription-elongation impairment and transcription-dependent hyper-recombination phenotypes (Aguilera and Klein 1990; Piruat and Aguilera 1998; Chavez et al. 2000; Jimeno et al. 2002). This is mainly attributed to an improper mRNP formation. Yeast cell lacking a functional THO/TREX produce a suboptimal mRNP which is able to hinder transcription elongation and generate genomic instability measured as an increase in recombination. This in turn can be largely explained by an increase in R-loop formation (Huertas and Aguilera 2003). A suboptimal mRNP particle in THO mutans increase the chance of nascent mRNA to hybridize back with the DNA template during transcription. Indeed, it

has been shown that there is an increase in RNA-DNA hybrids in THO-depleted cells, as determined by genetic approaches and by physical detection of RNA-DNA hybrids with molecular tools using specific antibodies (Huertas and Aguilera 2003; Bhatia et al. 2014). These R loops structures could be the responsible for the replication impairment observed in THO mutant, as determined by 2D gel and Rrm3 helicase ChIP-chip analyses (Wellinger et al. 2006; Gomez-Gonzalez et al. 2011). Moreover, this phenomenon seems to be conserved since similar phenotypes are observed in *C. elegans* and human cells depleted of THO (Dominguez-Sanchez et al. 2011a; Castellano-Pozo et al. 2012). hTHO depletion in different cell lines increases DNA breaks, as measure by single-cell electrophoresis, γH2AX and 53BP1 foci; enhances class-switching recombination and causes replication impairment, as determined by DNA combing (Dominguez-Sanchez et al. 2011a). Replication impairment has also been observed in THO mutants in *C. elegans*, by incorporation of labeled deoxyribonucleotides (Castellano-Pozo et al. 2012). In all these organism data support that increased DNA damage, hyper-recombination, transcription elongation defects and replication impairment are mediated by R-loop formation since all these phenotypes are suppressed by RNaseH and/or enhances by AID overexpression (Figure I9) (Dominguez-Sanchez et al. 2011a; Gomez-Gonzalez et al. 2011; Castellano-Pozo et al. 2012). Recently, a connection between R loops and chromatin has been shown since lack of THO accumulates high levels of H3S10-P in an R-loop-dependent manner in yeast, *C. elegans* and human cells (Castellano-Pozo et al. 2013).

Importantly, later studies have shown that this connection between transcription, mRNP biogenesis/export and genome instability is not only established exclusively through THO/TREX but also extends to many other factors involved in different nuclear mRNA processes. These include proteins involved in splicing, as is the case of SRSF1 (serine/arginine splicing factor 1) (Li and Manley 2005); in mRNA 3´ end processing and degradation (Luna et al. 2005; Stirling et al. 2012; Gavalda et al. 2013; Pefanis et al. 2015); mRNP biogenesis and export such as THSC/TREX2 among others (Jimeno et al. 2002; Gallardo et al. 2003; Gonzalez-Aguilera et al. 2008); helicases such as SETX/Sen1, DDX19, DDX23 and others factors (Paulsen et al. 2009; Skourti-Stathaki et al. 2011; Wahba et al. 2011; Santos-Pereira and Aguilera 2015; Hodroj et al. 2017; Sridhara et al. 2017).

4.4. Other roles of THO/TREX complex

THOC1 (hHpr1 or p84N5), one of the conserved subunits of THO, was originally identified as a nuclear matrix protein involved in RNA metabolism that interacts with the tumor suppressor retinoblastoma protein (pRB) (Durfee et al. 1994). Different reports indicate that THO plays a role in cell proliferation and apoptosis, which can be mediated by the association of THOC1 with pRB, and the death domain of this subunit (Doostzadeh-Cizeron et al. 1999) (reviewed in (Heath et al. 2016)). Interestingly, THOC1 has been reported that is de-regulated in several cancer types (Guo et al. 2005; Li et al. 2007; Dominguez-Sanchez et al. 2011b; Guo et al. 2012).

Moreover, THO complex can play a role in cell differentiation. Although THO is not essential for the viability of unicellular yeasts, the orthologous Thoc1 protein is required for viability of the early mouse embryo (Wang et al. 2006). This complex is also involved in development of adult cells, as determined by defects in spermatogenesis of hypomorfic alleles of Thoc1 mice (Wang et al. 2009a) and in bone marrow and hematopoiesis differentiation, as shown in conditional THOC5 knock-down mice (Mancini et al. 2010). These reports suggest that THO/TREX could influence different aspects of differentiation and proliferation probably by either affecting transcription or the export of specific mRNAs (Tamura et al. 1999; Luna et al. 2012).

Recently, a link between THO complex and syndromic intellectual disability has been also shown as mutations of different THO subunits has been related to this neurodevelopmental disorder (Chi et al. 2013; Kumar et al. 2015).

THO/TREX may also have a similar role in mRNP biogenesis and export or different functions in other organisms since this complex is not only present in yeast, vertebrates and mammals but also it has been found in plants and in the protozoa *Trypanosoma* (Furumizu et al. 2010; Serpeloni et al. 2011). In *Arabidopsis*, THO/TREX has been shown to be involved in mRNA export and alternative splicing and other functions such as transgene and endogenous siRNA and microRNA biosynthesis (Yelina et al. 2010; Francisco-Mangilet et al. 2015; Sorensen et al. 2017). Moreover, this complex has been shown to participate in the nuclear export of viral mRNA transcript (Koffa et al. 2001; Lischka et al. 2006).

4.5. Human THO/TREX-associated partners

During mRNP biogenesis, THO interacts with a variety of proteins that either contribute to THO binding to the pre-mRNA or use THO as a landing platform and subsequently binding to nascent pre-mRNA. In addition to UAP56 and ALY, several proteins have been found to be associated with TREX during the last years. Most of them are proteins as RNA helicases (UAP56 and its paralog URH49), hnRNPs as CIP29, and other proteins that acts as mRNA export adaptors among others (Luna et al. 2012; Heath et al. 2016). CIP29 (also known as SARNP) has been purified together with THO (Dufu et al. 2010). It is the human homologue of yeast Tho1, an hnRNP that, as well as Sub2, was identified as a multicopy suppressor of THO mutants defects in transcription and mRNA export (Piruat and Aguilera 1998; Jimeno et al. 2006). Although CIP29 may interact with both UAP56 and URH49, it has been shown that UAP56 is preferentially associated with ALY to form TREX, whereas URH49 is preferentially associated with CIP29 to form the AREX complex (alternative mRNA export complex) (Yamazaki et al. 2010). These are just some examples among the new putative TREX subunits or associated proteins identified over the last few years. This suggests that mRNP biogenesis is a dynamic process in which THO is a key component that acts as a platform that connects different steps of mRNA metabolism (Luna et al. 2012; Heath et al. 2016).

Deciphering the nature and relevance of the new physical and genetic interactions of THO would help define a group of factors working at the same nuclear process that may be the key to understand the mechanism of mRNP biogenesis and its coupling with genome instability.

During the course of this Thesis we have identified a previously unknown link between THO/TREX and the histone deacetylase Sin3A complex. A brief introduction to this complex is included in Introduction section 5.

5. THE HUMAN SIN3A HISTONE DEACETYLASE COMPLEX

Eukaryotic gene expression is a highly orchestrated process that involves coordinated activation and silencing of genes mediated by complex interactions between the chromatin and transcription regulatory proteins. The Sin3 co-repressor complex is a highly conserved multiprotein complex that mediates transcriptional repression via histone deacetylation (Grzenda et al. 2009). In general, acetylation of histone tails promotes a more open chromatin conformation and correlates with active transcription whereas deacetylation of histone tails causes a more closed/compacted chromatin and correlates with transcriptional repression (Shahbazian and Grunstein 2007). SIN3 protein through its interaction with DNA-binding factors, acts as a scaffold protein that recruits histone deacetylases (Class I HDACs; HDAC1 and HDAC2) and other chromatin-modifying enzymes onto the target genes. Thus, the core Sin3-HDAC complex interacts with a wide variety of repressors and corepressors, providing flexibility and expanded specificity in modulating chromatin structure and transcription. As result, Sin3-HDAC corepressor complex functions in transcriptional regulation of several genes and is therefore implicated in the regulation of key biological processes, such as cellular proliferation, differentiation, apoptosis as well oncogenic transformation (Kadamb et al. 2013). Although a repressive role for Sin3-HDAC in regulating gene expression is well established, accumulating evidence also points to a role in gene activation.

Sin3/HDAC complex is conserved from yeast to humans, and the presence of several paralogs and alternatively spliced isoforms of SIN3 in different organisms potentially contribute to the complexity and diversity of SIN3 functions. In mammalian cells, the Sin3A complex consists of eight core subunits, SIN3, HDAC1, HDAC2, RBBP4, RBBP7, SUDS3, SAP30 and SAP18 and several additional subunits, including RPB1, SAP180, SAP130, ING1/2, BRMS1 and SAP25 (Table I1) (Grzenda et al. 2009).

Table I1. The mammalian Sin3A complex.

Mammalian mSin3A complex		
Subunit	**Function**	
SIN3A	Scaffold protein	Core components
HDAC1 / HDAC2	Catalytic subunits	
RBBP4 / RBBP7	Stabilization with the nucleosome	
SUDS3	Integrity and activity	
SAP30	Stability and bridge with co-repressors	
SAP18	Stabilization and activity	
RBP1, SAP180, SAP130, ING1/2, BRMS1, SAP25	Non-core components. Recruitment of the complex to chromatin and interaction with other proteins	

SIN3 protein is a key piece of the complex since it mediates interactions with DNA-binding factors and acts as a platform for the rest of the complex and other chromatin modifiers. To establish all these interactions SIN3 possesses six highly conserved regions including four Paired Amphipathic Helices (PAH1-4), one Histone Deacetylase Interaction Domain (HID) and one Highly Conserved Region (HCR) (Silverstein and Ekwall 2005). In mammals, there are two isoforms, SIN3A and SIN3B, encoded by two separate genes (Ayer et al. 1995). Both isoforms are approximately 57% identical in their amino acid sequence with the highest regions of homology localized in the PAH and HID regions. The most pronounced difference between them is that SIN3B possesses has a shorter amino-terminal region. Although both SIN3A and SIN3B may be found expressed within the same cells and tissues and have common functions and interactors, they have also been found to associate with different proteins and affect the regulation of different target genes (van Oevelen et al. 2008; van Oevelen et al. 2010; Jelinic et al. 2011; Kadamb et al. 2013; Lewis et al. 2016).

Histone deacetylases 1 and 2 (HDAC1/2) belong to the class I HDAC family (Grozinger and Schreiber 2002). HDAC1/2 constitute the major catalytic subunits of the complex and are involved in transcriptional silencing in the nucleus. The retinoblastoma-binding proteins 4 and 7 (RBBP4/7, formerly RbAp48/46) contribute to the physical stabilization of the contact between the complex and the nucleosome. Suppressor of defective silencing 3 (SUDS3) (also known as SDS3) is required for the integrity and activity of the complex. The Sin3-associated protein (SAPs) SAP30 has a role in stabilizing the complex and may act as a bridge between the Sin3A complex and

other co-repressors. SAP18 has also a role in the stabilization of the complex, especially the HDAC1-SIN3 interaction, and may enhance the enzymatic activity as well as it mediates the interaction with other proteins (Silverstein and Ekwall 2005; Grzenda et al. 2009). In addition to the core components, several other proteins form part of the complex, among them SAP130 a subunit that has been identified in this Thesis as a new THO interactor (Table I1).

Therefore, SIN3 is recruited to chromatin by DNA-binding proteins and acts as a platform. SIN3 and the rest of the different components of the complex provide structural stability and expand the action of the complex by establishing interactions with others DNA-binding transcriptional repressors and other transcriptional regulatory proteins (Kadamb et al. 2013).

In addition to transcription regulation, Sin3 complex also mediates post-translational stabilization of several non-histone proteins (Zilfou et al. 2001; Icardi et al. 2012), formation of heterochromatin and chromosome segregation and rDNA silencing (Grzenda et al. 2009). In *S. cerevisiae*, Sin3 complex has been shown to be important in other processes as replication timing, since it avoids early activation of late origins (Aparicio et al. 2004); and in DNA repair, since is recruited to create a local region of hypoacetylation necessary for double-strand break repair by NHEJ (Jazayeri et al. 2004).

In conclusion, Sin3A is involved in the regulation of key biological processes mainly due to its histone deacetylase activity and the transcriptional regulation of a wide variety of genes (Silverstein and Ekwall 2005; Kadamb et al. 2013). In this thesis we will describe the physical and functional interaction of this complex with the THO mRNP complex and its biological relevance on the maintenance of genome integrity.

OBJECTIVES

The main goal of this thesis is to further explore how mRNP biogenesis contributes to prevent genome instability. For this purpose, we addressed the following specific objectives:

1. To identify new partners of human THO/TREX.

2. To analyze the role of the newly identified partners in the maintenance of genome integrity and gene expression, with the aim of getting new insights into the molecular mechanisms of THO/TREX in preventing genome instability.

RESULTS

1. IDENTIFICATION OF NEW THO/TREX-INTERACTING FACTORS

To get further insight into the connection between mRNP biogenesis and genome dynamics, we decided to perform two-hybrid screenings using a human cDNA library to look for new partners of human THO/TREX that could help us to better understand how this complex contributes to the maintenance of genome stability.

1.1. Identification of putative THO/TREX interactors by two-hybrid screening

We searched for factors interacting with THO/TREX by the yeast two-hybrid system using different subunits as bait: THOC1, UAP56, URH49 and ALY. THOC1 is a core subunit of the THO complex, UAP56 and URH49 are two closed related RNA helicases and ALY is an mRNA export adaptor protein (Luna et al. 2012). In order to perform the screening, the cDNAs of the selected subunits were cloned into pGAL4BD bait vector fused to the DNA-binding domain of the transcriptional factor Gal4 (pGAL4BD-*BAIT*) (Figure R1A). As a result, bait protein is expressed under the control of *ADH1* promoter and fused with Gal4 DNA-BD and c-Myc epitope tag. Bait constructions were checked by sequencing and the expected protein molecular weight (kDa) was confirmed by western blot with c-Myc antibody (Figure R1B).
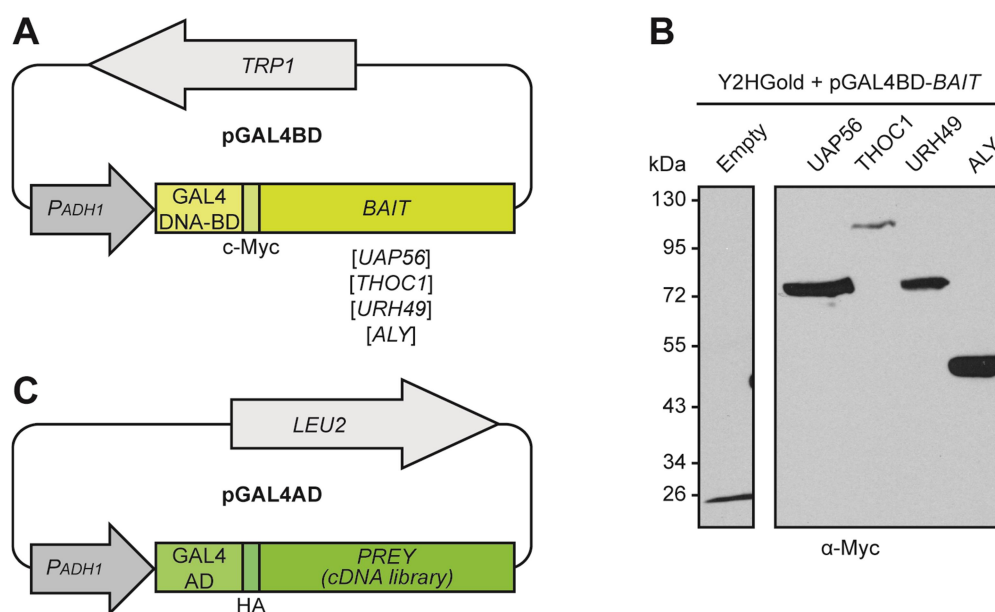
Figure R1. Yeast two-hybrid vectors and expression of bait fusion proteins.
(A) Scheme of the pGAL4BD bait vector. This vector allows the expression of the Bait protein fussed with the Gal4 DNA-binding domain (DNA-BD) and c-Myc epitope tag under the control of *ADH1* promoter. It contains the *TRP1* gene as the selection marker in yeast. **(B)** Western blot analysis of bait proteins (UAP56, THOC1, URH49 and ALY) fussed with the Gal4 DNA-BD and c-Myc epitope tag. Fusion bait proteins were detected using c-Myc antibody. **(C)** Scheme of the pGAL4AD prey vector. This vector allows the expression of the human cDNA library (prey) fussed with the Gal4-activation domain (AD) and HA epitope tag under the control of *ADH1* promoter. Prey vector contains the *LEU2* gene as the selection marker in yeast.

Prey proteins were expressed from a universal and normalized human cDNA library cloned into pGAL4AD prey vector fused to that carries the Gal4 activation domain and HA epitope tag (pGAL4AD-cDNA library) (Figure R1C). We transformed yeast strain Y2HGold with each vector pGALBD-*BAIT* and after mating with the Y187 strain transformed with the cDNA library positive interactions were selected plating the culture onto high stringency medium (see Materials and Methods 6.1). Prior to perform the two-hybrid screening it was tested that bait proteins were not able to autonomously activate the reporter genes (no growth was observed when Y2HGold strain co-transformed with pGALBD-*BAIT* and empty pGAL4AD was plated on restrictive medium).

More than 2×10^6 clones were screened with each selected bait (Table R1). Total number of clones was estimated according to the number of diploids formed in the mated cultured (see Materials and Methods 6.1). All clones that grew and showed blue staining onto this high stringency medium were analyzed (Figure R2). Library plasmids responsible for activation of reporters (positive clones) were rescued and positive interactions were confirmed with different yeast transformants. Positive clones showed growth and blue staining in selective medium when bait was present while no growth was observed in its absence (empty plasmid pGAL4BD) (Figure R2). Finally, cDNA inserts of the candidates were sequenced. cDNA sequence analysis allowed the identification of the proteins that interact with the subunits of THO/TREX tested by two-hybrid system and discard those cDNAs that were not cloned in frame, and therefore were false positives. Nineteen proteins were identified as new putative partners of THOC1 (see Table R2). In the screenings with UAP56 and URH49 we found several clones encoding proteins that have been previously reported to interact with these RNA helicases. Six clones encoding URH49 were found using UAP56 as bait, and four clones encoding the heterogeneous ribonucleoprotein hnRNP CIP29 were isolated with URH49 as bait (Leaw et al. 2004; Lehner et al. 2004; Dufu et al. 2010; Yamazaki et al. 2010). In the case of the screening with ALY, one clone encoding ZBTB44 was found as a new putative interactor. Number of clones and information about the encoded protein fragment isolated from the screening is shown in Table R2.

Table R1. Number of clones screened with each bait.

Bait	Clones screened	Number of positive clones
THOC1	6.96×10^6	28
UAP56	2.7×10^6	6
URH49	4.94×10^6	4
ALY	2.15×10^6	1
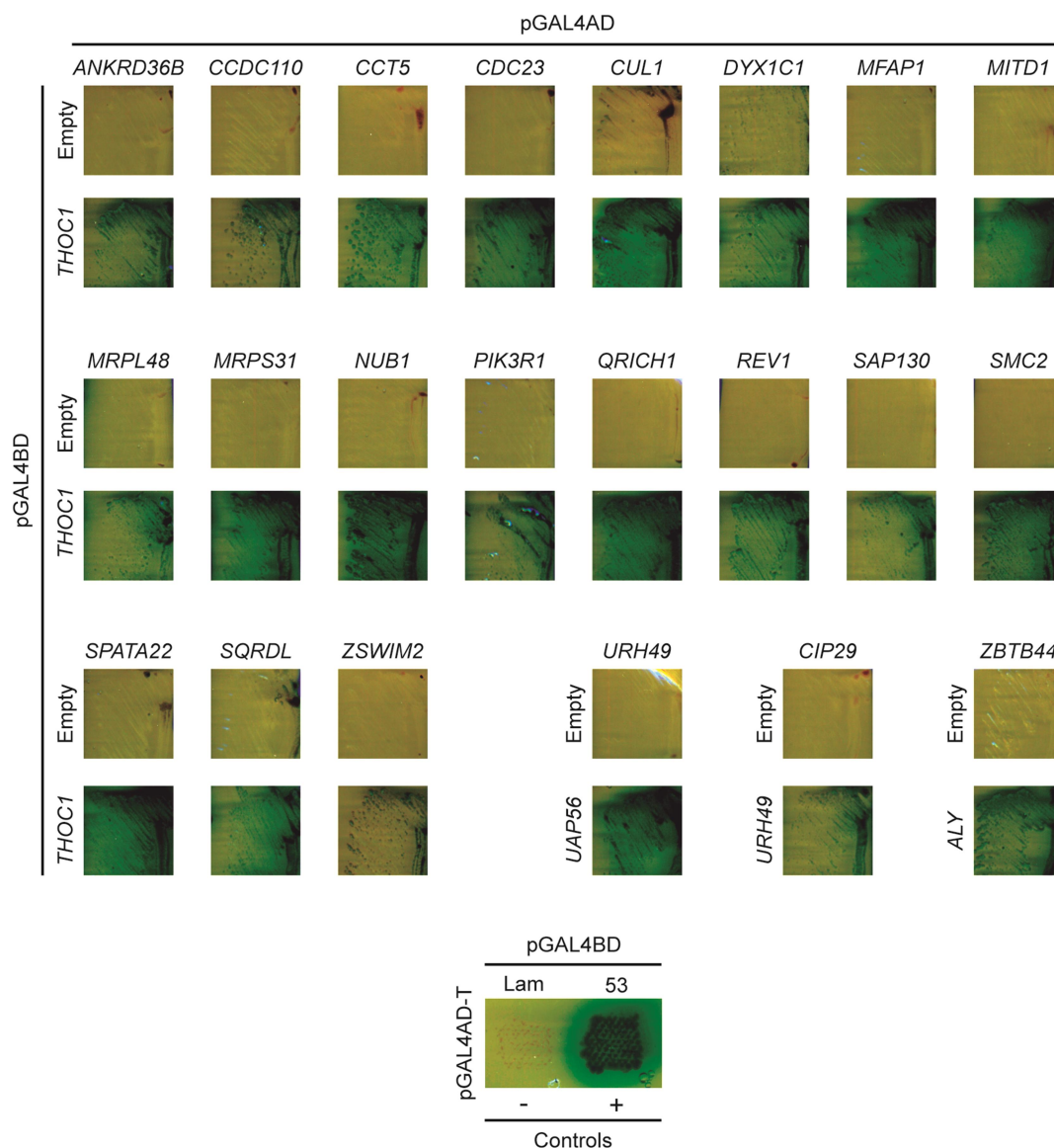
Figure R2. Identification of human THOC1, UAP56, URH49 and ALY partners by Two-Hybrid system.
Positive clones isolated from a human cDNA library as interactors of THOC1, UAP56, URH59 and ALY. The two-hybrid interaction of bait proteins with those encoded by the indicated human cDNAs allows growth in high stringency medium (SC-Ade-His-Trp-Leu+(X-α-Gal)+AbA). The interaction is also observed by enzymatic activity (blue colonies). No growth is observed when yeast cells are co-transformed with pGAL4AD-*cDNA* and empty pGAL4BD-. Positive and negative controls are shown at the bottom of the figure. For more details see Materials and Methods 6.1 and Table M6.

Table R2. Description of the interactors identified in the screening.

Protein	Description	Number of isolated clones	Accession	Protein fragment isolated (related to full length). (aa.=amino acids)	Domains and motifs	Cellular localization (N, nucleus; C, cytosol)	References
Y2H screening using THOC1 as bait							
ANKRD36B	Ankyrin Repeat Domain 36B.	1	Q8N2N9	971-1353/1353 aa.	CC domain	-	(Li et al. 2006)
CCDC110	Coiled-Coil Domain Containing 110	1	Q8TBZ0	322-690/833 aa.	CC domain	N,C	(Monji et al. 2004)
CCT5	Chaperonin Containing TCP1, Subunit 5 (Epsilon)	1	P48643	120-423/540 aa.	-	N,C	(Spiess et al. 2004)
CDC23*	Cell Division Cycle 23. Component of anaphase-promoting complex (APC).	1	Q9UJX2	1-500/596 aa.	TPR domain	N	(Nakayama and Nakayama 2006)
CUL1*	Cullin 1. Role in protein ubiquitination and degradation.	2	Q13616	624-776/776 aa. 639-776/776 aa.	-	N,C	(Nakayama and Nakayama 2006)
DYX1C1	Dyslexia Susceptibility 1 Candidate 1. Interacts with estrogen receptors and heat shock proteins.	1	Q8WXU2	1-320/420 aa.	- CS domain - TPR repeats - Region that mediates interaction with ESR1 and STUB1	N,C	(Taipale et al. 2003)
MFAP1*	Microfibrillar Associated Protein 1. Pre-mRNA splicing factor. Spliceosome.	1	P55081	289-438/438 aa.	-	N	(Hegele et al. 2012)

Protein	Description	Number of isolated clones	Accession	Protein fragment isolated (related to full length). (aa.=amino acids)	Domains and motifs	Cellular localization (N, nucleus; C, cytosol)	References
Y2H screening using THOC1 as bait								
MITD1	Microtubule Interacting And Trafficking Domain Containing 1. Required for proper cytokinesis.	4	Q8WV92	1-249/249 aa.	- MIT domain - Region for association with membranes	C	(Lee et al. 2012)
MRPL48	Mitochondrial Ribosomal Protein L48. Protein synthesis within the mitochondrion.	1	Q96GC5	1-212/212 aa.	-	C	(Chen 2015)
MRPS31	Mitochondrial Ribosomal Protein S31. Protein synthesis within the mitochondrion.	1	Q92665	44-395/395	-	C	(Chen 2015)
NUB1*	Negative Regulator Of Ubiquitin-Like Proteins 1. Regulation of NEDD8 degradation.	1	Q9Y5A7	82-466/615 aa.	- CC domain - UBA domain - Nuclear localization signal - NEDD8-binding	N	(Kito et al. 2001)
PIK3R1	Phosphoinositide-3-Kinase Regulatory Subunit 1. Role in the metabolic actions of insulin.	1	P27986	329-723/723 aa.	- SH2 domain	C	(Hansen et al. 1997)
QRICH1	Glutamine Rich 1. Inflammation and apoptosis.	1	Q2TAL8	652-776/776 aa.	-	N,C	(Jang et al. 2015)

Protein	Description	Number of isolated clones	Accession	Protein fragment isolated (related to full length). (aa.=amino acids)	Domains and motifs	Cellular localization (N, nucleus; C, cytosol)	References
Y2H screening using THOC1 as bait								
REV1*	Translesion DNA synthesis (Y-family polymerase).	1	Q9UBZ9	1-433/1251 aa.	- BRCT domain - Interaction with target DNA - dCTP binding	N	(Nelson et al. 1996)
SAP130*	Sin3A Associated Protein 130kDa Transcriptional repression (Sin3A histone deacetylation complex).	1	Q9H0E3	903-1083/1083 aa.	Interaction with SIN3A and HDAC1	N	(Fleischer et al. 2003)
SMC2*	Structural Maintenance Of Chromosomes 2. Condensin. Chromosome condensation.	1	O95347	867-1197/1197 aa.	- CC domain - Ala/Asp-rich (DA-box)	N,C	(Losada and Hirano 2005)
SPATA22*	Spermatogenesis Associated 22. Gametogenesis.	6	Q8NHS9	2-363/363 aa. 2-226/363 aa. 2-157/363 aa. 58-363/363 aa. 130-363/363 aa.	-	N	(La Salle et al. 2012)
SQRDL	Sulfide Quinone Reductase-Like (Yeast). Mitochondria metabolism.	1	Q9Y6N5	335-450/450 aa.	FAD nucleotide-binding motif	C	(Jackson et al. 2012)
ZSWIM2	Zinc Finger SWIM-Type Containing 2. E3 ubiquitin-protein ligase involved in apoptosis.	1	Q8NEG5	286-633/633 aa.	RING-finger domain	C	(Nishito et al. 2006)

Protein	Description	Number of isolated clones	Accession	Protein fragment isolated (related to full length). (aa.=amino acids)	Domains and motifs	Cellular localization (N, nucleus; C, cytosol)	References
Y2H screening using UAP56 as bait							
URH49	DEAD-Box Helicase 39A. RNA helicase.	6	O00148	8-426/426 aa.	- Helicase ATP-binding - Helicase C-terminal - Q motif - DECD box	N	(Lehner et al. 2004; Yamazaki et al. 2010)
Y2H screening using URH49 as bait							
CIP29/THO1	Cytokine Induced Protein 29 KDa. Component of the TREX complex.	4	P82979	1-209/209 aa.	SAP domain	N	(Yamazaki et al. 2010)
Y2H screening using ALY as bait							
ZBTB44[*]	Zinc Finger And BTB Domain Containing 44. Unknown. May be involved in transcriptional regulation because of its BTB domain.	1	Q8NCP5	215-570/570 aa.	C2H2-type 1-4 (zinc finger)	N	(Beaulieu and Sant'Angelo 2011; Lee and Maeda 2012)

* Prey proteins selected for validation of the association with the bait protein in human cells

Abbreviations: CC, Coiled-Coil; TRP, Tetratricopeptide; CS, CHORD-containing proteins and SGT1; ESR1, Estrogen Receptor 1; STUB1, STIP1 homology and U-box containing; MIT, Microtubule Interacting and Trafficking; UBA, Ubiquitin-Associated; SH2, Src homology2; BRCT, BRCA1 C-terminal; dCTP, Deoxycytidine triphosphate; HDAC1, Histone Deacetylase 1; FAD, Flavin Adenine Dinucleotide; RING, Really Interesting New Gene; Q motif, GFXXPXPIQ motif with invariant glutamine (Q); DECD, variant of DEAD (Asp-Glu-Ala-Asp) box; SAP, Scaffold-Associated Protein.

1.2. Validation of protein interactions in human cells

In order to analyze whether the novel interactions identified from the two-hybrid screening in yeast also take place *in vivo* in human cells we performed co-immunoprecipitation assays. We focus our analysis on those candidates (for THOC1 and ALY) that act in the nucleus, play a role that could affect genome stability and whose interaction with the bait protein has not been previously reported (see Table R2). Full-length cDNAs from IMAGE clones of prey candidates were cloned into pEGFP-C2 mammalian expression vector, fused to EGFP tag (see Table M3). For Co-immunoprecipitation assays, HEK293T were transfected with pEGFP-*cDNA prey* vector for 24 h and immunoprecipitations were performed with an anti-GFP antibody. HEK293T transfected with empty pEGFP-C2 was used as negative control. The presence of bait protein in the input and immunoprecipitated fraction was revealed by immunoblot with specific antibodies. Among the positive clones selected by two-hybrid, two new THOC1-interactors, MFAP1 (Microfibrillar-Associated Protein 1) and SAP130 (Sin3A Associated Protein 130 kDa), were confirmed by co-immunoprecipitation in human cells (Figure R3A, B). The rest of candidates could not be confirmed by biochemical approaches, at least at the conditions assayed (see Materials and Methods 6.2). THOC1 was detected by immunoblot in the immunoprecipitate of HEK293T cells transiently transfected with the plasmid encoding GFP-MFAP1 fusion by pull down with an anti-GFP antibody (Figure R3B). Immunoprecipitation assays with available polyclonal SAP130 antibody and HEK293T cell extracts also revealed an *in vivo* interaction between THOC1 and SAP130 protein (Figure R3B).

Furthermore we showed the *in vivo* association of these proteins in HeLa cells by proximity ligation assays (PLA), which allows the detection of cellular protein–protein associations *in situ* (see Materials and Methods 6.3) (Soderberg et al. 2006). A strong PLA signal was detected using anti-MFAP1 and anti-THOC1 (Figure R4A) and anti-SAP130 and anti-THOC1 (Figure R4B) specific antibodies, thus confirming a close association of THOC1 and MFAP1 and SAP130 in human cells.
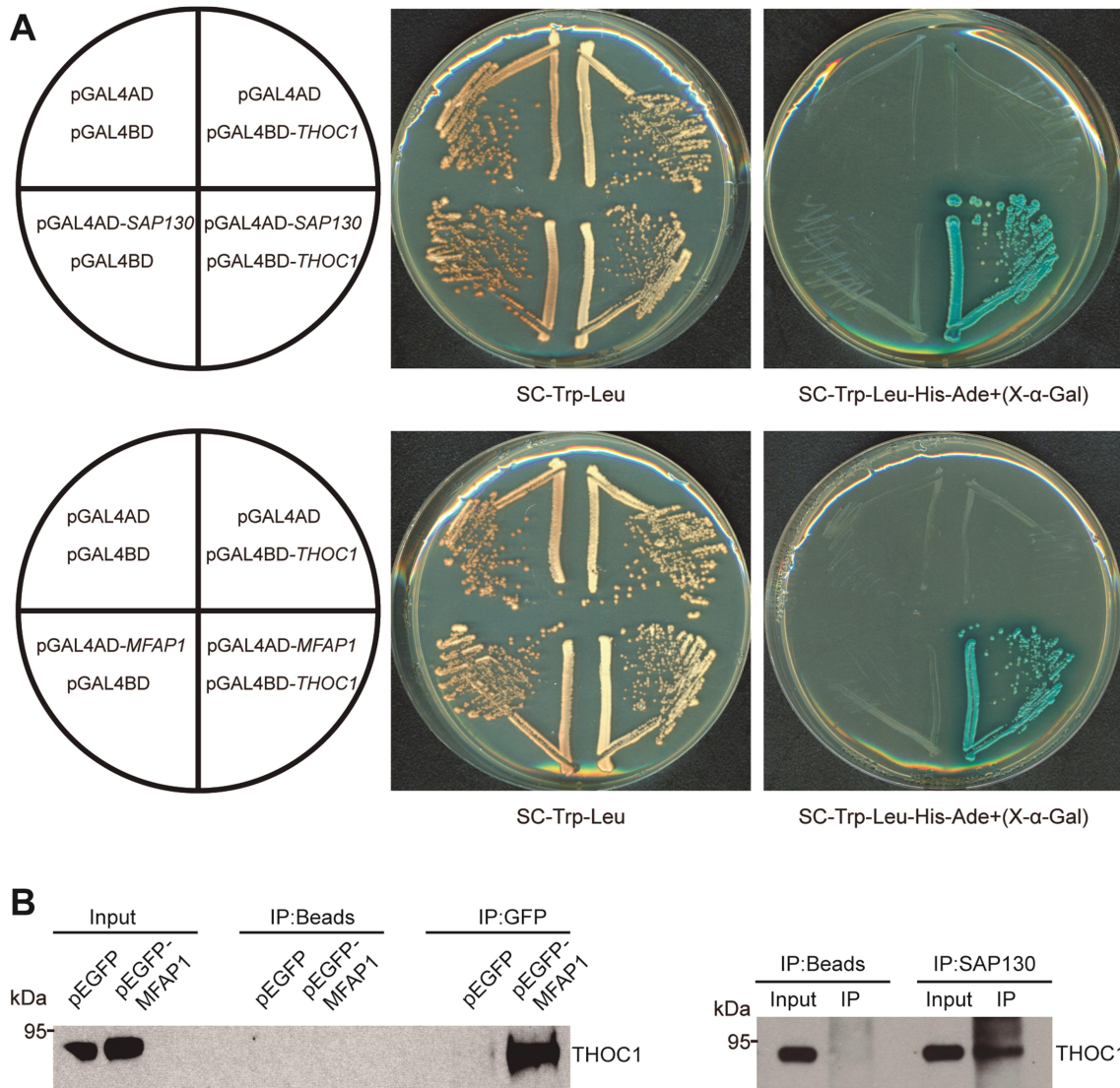
Figure R3. Physical interaction of THOC1 with MFAP1 and SAP130.
(A) Two-hybrid interaction between MFAP1 and THOC1 (upper panel) and SAP130 and THOC1 (lower panel) determined by cell growth and enzymatic activity (blue colonies). **(B)** MFAP1 and THOC1 protein interaction detected by Co-Immunoprecipitation (Co-IP) with anti-GFP antibody from whole cell extracts of HEK293T cells after transient expression of the GFP-MFAP1 fusion protein (left panel). THOC1 and SAP130 interaction detected by Co-IP assays in whole cell extracts of HEK293T with anti-SAP130 antibody (right panel). Input extract and total immunoprecipitate (IP) were analyzed by western with anti-THOC1 antibody (B).
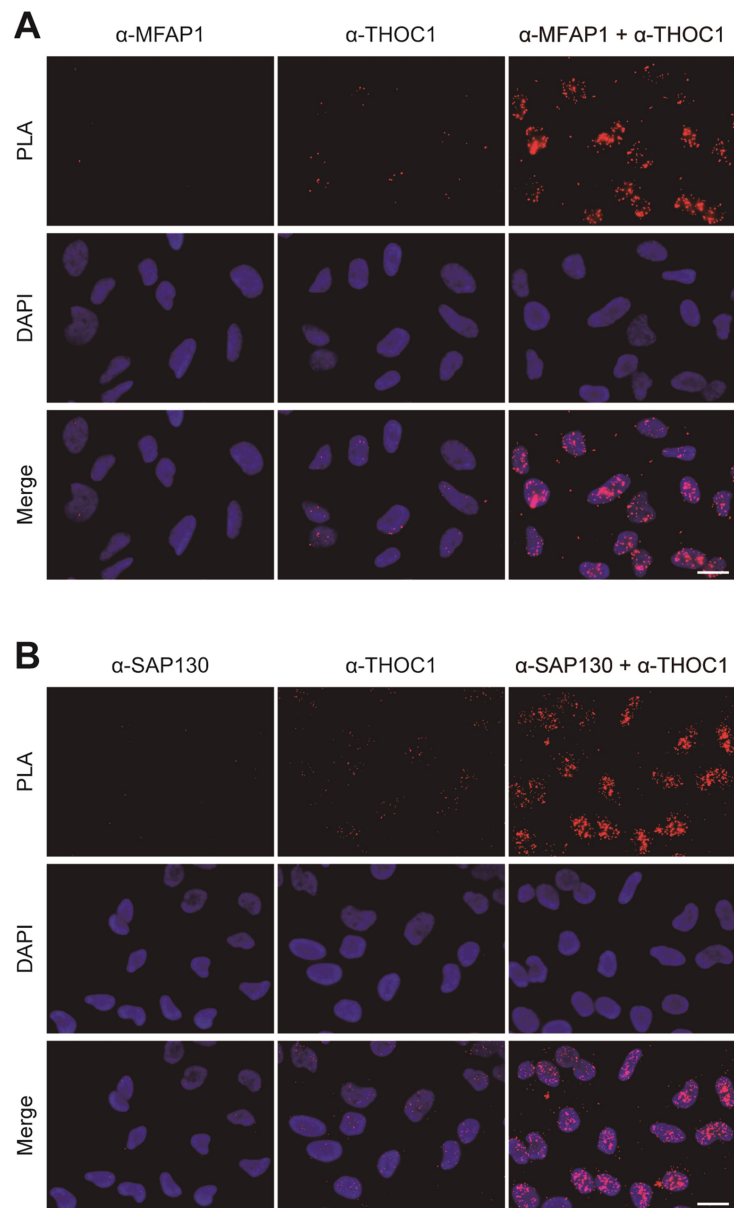
Figure R4. In situ proximity ligation assay between endogenous MFAP1 and THOC1 and SAP130 and THOC1.
(A) Proximity ligation assay (PLA) showing specific interactions of THOC1 and MFAP1 endogenous proteins. **(B)** PLA signal (red spots) indicating specific interaction between THOC1 and SAP130 endogenous proteins. Negative controls with only one of the antibodies are also shown (A, B). DNA was stained with DAPI. Scale bars, 20 μm.

2.   ROLE OF HUMAN SIN3A COMPLEX IN THE MAINTENANCE OF GENOME STABILITY

Given the physical interaction between SAP130 and THOC1, we wondered whether SAP130 plays any kind of role in the maintenance of genome integrity, as is the case of THO/TREX complex. For this purpose we carried out a functional analysis using RNA interference (RNAi) in human cells.

2.1. Analysis of cell proliferation and cell cycle in SAP130-depleted cells

SAP130 is part of the human Sin3A corepressor complex (Fleischer et al. 2003). This complex plays a role in transcription regulation through deacetylation of nucleosomes by histone deacetylases (HDACs) and the recruitment of others chromatin remodelers and therefore it is implicated in the regulation of key biological processes (Grzenda et al. 2009). It has been reported that loss of Sin3A causes defects in cell cycle and proliferation (Pile et al. 2002; Cowley et al. 2005). Given that SAP130 is a subunit of the Sin3A complex, we wonder whether it could be also relevant for cell cycle and cell proliferation. To address this question we depleted SAP130 via small interfering RNA (siRNA) from HeLa cells.
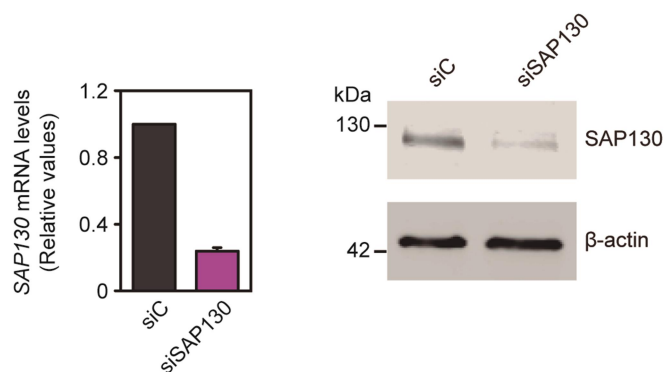
Figure R5. Analysis of siRNA-mediated SAP130 depletion.
Relative *SAP130* mRNA quantification measured by RT-qPCR in siRNA-transfected HeLa cells (left panel). *SAP130* mRNA expression values were normalized respect to the expression of the *HPRT* housekeeping gene. Means and SEM of three independent experiments are depicted. Western blot analysis showing SAP130 expression in HeLa cells treated with control and SAP130 siRNAs (right panel). β-actin protein was used as a loading control.

First we checked the efficiency of a siRNA pool against SAP130 by RT-PCR and western-blot. A clear reduction of SAP130 mRNA and protein levels was observed in HeLa cells after 72 hours of transfection (Figure R5). Then, we performed proliferation curves of control and SAP130-transfected HeLa cells. SAP130 depletion severely reduced cell proliferation (Figure R6A). Next, we considered the possibility that cell cycle could be affected in SAP130-depleted cells. To address this question we first analyzed the cell cycle distribution of asynchronously growing siRNA-transfected HeLa cells. Before collecting samples, cells were incubated with EdU to allow its incorporation into the DNA during replication and then subjected to FACS analysis. No differences were found in the percentages of cells in G1, S and G2 phase between control and SAP130-depleted cells, thus discarding a cell cycle arrest (Figure R6B). Since the analysis of asynchronous cultures is sometimes not sensitive enough to detect cell cycle alterations we decided to analyze cell cycle progression after synchronization. We synchronized the cells in G1 phase by double thymidine block and focused on S phase progression in order to see whether SAP130 depletion affects replication progression as previously reported in the absence of Sin3 and THO/TREX complex components (Aparicio et al. 2004; Dominguez-Sanchez et al. 2011a). Percentages of G1, S and G2 cell populations in the different time points analyzed after release were very similar in siRNA control and SAP130-depleted cells (Figure R6C). S phase progression was found to be slightly faster in siSAP130 than in siC cells. Considering that proliferation is seriously affected but cell cycle is not, we also analyzed the apoptotic cell population. As shown in Figure R6D the amount of apoptotic cells, as measured by sub-G_1 DNA content, was similar in siC- and siSAP130-depleted cells. Thus, although cell proliferation is affected in SAP130-depleted cells, as is the case of Sin3-deficient cells, this does not seem to be due to major problems in the cell cycle or to an increase of apoptosis.
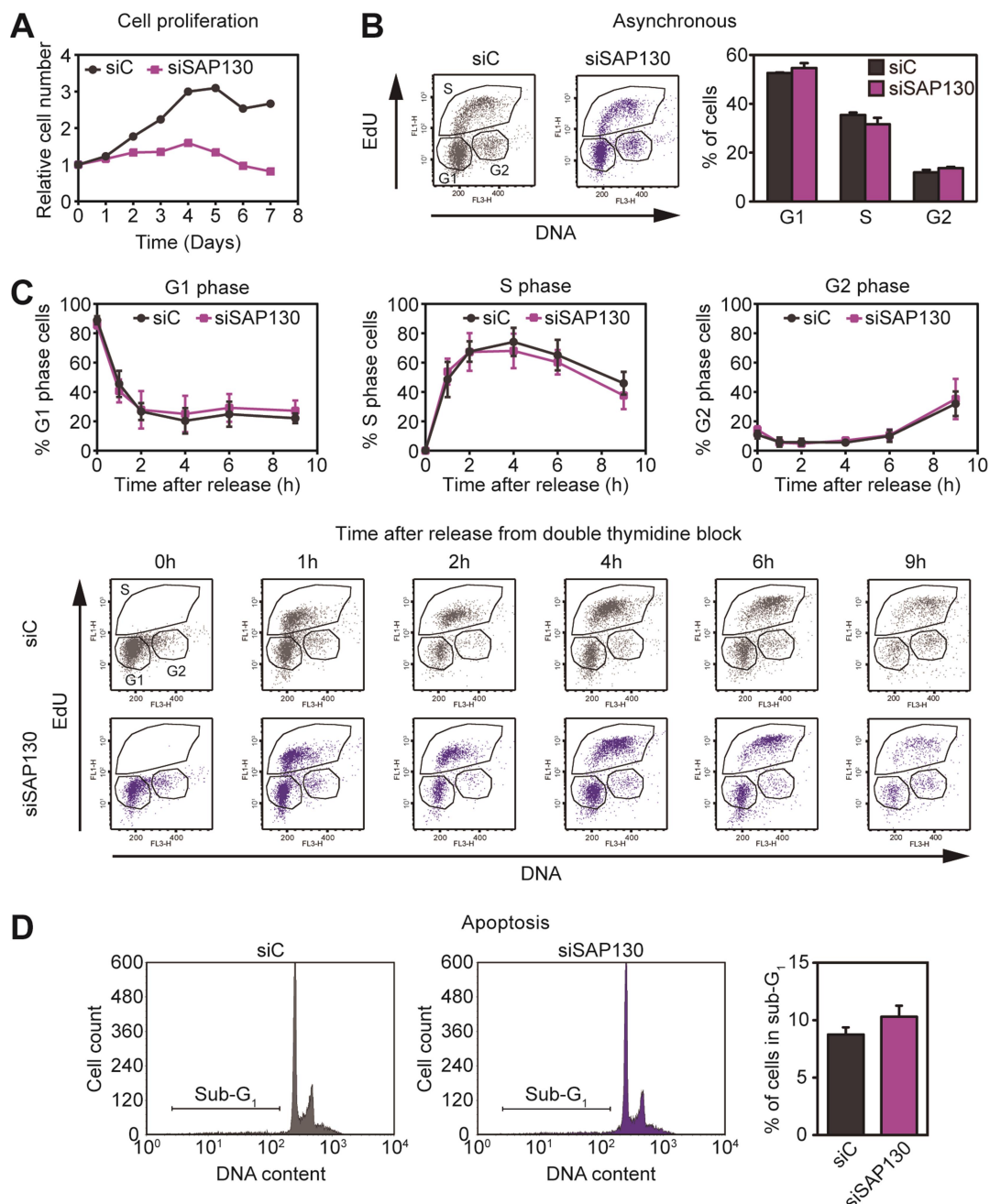
Figure R6. Defective proliferation of cells depleted of SAP130.
(A) Cell proliferation curves of HeLa cells after SAP130 depletion. Cells were plated 48h after siRNA depletion (day 0). The relative number of living cells was determined by measuring the metabolic activity of mitochondrial dehydrogenases. Values were normalized to the day 0 value. A representative experiment is shown. (B) FACS analysis of asynchronously growing HeLa cells depleted of SAP130 after 20 min of incubation with 20 µM EdU. Cells were separated based on their DNA content (7-AAD staining) and replication was analyzed by EdU incorporation. (C) FACS analysis of cell cycle progression in control and SAP130-depleted cells after G1-phase synchronization by double thymidine block. Samples were collected at the indicated times (hours) after release and processed for flow cytometry. G1, S and G2 cell population were separated based on their DNA content and replication activity as in (B). The percentages of cells in G1, S and G2 phases are depicted (B, C). (D) Analysis of apoptosis in siC- and siSAP130-transfected HeLa cells as determined by FACS analysis of cells displaying subG1-DNA content. Means and SEM of three independent experiments are depicted (B, C, D). See Materials and Methods 7-9.

2.2. Recruitment of SAP130 and SIN3 to active transcribed chromatin

Given the interaction between THO and SAP130, we first studied the recruitment of SAP130 to active chromatin by chromatin immunoprecipitation. Sin3 complex has been shown to be recruited at promoter regions in agreement with its role in transcription repression by deacetylation (Kadosh and Struhl 1997; van Oevelen et al. 2008; Smith et al. 2010). Nevertheless, recent reports show that Sin3 complex is also recruited to ORFs of transcribed genes (Carrozza et al. 2005; van Oevelen et al. 2010; Jelinic et al. 2011) as is the case of THO complex (Luna et al. 2012). We performed chromatin immunoprecipitation (ChIP) assays along the highly transcribed *GAPDH* gene (Figure R7A). Both SAP130 and SIN3, the core subunit of the Sin3A complex, were recruited to chromatin with a similar profile as shown in Figure R7B. Thus, we confirmed that SAP130 and SIN3 are recruited not only to promoter but also along the transcribed gene, preferentially at regions closed to 5´end.
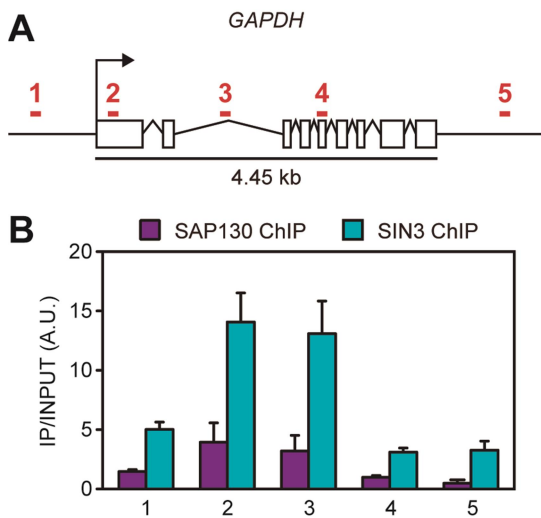
Figure R7. SAP130 and SIN3 are recruited to chromatin along actively transcribed genes.
(A) Schematic diagrams of *GAPDH* gene, with the exons depicted as open boxes. The arrow indicates the start of transcription and numbered red lines indicate the regions where the ChIP analyses were performed. **(B)** ChIP analysis of SAP130 and SIN3 within the *GAPDH* gene in HeLa cells. Values represent the ratios of precipitated DNA (IP) to input DNA (INPUT). The positions of amplified regions are indicated on the x axis. Means and SEM of three independent experiments are depicted.

2.3. Chromatin immunoprecipitation analyses of histone acetylation levels, RNAPII and THOC1 in SAP130-depleted cells

Since SAP130 is part of the histone deacetylase complex Sin3A, we examined the histone acetylation levels after its depletion. For this purpose, we performed ChIP analysis of total histone H3 and histone H4 acetylation along *GAPDH* gene in cells depleted of SAP130. Although no significant differences were found with respect to siC

control cells (Figure R8), a slight increase in histone acetylation was observed in SAP130-depleted cells.

Since lack of functional yeast and human THO causes transcription defects (Chavez et al. 2000; Rondon et al. 2003; Dominguez-Sanchez et al. 2011a), next we wondered whether SAP130 depletion could also affect transcription. To test this possibility, we performed ChIP analysis of total RNAPII in *GAPDH* gene, and also in *β-actin* gene (Figure R9) in SAP130-depleted cells. No significant differences in RNAPII levels were observed between control and cells depleted of SAP130 (Figure R9C, D). These data suggest that transcription seems not to be affected in SAP130-depleted cells. In parallel we also performed ChIP analysis of THOC1 to determine whether SAP130 is involved in the recruitment of THOC1 to chromatin. Although no significant differences were observed, THOC1 levels are slightly higher in SAP130-depleted cells (Figure R9E, F). Thus, we concluded that SAP130 is not required for the recruitment of THOC1 to chromatin.
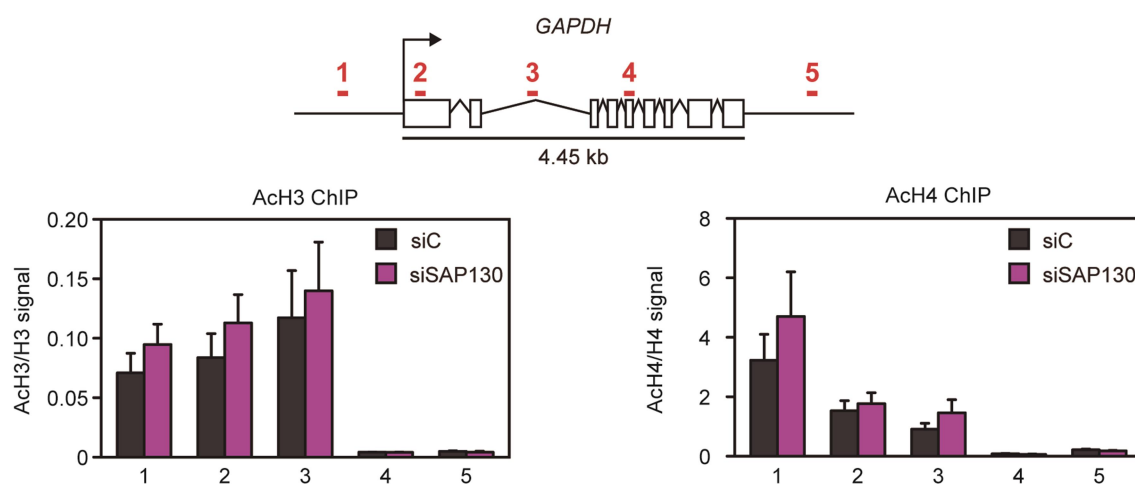
Figure R8. Histone acetylation levels in SAP130-depleted cells.
ChIP analysis of histone H3 and H4 acetylation in siControl- and siSAP130-transfected HeLa cells within the *GAPDH* gene. Acetyl-Histone H3 and H4 IP/INPUT signals, relative to H3 or H4 IP/INPUT signals, respectively, are plotted. Means and SEM of three independent experiments are shown. Schematic diagram of *GAPDH* gene is depicted. Numbered red lines indicate the regions where the ChIP analyses were performed.
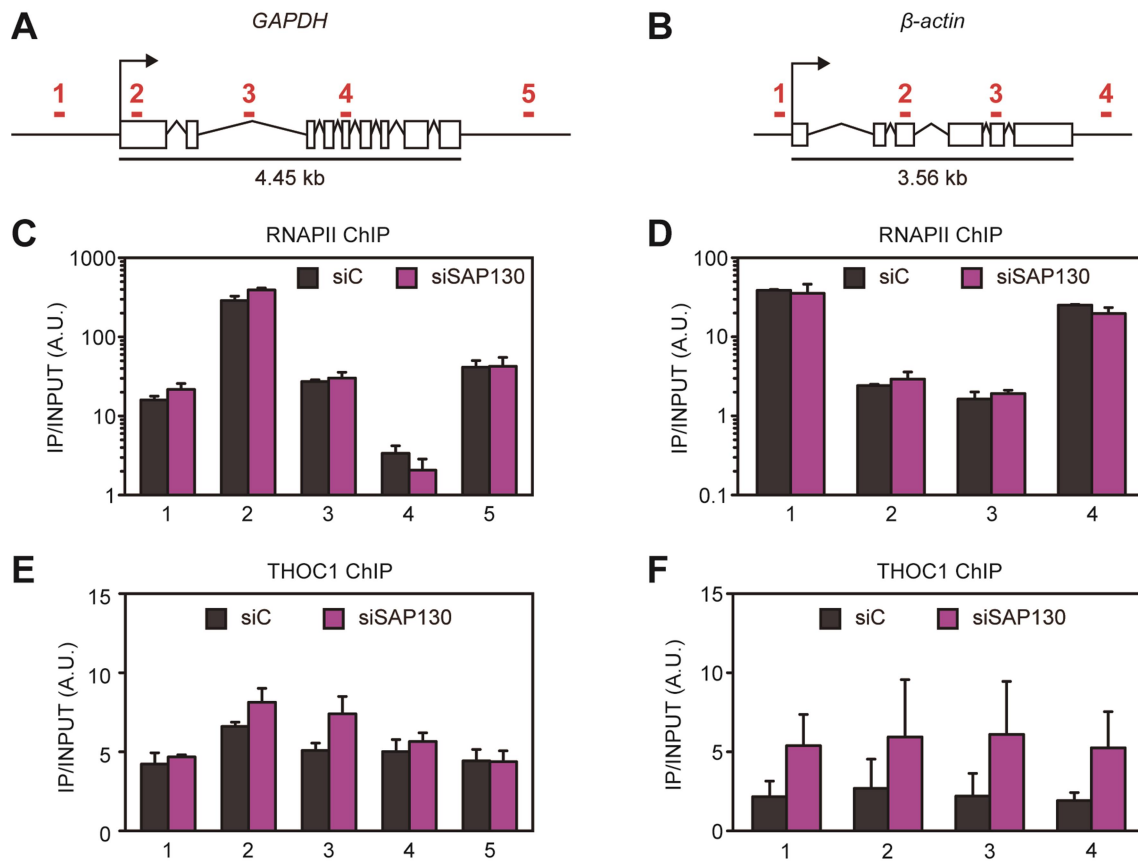
Figure R9. RNAPII and THOC1 recruitment to chromatin in SAP130-depleted cells.
(A, B) Schematic diagrams of (A) *GAPDH* and (B) *β-actin* genes. **(C, D)** Quantitative ChIPs analyses performed with antibody against total RNAPII in control and SAP130-depleted HeLa cells within the *GAPDH* and *β-actin* gene, respectively. **(E, F)** ChIP analysis of THOC1 in siC- and siSAP130-transfected HeLa cells within the *GAPDH* and *β-actin* gene, respectively. Means and SEM of three independent experiments are depicted (C, D). Other details as in Figure R7.

2.4. Association between THO/TREX and Sin3A complexes

Once established that THOC1 associates with SAP130, since both proteins are subunits of two large complexes, we decided to investigate the interaction between them. First, we performed co-immunoprecipitation assays and PLA between THOC1 and SIN3, the core component of the Sin3A complex. SIN3-THOC1 interaction was validated in human cell by co-immunoprecipitation with an anti-SIN3 antibody as shown in Figure R10A. Moreover, a strong PLA signal was observed between THOC1 and SIN3 (Figure R10B).
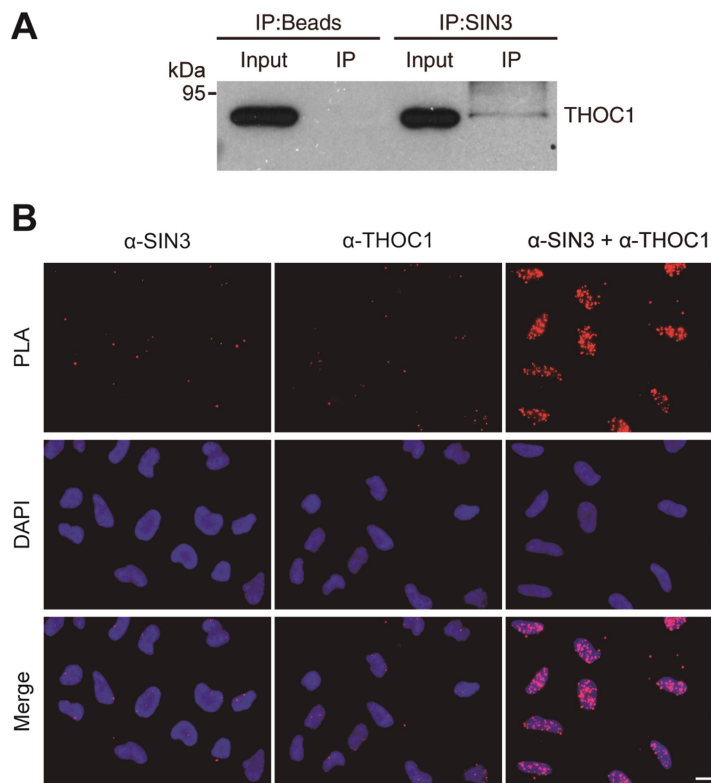
Figure R10. Physical interaction of THOC1 with SIN3. **(A)** THOC1 and SIN3 interaction detected by Co-IP assays in whole cell extracts of HEK293T with anti-SIN3 antibody. Input extract and total immunoprecipitate (IP) were analyzed by western with anti-THOC1 antibody. **(B)** PLA showing the association of THOC1 and SIN3 endogenous proteins. Red spots are indicative of a positive PLA signal. Negative controls with only one of the antibodies are also shown. Scale bar, 20 μm.

Next we decided to study the interaction between SIN3 and other subunits of the THO/TREX complex. We chose THOC2, THOC5 and UAP56 as representative subunits and PLA experiments were performed with specific antibodies (Figure R11, upper panel). Background signal of each antibody is also shown in Figure R11, lower panel. As expected a clear interaction of THOC1 with THOC2, THOC5 and UAP56 was observed. Negative PLA was observed between THOC2 or THOC5 with SIN3 (Figure R11, upper panel). The absence of PLA signal with some subunits is not incompatible with the association between these two complexes, since different subunits of each complex could be not close enough to result in a positive PLA signal, overall considering the large size of THO/TREX and Sin3A complexes. Although no positive signal could be detected for THOC2 and THOC5 proteins with SIN3, PLA assays revealed that the mRNA export factor UAP56 associates with SIN3 (Figure R11, upper panel). Altogether this data supports an association between THO/TREX and Sin3A-HDAC complexes.
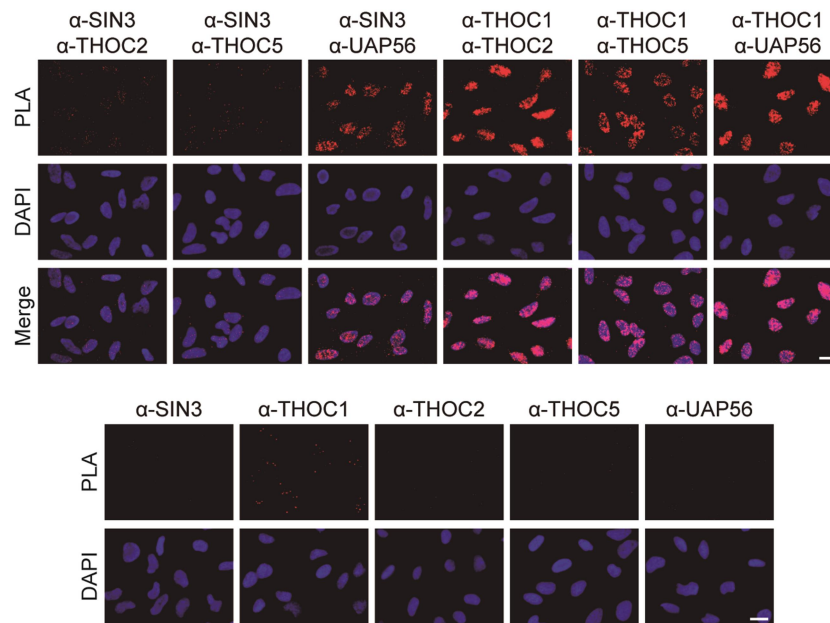
Figure R11. PLA between SIN3 and different subunits of the THO/TREX complex.
PLA performed with SIN3 and THOC1 against THOC2, THOC5 and UAP56 subunits of the THO/TREX complex (upper panel). Red spots are indicative of a positive PLA signal. Negative controls with only one of the antibodies are also shown (lower panel). Scale bars, 20 μm.

2.5. Genome instability in Sin3A-depleted cells

To determine whether the physical interaction of THO/TREX and Sin3A complexes had any functional implication we assayed first whether SAP130 plays any kind of role in the maintenance of genome integrity, as is the case of THOC1 (Dominguez-Sanchez et al. 2011a). For this purpose we first determined γH2AX and 53BP1foci, which accumulate at sites of DBSs and are used as markers of DNA damage/repair, in HeLa cells depleted of SAP130. No differences were found with respect to siC control in the percentages of γH2AX and 53BP1 foci (Figure R12A). Next, we performed more sensitive assays as alkaline single-cell electrophoresis which detects DNA breaks. Importantly, alkaline single-cell electrophoresis revealed a significant increase in tail moment (2.6 fold) in siSAP130-depleted cells indicating an accumulation of DNA breaks (Figure R12B). Moreover, depletion of other subunits of the Sin3A complex, such as SIN3, SAP30, which acts as a bridge with other co-repressors, and SUDS3, required for the integrity and activity of the complex (Grzenda et al. 2009), leads to a high significant increase in DNA breaks (Figure R12B). An increase in DSBs was also observed by neutral single-cell electrophoresis assays in cells depleted by siRNAs of different subunits of the Sin3A complex as well as in siTHOC1 cells (Figure R12C). Altogether this data suggest that the Sin3A complex plays a role in preventing DNA break, similar to those of the THO complex.

84
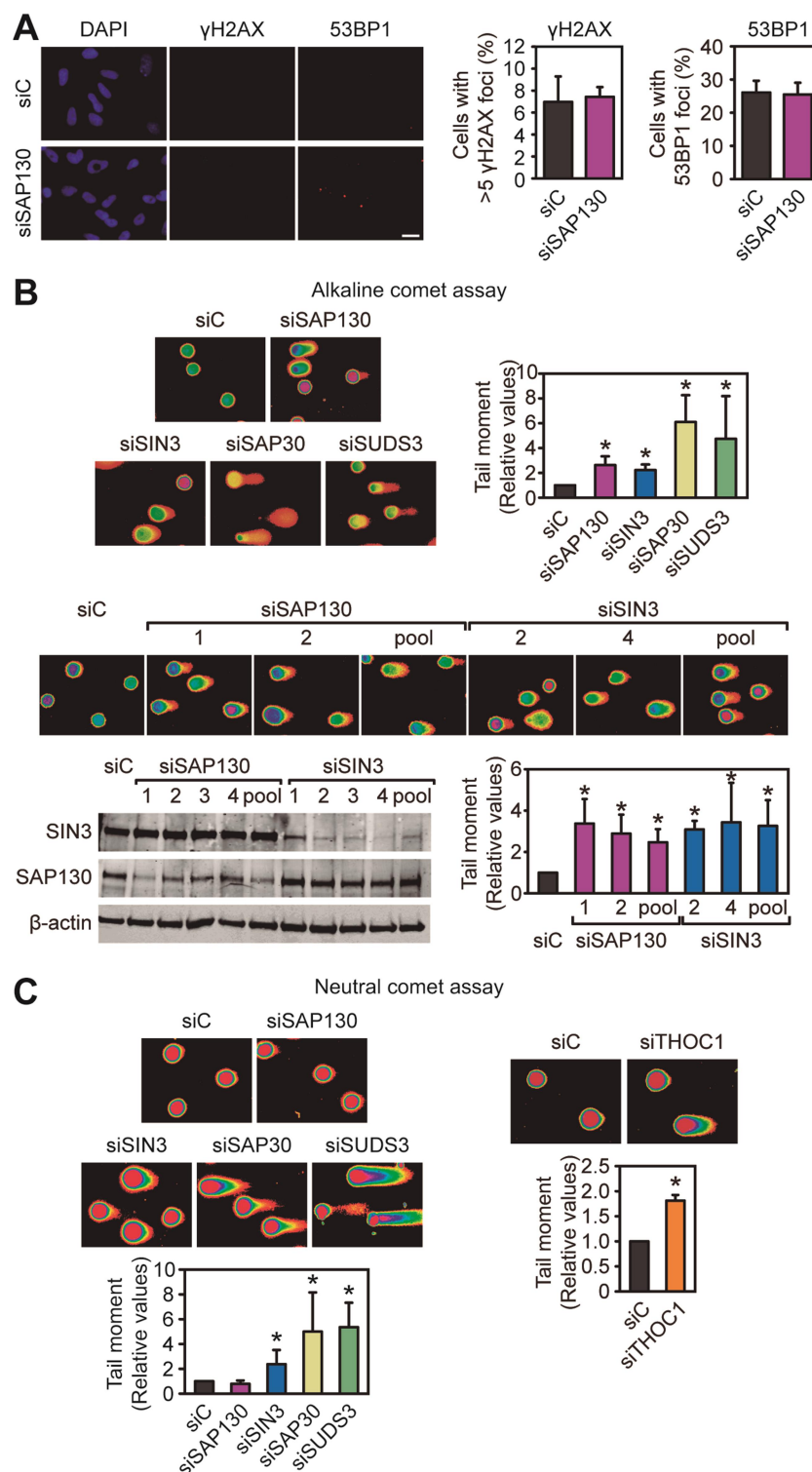
Figure R12. DNA breaks caused by Sin3A complex depletion.
(A) Immunofluorescence of γH2AX and 53BP1 after transfection with control and SAP130 siRNAs. Nuclei were stained with DAPI. Percentages of cells with >5 H2AX foci and cells with 53BP1 foci are shown. Means and SEM are plotted (n=3). Scale bar, 20 μm. **(B)** Alkaline single cell electrophoresis in HeLa cells transfected with siRNAs against siC (control), siSAP130, siSIN3, siSAP30, and siSUDS3 (upper panel) or different independent siRNAs against SAP130 or SIN3 (lower panel). Relative comet-tail moments are plotted as means and SEM (n≥3). **(C)** Neutral single cell electrophoresis of HeLa cells transfected with the indicated siRNAs. Relative comet-tail moments are plotted as means and SEM (n=3). *, P < 0.05 (Mann-Whitney U test) (B, C).

2.6. Transcription and R loop-dependent genome instability in Sin3A-depleted cells

Since lack of functional yeast and human THO lead to genome instability in a transcription and R loop-dependent manner (Chavez et al. 2000; Huertas and Aguilera 2003; Dominguez-Sanchez et al. 2011a), we next considered the possibility that the accumulation in DNA breaks caused by Sin3A complex depletion was also transcription dependent. To asses this possibility we performed alkaline single-cell electrophoresis in the presence of the transcription inhibitor cordycepin, which specifically inhibits RNA chain elongation. A full suppression of the increase in DNA breaks was observed after depletion of SAP130 and SIN3, taken as representative subunits of the Sin3 complex, (Figure R13A, upper panel). Similar results were also obtained with the transcription inhibitor 5,6-dichloro-1-β-D-ribofurosylbenzimidazole (DRB) (Yamaguchi et al. 1998), indicating that the effect is not specific of inhibitor and that genome instability induced by depletion of these factors is mediated by transcription (Figure R13A, lower panel). Consequently, we wondered whether the increase in DNA breaks was due to co-transcriptional RNA-DNA hybrids, as in THO-depleted cells. Alkaline single-cell electrophoresis assays of cells overexpressing or not RNase H1 (RNH1), a ribonuclease that specifically degrades the RNA moiety of RNA-DNA hybrids, revealed that the increase in DNA breaks was completely suppressed by RNH1 overexpression in siSAP130 and siSIN3 cells (Figure R13B). From these experiments we can conclude that DNA breaks accumulation in Sin3A-depleted cells is dependent on transcription and RNA-DNA hybrid formation.
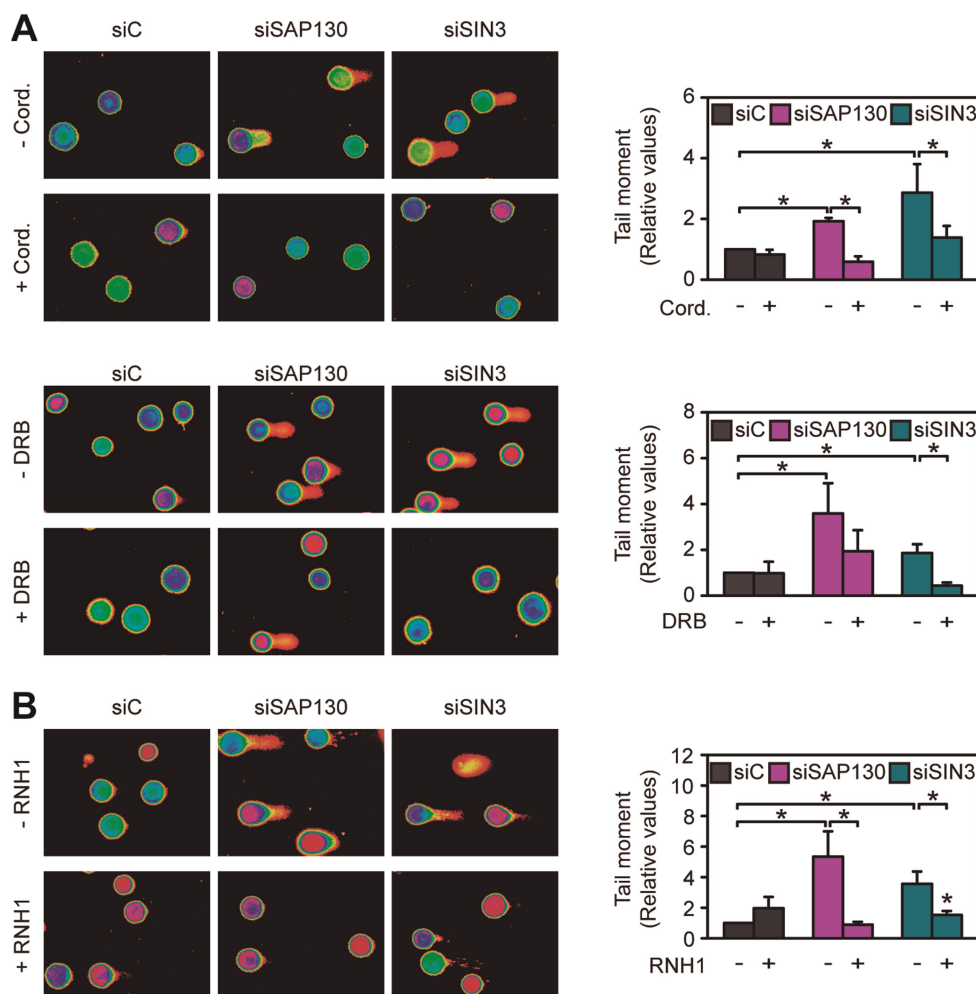
Figure R13. DNA breaks accumulation in Sin3A-depleted cells in relation to transcription and RNA-DNA hybrids.
(A) Alkaline comet assay in cells transfected with siC, siSAP130 or siSIN3 untreated or treated with 50 µM cordycepin (upper panel) or with 100 µM DRB (lower panel) for 4 h. Relative comet-tail moments are plotted (n≥3) as means and SEM. **(B)** Alkaline comet assay in siC, siSAP130 and siSIN3 cells transfected with pcDNA3 (-RNH1) or pcDNA3-RHaseH1 (+RNH1) during 48h. Relative comet-tail moments are plotted (n=3) as means and SEM. *, $P < 0.05$ (Mann-Whitney U test) (A, B).

2.7. R loop accumulation in Sin3A-depleted cells

Consequently, we assayed whether indeed R loops were increased in siSAP130 and siSIN3 cells by immunofluorescence using the anti-RNA-DNA hybrid S9.6 monoclonal antibody. A significant enrichment of the S9.6 nuclear signal was observed in both SAP130 and SIN3-depleted cells (Figure R14), confirming the accumulation of RNA-DNA hybrids at the cellular level. Then we assayed whether we could also detect an increase in R loop accumulation at the molecular level by DRIP-qPCR (Ginno et al. 2012) using the S9.6 monoclonal antibody in four human genes (*APOE*, *RPL13A*, *BTBD19* and *EGR1*) that have been previously validated for R-loop detection (Bhatia et

al. 2014; Herrera-Moyano et al. 2014; Garcia-Rubio et al. 2015). Results clearly show that R loops accumulate at significantly higher levels in siSAP130 and siSIN3 than in siC control HeLa cells (Figure R15A, B). These high levels of RNA-DNA hybrids were not due to an increase in transcription, since no significant differences in mRNA levels, as detected by RT-qPCR, or in RNAPII occupancy, as determined by ChIP analyses, were observed (Figure R15C). Since SIN3A associates predominantly at the transcription start site (TSS) of human genes (van Oevelen et al. 2010; Williams et al. 2011; Sanz et al. 2016), and a similar distribution at 5´end of the *GAPDH* gene was observed (Figure R7), we wondered whether Sin3A complex could have a role in suppressing R-loops at TSS proximal regions. To address more accurately R-loop formation along the gene we performed DRIP-qPCR analysis using different primers from 5´ to 3´ end of *APOE* and *RPL13A* genes, which show the highest levels of RNA-DNA hybrids. The results revealed that R loops accumulated at the different regions tested, from 5' to 3' ends, in agreement with a role of the Sin3A complex in preventing co-transcriptional R-loops along the gene (Figure R16).
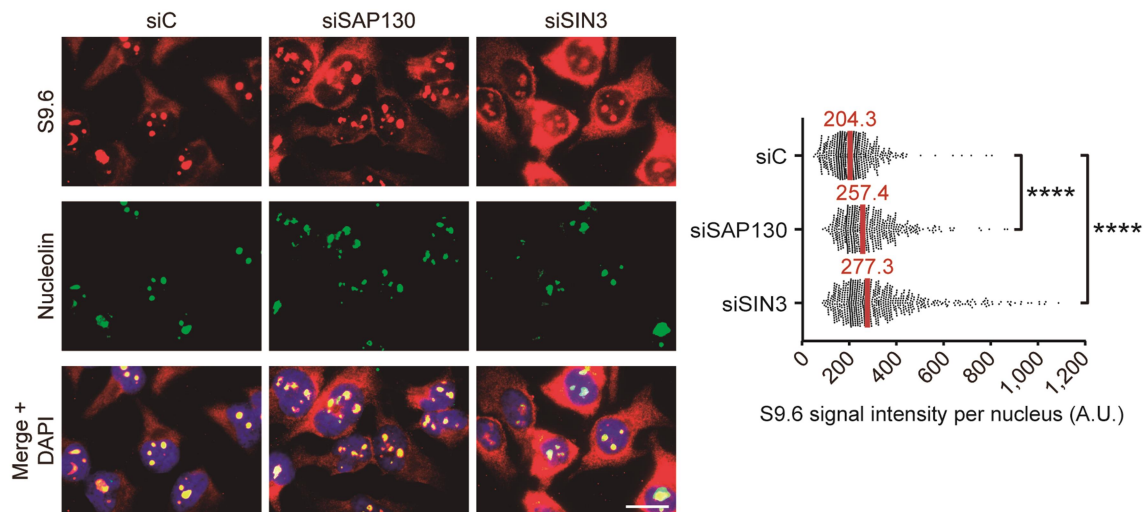
Figure R14. Analysis of nuclear RNA-DNA hybrid accumulation in Sin3A complex-depleted cells by S9.6 immunofluorescence microscopy.
Immunostaining with S9.6 (red) and anti-nucleolin (green) antibodies in siC, siSAP130 and siSIN3 transfected HeLa cells. The median of S9.6 signal intensity per nucleus after nucleolar signal removal is shown (n=3). ****, $P < 0.0001$ (Mann-Whitney U test, two-tailed). Scale bar, 20 µm.
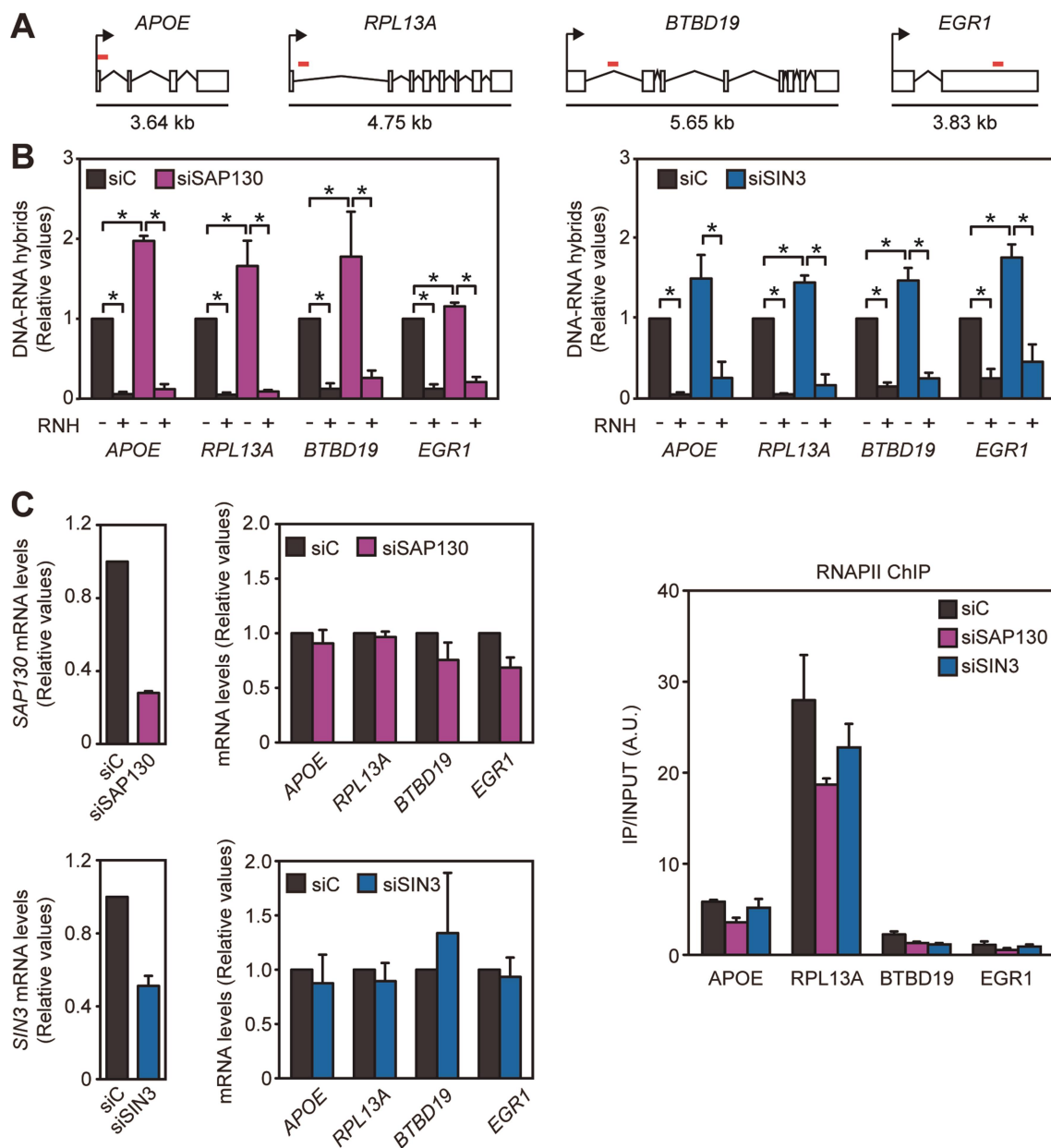
Figure R15. Effect of Sin3A depletion on R-loop accumulation as determined by DRIP assays.
(A) Schematic diagrams of *APOE*, *RPL13A*, *BTBD19* and *EGR1* genes. Exons are depicted as open boxes, the arrows indicate the start of transcription and red lines indicate the regions where the DRIP-qPCR analyses were performed. **(B)** DRIP-qPCR using the anti-RNA-DNA hybrids S9.6 monoclonal antibody in SAP130-depleted cells (left panel) and SIN3-depleted cells (right panel) at indicated regions. Signal values normalized with respect to the siC control are plotted (n=3) as means and SEM. **(C)** Validation of siRNAs and relative mRNA quantification of *APOE*, *RPL13A*, *BTBD19* and *EGR1* genes by RT-qPCR in SAP130-depleted cells and SIN3-depleted cells (left panel). mRNA expression values were normalized to the expression of the *HPRT* housekeeping gene. RNAPII ChIP analysis in siSAP130 and siSIN3 transfected HeLa cells or in siC transfected cells at A*POE*, *RPL13A*, *BTBD19* and *EGR1* genes (right panel). Data are plotted as mean and SEM. *, P < 0.05 (Mann-Whitney U test) (B).
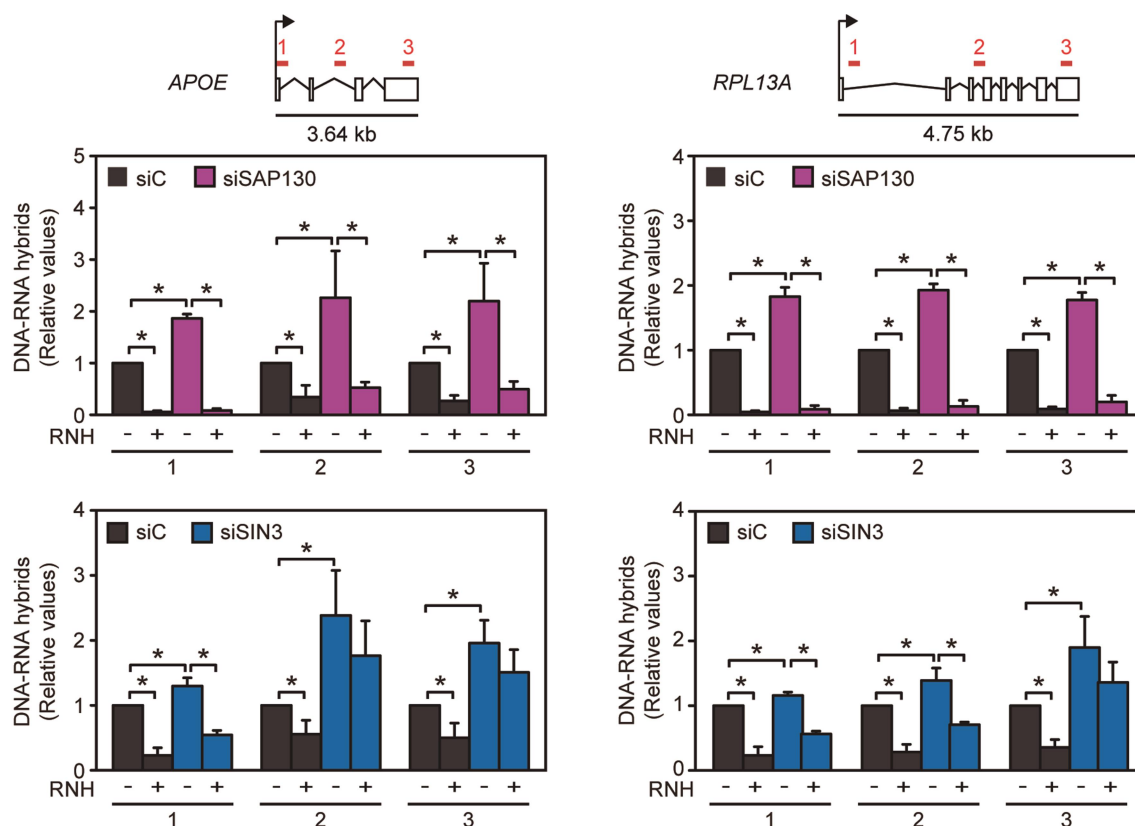
Figure R16. Analysis of R-loop accumulation along the gene after Sin3A complex depletion by DRIP assays.
DRIP-qPCR in siC, siSAP130 and siSIN3 cells at different regions of *APOE* and *RPL13A* genes. Schematic diagrams of genes are depicted at the bottom. Numbered red lines indicate the regions where PCR analyses were performed. Signal values normalized with respect to the siC control are plotted (n=3) as means and SEM. *, $P < 0.05$ (Mann-Whitney U test).

Next, we assayed R-loop accumulation by S9.6 immunofluorescence in cells transfected with siRNA against both THOC1 and SIN3. Interestingly, a double-depletion resulted in significant increase of R loops compared to the siC control (Figure R17). Moreover, RNA-DNA levels were higher in double-depleted cells than in the single depletion of each factor (Figure R17). These results support a functional interaction between Sin3A complex and THO preventing R-loop formation.
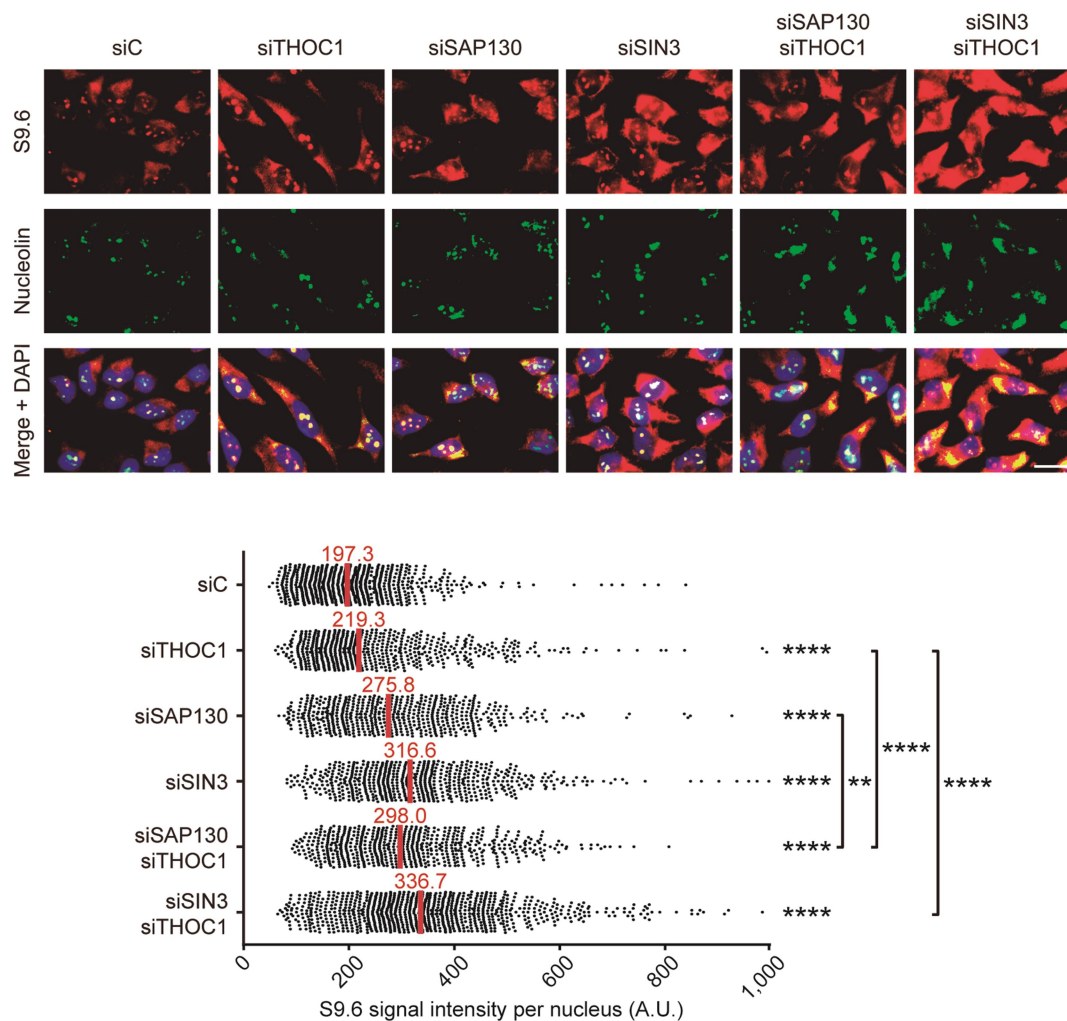
Figure R17. DNA-RNA hybrids accumulation after single and double depletion of THOC1 and Sin3A complex subunits.
Immunostaining with S9.6 (red) and nucleolin (green) antibodies in siRNA-transfected HeLa cells. The median of the S9.6 signal intensity per nucleus after nucleolar signal removal (n=3) is shown. ****, $P < 0.0001$; **, $P < 0.01$ (Mann-Whitney U test, two-tailed). Scale bar, 20 μm.

2.8. R-loop accumulation related to high levels of histone acetylation

Since the Sin3A complex plays a role in transcription regulation through deacetylation of nucleosomes by histone deacetylases (HDACs) and the recruitment of other chromatin remodelers (Grzenda et al. 2009), our results suggest that histones would need to be deacetylated after transcription to prevent R loops. Consequently, we used two different HDAC inhibitors, trichostatin A (TSA) (Yoshida et al. 1990) and suberoylanilide hydroxamic acid (SAHA) (Smith et al. 2010), to determine whether high levels of acetylated histones facilitate R-loop accumulation. Immunofluorescence assays revealed a significant increase in nuclear S9.6 signal in cells treated with 250 nM

91

TSA or 5µM-7.5µM SAHA during 3 h, correlating with the high levels of histone H3 acetylation reached, as determined by western blot (Figure R18A, B).
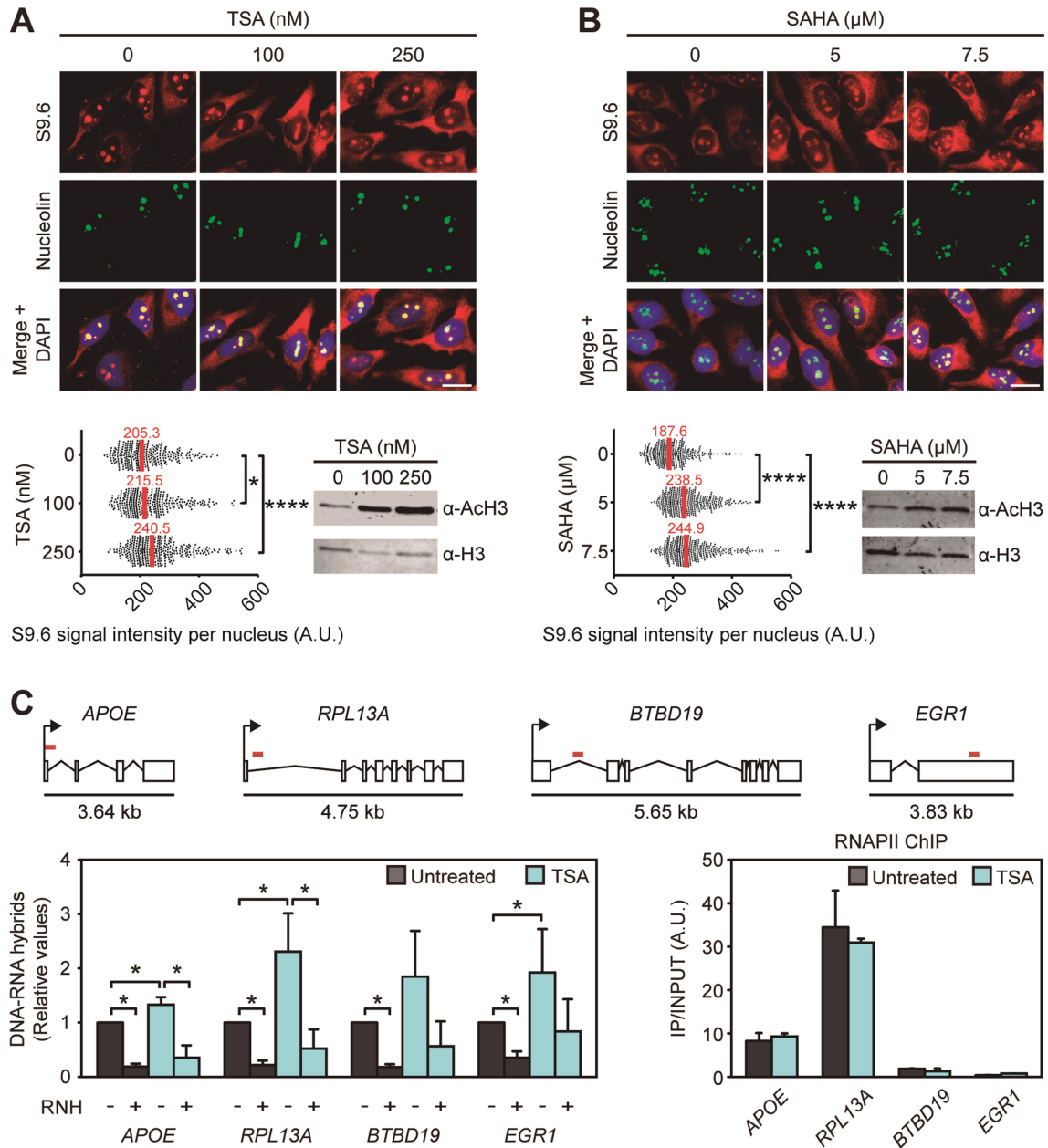
Figure R18. Effect of the inhibition of histone deacetylation by HDAC inhibitor treatments on R-loop levels.
(A) S9.6 immunofluorescence in HeLa cells untreated or treated for 3 h with 100, 250 nM trichostatin A (TSA) or (B) with 5, or 7.5 µM SAHA. (A, B) Western blot analysis of histone H3 acetylation in cells untreated or treated with HDAC inhibitor are shown. The median of the S9.6 signal intensity per nucleus after nucleolar signal removal is shown (n=3). *, P < 0.05; ****, P < 0.0001 (Mann-Whitney U test, two-tailed). Scale bar, 20 µm. (C) DRIP-qPCR using the S9.6 anti-RNA-DNA hybrid monoclonal antibody in cells untreated or treated for 3 h with 250 nM TSA (left panel). Signal values normalized with respect to the siC control are plotted (n=3) as means and SEM. RNAPII ChIP analysis at the same regions analyzed by DRIPq-PCR (right panel). Data are plotted as mean and SEM (n=3). Schematic diagrams of genes are depicted at the bottom. *, P < 0.05 (Mann-Whitney U test).

To confirm R loop accumulation at the molecular level, we performed DRIP analysis in TSA treated cells in the four human genes *APOE*, *RPL13A*, *BTBD19* and *EGR1*. Results clearly show that histone deacetylation inhibition leads to high levels of R-loops at the genes analyzed (Figure R18C). These high levels of RNA-DNA hybrids seems not to be due to an increase in transcription, since no significant differences in RNAPII occupancy, as determined by ChIP analyses, were observed (Figure R18C). Altogether the data suggest a functional relationship between histone acetylation and R-loop accumulation, consistent with the role of Sin3A in genome integrity.

2.9. Suppression of genome instability and R-loop accumulation in THOC1-depleted cells via histone acetylation inhibition

Next, we reasoned that THO could interact with Sin3A to trigger co-transcriptional histone deacetylation and transiently close chromatin thus preventing the nascent RNA to stably hybridize back with the DNA template. One possibility was that THO regulates Sin3A recruitment to chromatin at transcribed regions. For this purpose we performed ChIP analysis at R-loop forming genes. The levels of SAP130 and SIN3 factors in siTHOC1 depleted cells were similar to those of siC cells, as determined by ChIP analysis (Figure R19A, B). Thus, we conclude that THO is not necessary for the recruitment of Sin3 to chromatin. Another possibility was that THO was necessary for proper deacetylation. Interestingly, histone H3 acetylation levels in siTHOC1 cells were similar to those of Sin3A-depleted cells and slightly increased respect to those of siC cells, as assayed by ChIP and western blot (Figure R19C, D). Consequently, we reasoned that if THO could affect histone acetylation as a way to prevent co-transcriptional R loops, we should be able to suppress siTHO phenotypes by maintaining histones hypo-acetylated. To test this hypothesis we used a treatment with the histone acetyltransferase inhibitor (HATi) anacardic acid (AA) (Balasubramanyam et al. 2003), as a way to transiently maintain histones hypo-acetylated, in THOC1-depleted cells. Immunofluorescence assays of γH2AX foci revealed that the high levels of γH2AX foci accumulated in siTHOC1 cells was suppressed by treatment with AA (Figure R20A). Moreover, a clear suppression of the increase in DNA breaks caused by THOC1 depletion was also observed by single cell electrophoresis (Figure R20B). In order to determine whether such suppression of genome instability in cells depleted of THOC1 is due to a suppression of R-loop accumulation we performed DRIP analysis in

THOC1-depleted cells untreated or treated with AA. Importantly, AA treatment in siTHOC1 cells leads to a reduction in R loops, as shown by DRIP-qPCR (Figure R21A). In addition, we checked that this suppression is not caused by a transcription reduction because of AA treatment, as determined by RT-qPCR and RNAPII ChIP (Figure R21B). Therefore, inhibition of histone acetylation helps prevent the R-loop formation.

Taken together, our results demonstrate a physical and functional interaction between the THO complex and the Sin3A complex that serves to prevent transcription-associated and R-loop-mediated genetic instability in human cells.
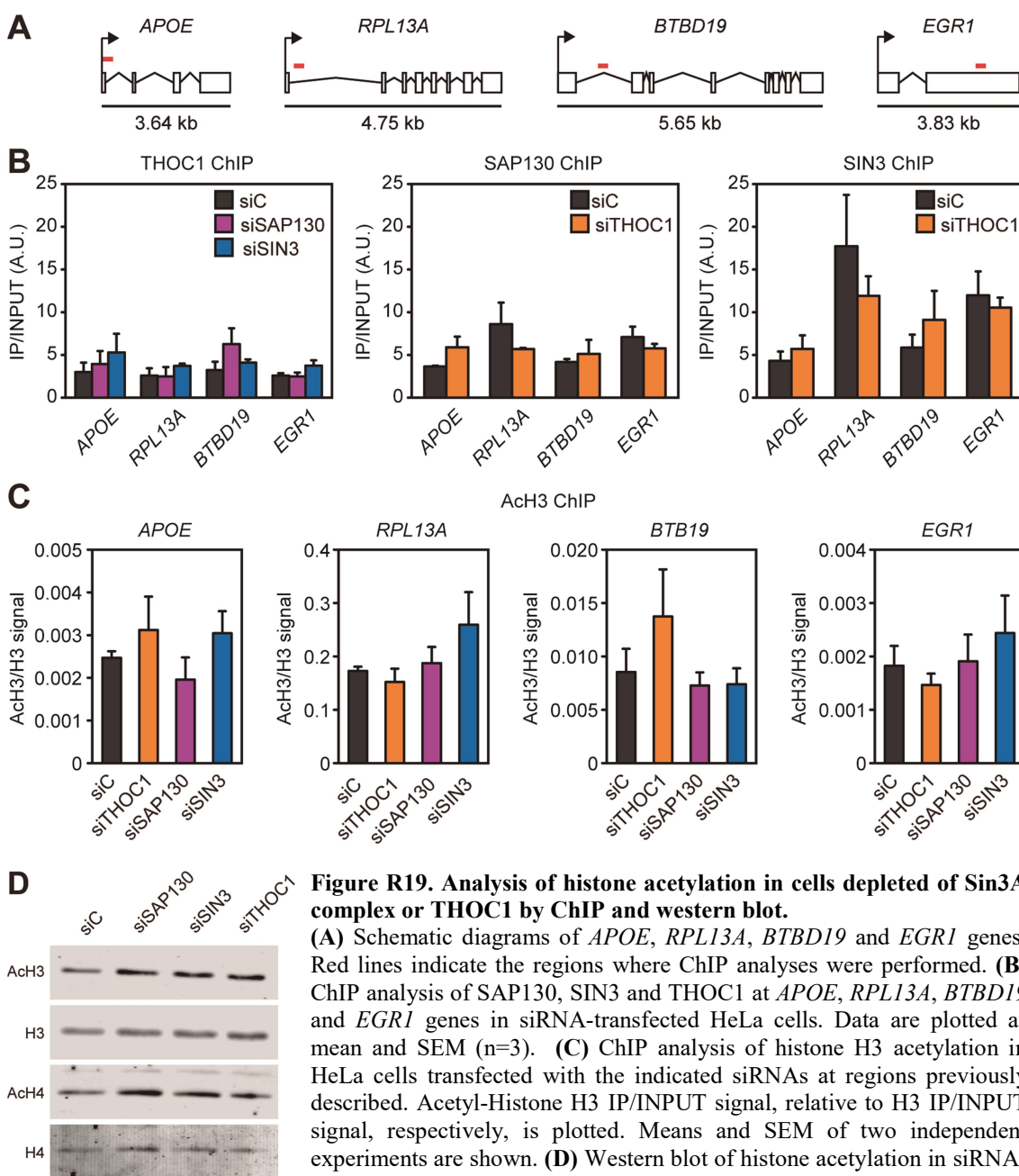
Figure R19. Analysis of histone acetylation in cells depleted of Sin3A complex or THOC1 by ChIP and western blot.
(A) Schematic diagrams of *APOE*, *RPL13A*, *BTBD19* and *EGR1* genes. Red lines indicate the regions where ChIP analyses were performed. (B) ChIP analysis of SAP130, SIN3 and THOC1 at *APOE*, *RPL13A*, *BTBD19* and *EGR1* genes in siRNA-transfected HeLa cells. Data are plotted as mean and SEM (n=3). (C) ChIP analysis of histone H3 acetylation in HeLa cells transfected with the indicated siRNAs at regions previously described. Acetyl-Histone H3 IP/INPUT signal, relative to H3 IP/INPUT signal, respectively, is plotted. Means and SEM of two independent experiments are shown. (D) Western blot of histone acetylation in siRNA-depleted cells.
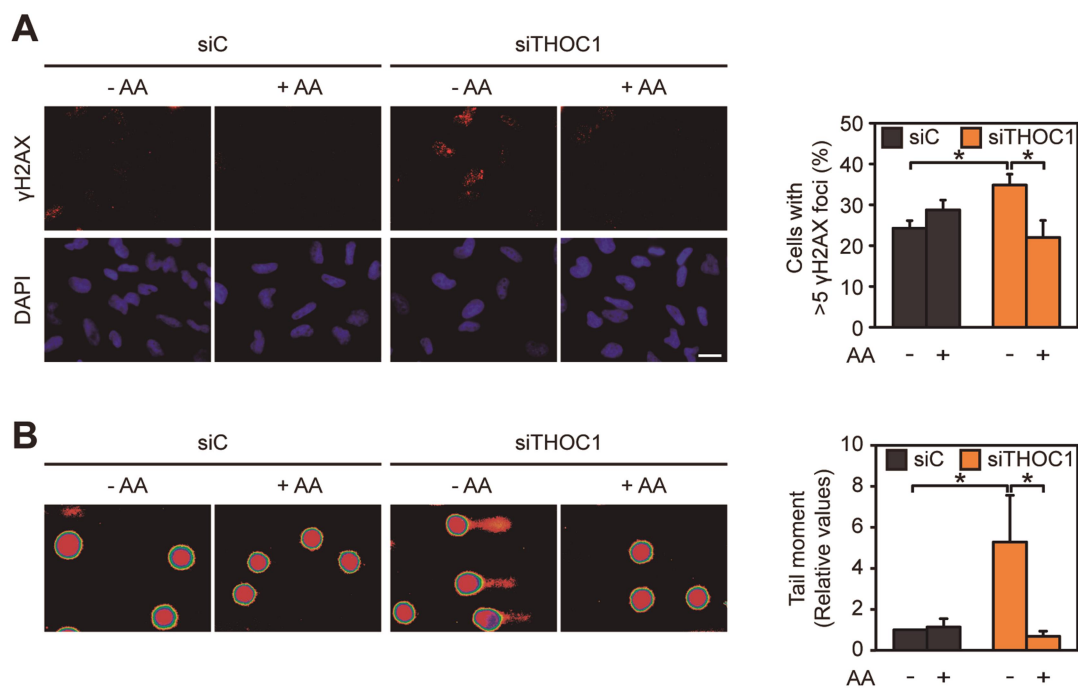
Figure R20. Effect of histone acetylation inhibition on THOC1 depletion-mediated genome instability.
(A) Immunofluorescence of γH2AX in siC and siTHOC1 depleted HeLa cells untreated or treated for 4 h with 30 μM anacardic acid (AA). Percentage of cells containing >5 γH2AX foci (n=4) is shown. Scale bar, 20 μm. **(B)** Alkaline single cell electrophoresis of THOC1-depleted cells untreated or treated with AA. Means and SEM are plotted (n=3). *, P < 0.05 as determined by Student's t-test (A) or Mann-Whitney U test (B).
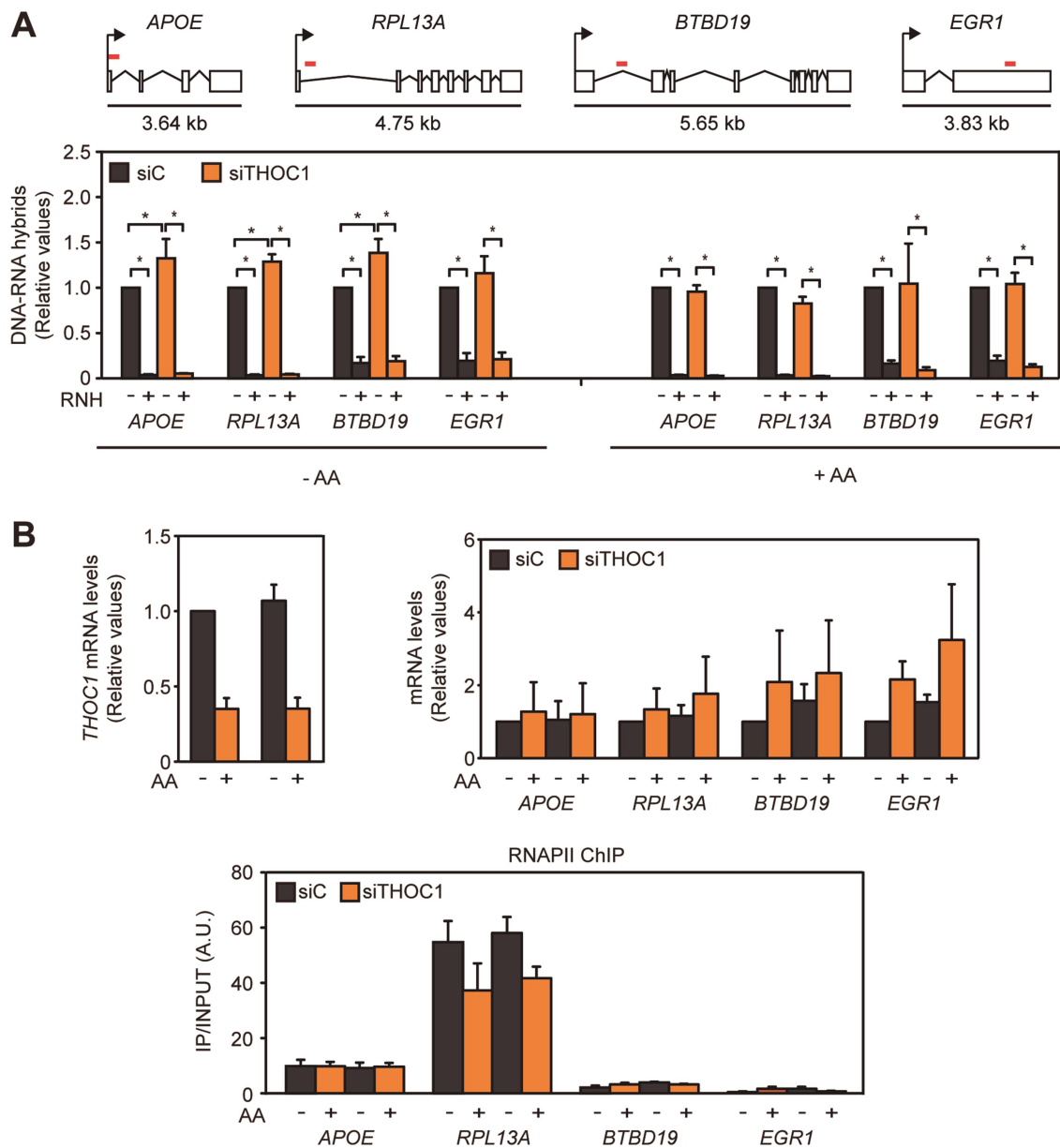
Figure R21. Analysis of R loop levels in siTHOC1 cells treated with a histone acetylation inhibitor. **(A)** DRIP-qPCR analysis in siC and siTHOC1 transfected cells untreated or treated with AA. Schematic diagrams of *APOE*, *RPL13A*, *BTBD19* and *EGR1* genes are depicted. Red lines indicate the regions where the qPCR analyses were performed. Signal values normalized with respect to the respective siC controls are plotted as mean ± SEM (n=3). **(B)** Validation of THOC1siRNA and relative mRNA quantification of *APOE*, *RPL13A*, *BTBD19* and *EGR1* genes by RT-qPCR in siC and siTHOC1 transfected cells untreated or treated with AA (upper panel). mRNA expression values were normalized to the expression of the *HPRT* housekeeping gene. RNAPII ChIP analysis at the same regions analyzed by DRIPq-PCR (lower panel). Data are plotted as means and SEM (n=3).*, $P < 0.05$ (Mann-Whitney U test) (A).

2.10. Suppression of R loop-induced chromatin compaction by Sin3A depletion

During the last years cumulative evidence indicate that chromatin organization could be a key factor in R loops-mediated genome instability by a dual role. On one hand chromatin organization plays an important role in preventing R-loop accumulation, as previously reported for the chromatin-reorganizing complex FACT (Herrera-Moyano et al. 2014). On the other hand, recent work indicates that R loops trigger chromatin compaction, which in turn would lead to genome instability (Castellano-Pozo et al. 2013). Since Sin3A is a chromatin remodeler complex we wondered whether in addition to help prevent R-loop formation, as our data show, this complex could be also involved in the chromatin compaction that R loops trigger. In order to test this possibility we first analyzed the percentage of cells with histone H3 phosphorylation at serine 10 (H3S10-P) foci, a mark of chromatin condensation that has been previously linked to R loops (Castellano-Pozo et al. 2013), in siSAP130, siSIN3 and siTHOC1 cells. Immunofluorescence with anti-H3S10-P antibody showed an increase in H3S10-P foci in THOC1-depleted cells, as it has been previously published (Figure R22), however, no significant differences were found in siSAP130 and siSIN3 with respect to siC control cells. Interestingly, siRNA depletion of SAP130 or SIN3 suppressed the increase in H3S10-P foci caused by THOC1 depletion (Figure R22).

Next, we analyzed the levels of H3K9 dimethylation (H3K9me2), a heterochromatin mark also linked to R loops (Groh et al. 2014; Skourti-Stathaki et al. 2014). For this purpose we performed immunofluorescence with anti-H3K9me2 antibody and measured nuclear H3K9me2 signal intensity. As shown in Figure R23, THOC1 depletion clearly led to an increase in H3K9me2 signal intensity whereas no significant differences were observed in siSAP130 and siSIN3 with respect to siC control cells (Figure R23). Interestingly, the increase in H3K9me2 signal caused by THOC1 depletion was partially suppressed by SAP130 depletion and completely suppressed by SIN3 depletion. These results were not due to a suppression of RNA-DNA hybrids levels, since high levels of S9.6 signal were detected by immunofluorescence after siRNA single and double depletion of THOC1 and Sin3A complex subunits (Figure R17). Finally, to assess whether suppression in chromatin compaction could lead to a reduction of DNA damage we performed immunofluorescence of γH2AX. SAP130 depletion completely suppressed the increase in γH2AX foci caused by THOC1 depletion while such suppression was not observed in

siSIN3-siTHOC1 cells (Figure R24). This could be due to the fact that depletion of SIN3 itself has a major impact on DNA damage than SAP130 depletion.

Taken together, these results suggest that Sin3A complex could participate in the process of chromatin compaction triggered by the R loops.
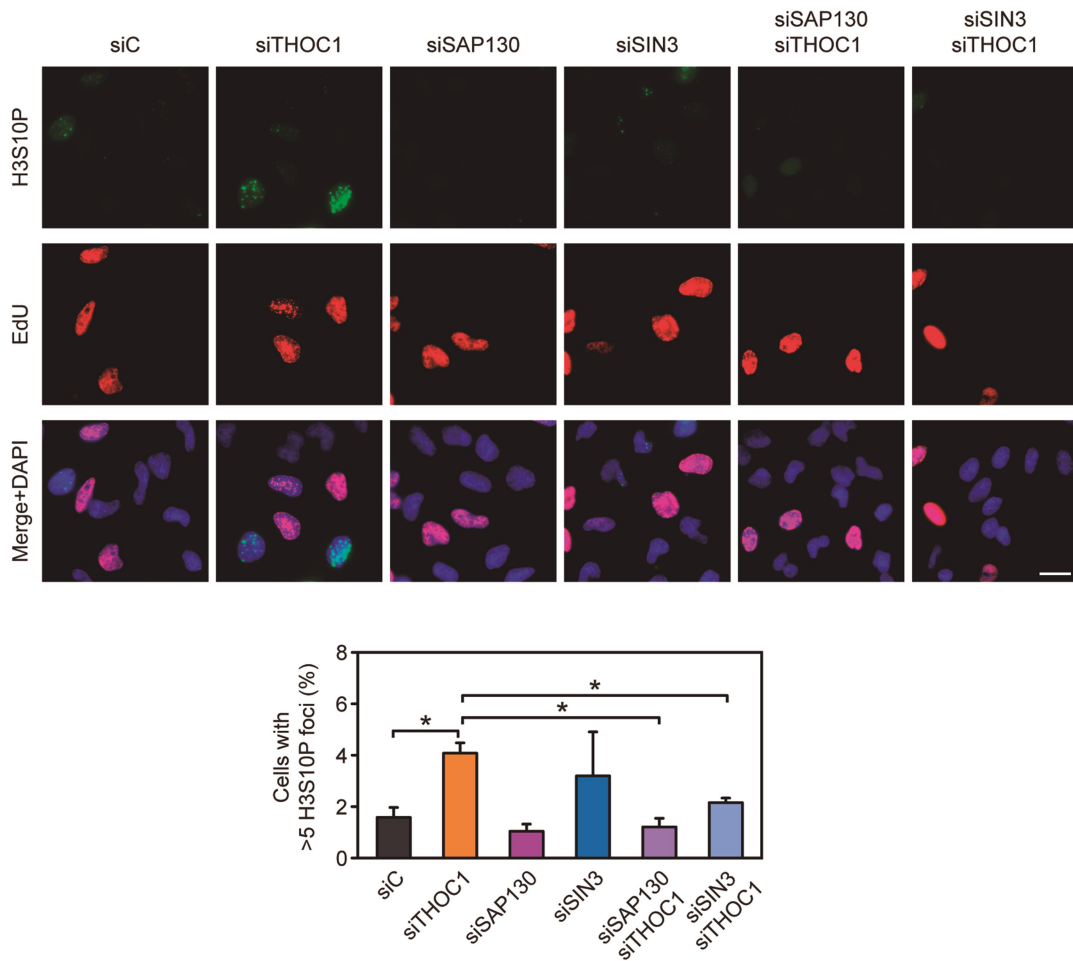
Figure R22. Analysis of H3S10-P in THOC1-depleted cells and cells depleted of THOC1 and Sin3A complex subunits.
Immunofluorescence of H3S10-P in HeLa cells transfected with the indicated siRNAs. Percentage of cells containing >5 H3S10-P foci excluding S-phase cells and mitotic cells, identified by EdU labeling or DAPI staining respectively. Data are plotted as means and SEM (n=3). Scale bar, 20 μm. *, P < 0.05 (Student's t-test).
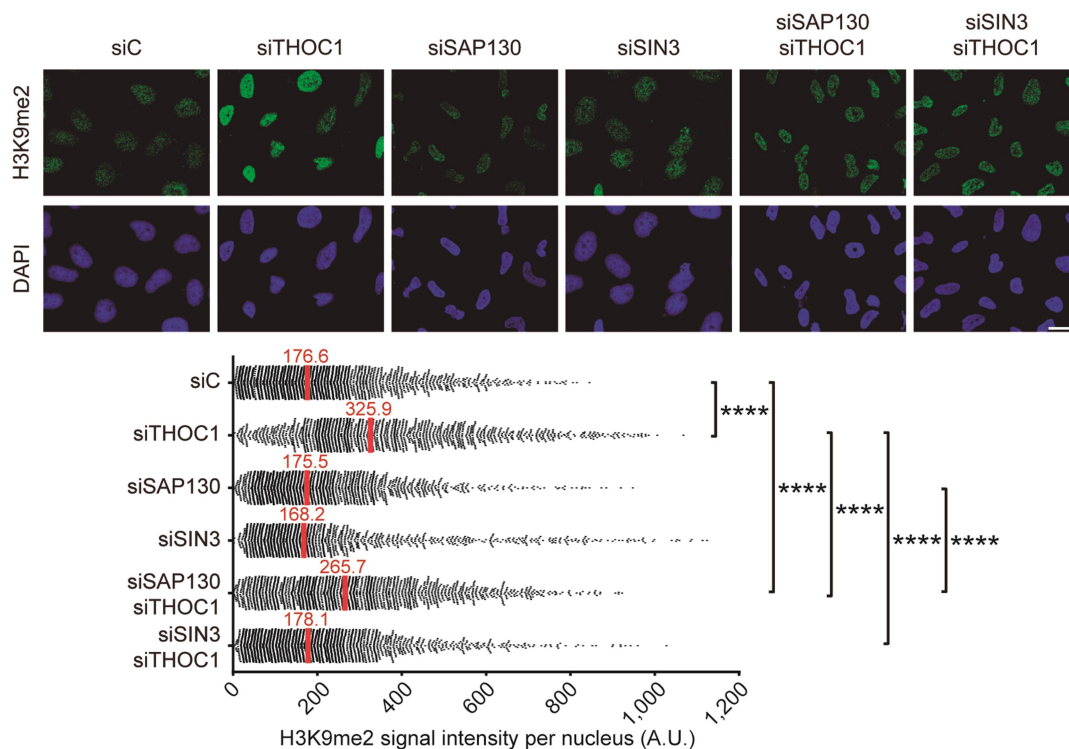
Figure R23. High levels of H3K9me2 caused by THOC1 depletion and suppression by double depletion of Sin3A complex subunits.

Immunofluorescence of H3K9me2 in HeLa cells transfected with the indicated siRNAs. The median of the H3K9me2 signal intensity per nucleus is shown (n=3). Nuclei were stained with DAPI. Scale bar, 20 μm. ****, $P < 0.0001$ (Mann-Whitney U test, two-tailed).
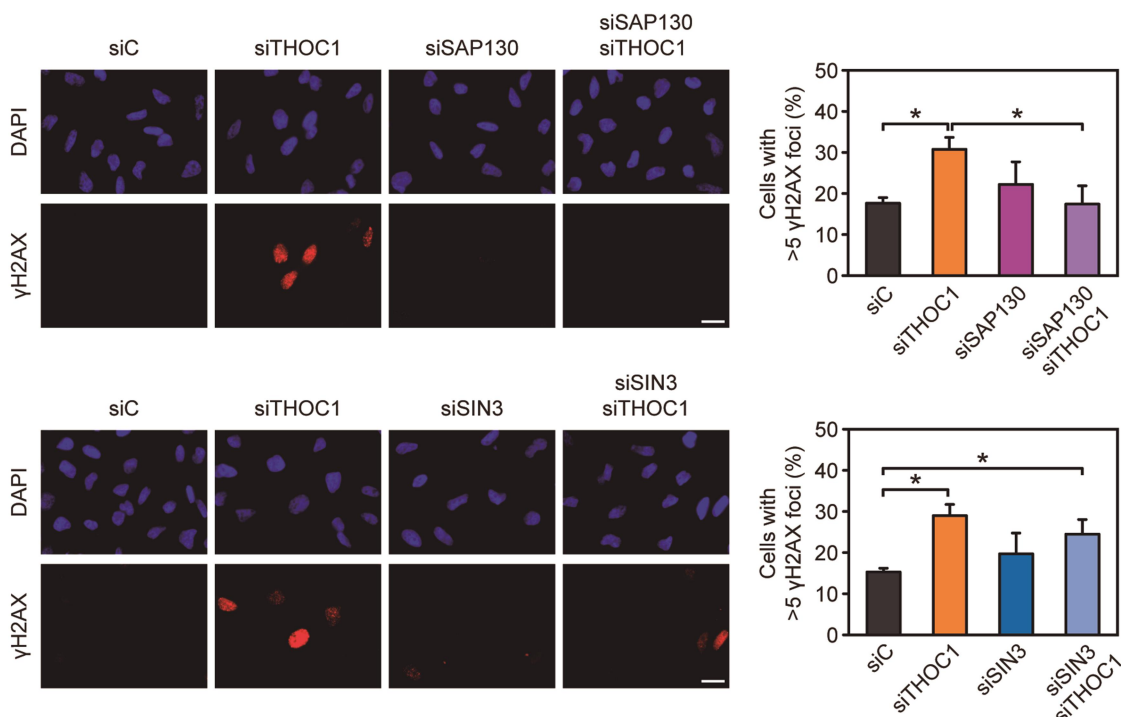
Figure R24. Suppression of the γH2AX foci accumulation of THOC1-depleted cells by SAP130 depletion.

Immunofluorescence of γH2AX in HeLa cells transfected with the indicated siRNAs. Percentage of cells containing >5 γH2AX foci (n=5) is shown. Nuclei were stained with DAPI. Scale bar, 20 μm. *, $P < 0.05$ (Student's t-test).

99

3. ROLE OF MFAP1 IN THE MAINTENANCE OF GENOME STABILITY

In addition to SAP130, the two hybrid screening and subsequent validations in human cells revealed MFAP1 as a new partner of THOC1. The interaction of the spliceosome-associated factor MFAP1 with THOC1 is consistent with the predicted role of THO in mRNA biogenesis. We carried out a functional analysis using RNAi and a genome-wide analysis with microarrays to explore whether this interaction has any functional relevance on gene expression, splicing and on the maintenance of genome stability.

3.1. Localization of the human pre-mRNA splicing factor MFAP1 at nuclear speckles

MFAP1 (Microfibrillar-associated protein 1) was initially identified as a putative extracellular matrix protein (Horrigan et al. 1992). However, MFAP1 has been recently found as a nuclear protein associated to the spliceosome, as determined in global screenings (Makarov et al. 2002; Hegele et al. 2012). Then we analyzed the association of MFAP1 with the spliceosome by immunofluorescence assays using the SC35 antibody, raised specifically against a serine rich protein and used as a marker of splicing nuclear speckles. As can be seen in Figure R25 MFAP1 and SC35 colocalize, confirming the cytological association of MFAP1 with the spliceosome.
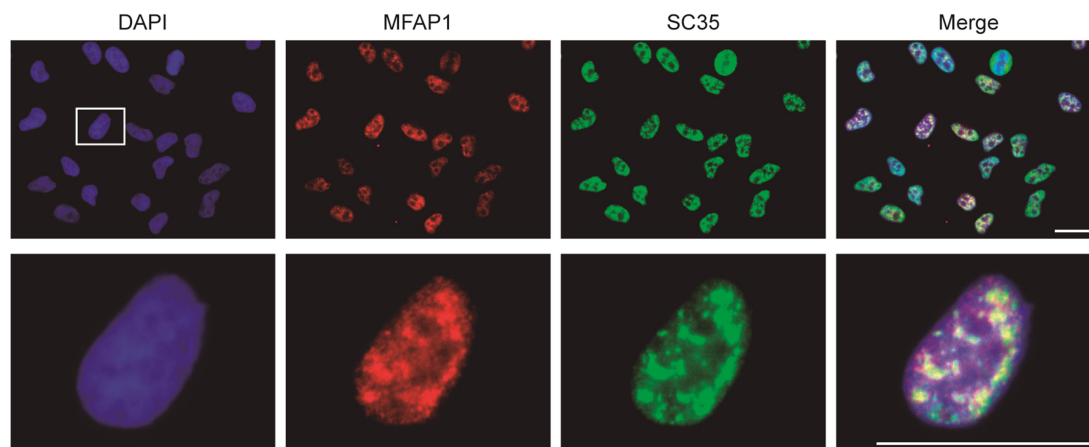
Figure R25. MFAP1 localizes at nuclear speckles.
Colocalization of endogenous MFAP1 and splicing speckles marker visualized by the SC35 antibody. HeLa cells were labeled by double-immunofluorescence with specific antibodies. Lower panels are a magnification of the region indicated in the first upper panel. Nuclei were stained with DAPI. Scale bars, 20 μm.

3.2. Proliferation and cell cycle defects in MFAP1-depleted cells

In addition to its association with the spliceosome, little is known about other possible roles of MFAP1. It has been reported that lack of functional MFAP1 affects cell cycle progression, especially the G2/M transition in *Drosophila*, and in human cells MFAP1 was isolated in a siRNA screen for cell division related genes (Kittler et al. 2004; Andersen and Tapon 2008). We started our functional studies by analysis of proliferation and cell cycle progression in HeLa cells transfected with siRNAs against MFAP1. As shown in Figure R26, siRNA-transfected HeLa cells showed a clear reduction in MFAP1 protein and mRNA levels. First, we performed proliferation curves of siC and siMFAP1-transfected HeLa cells. Cell proliferation was severely affected in cells depleted of MFAP1 (Figure R27A). Next, we analyzed the cell cycle distribution of asynchronously growing MFAP1-depleted cells. FACS analysis revealed an accumulation of cell in G2 phase together with a decrease of cells in S phase in MFAP1-depleted cells (Figure R27B). In order to analyze in detail the cell cycle progression we synchronized the cells in G1 phase by double thymidine block and collected samples at different time points after release. Results showed that MFAP1 depletion slows down cell cycle progression as well as leads to an increase in the percentage of arrested cells in G2 (Figure R27C). Due to these alterations in proliferation and cell cycle we decided to analyze apoptosis. An increase in apoptotic cells, as measured by sub-G_1 DNA content, was found in siMFAP1-transfected HeLa cells (Figure R27D). Therefore, these results confirm that depletion of MFAP1 leads to defective proliferation and cell cycle in human cells, which is consistent with previous studies in *Drosophila* model and genome-wide siRNA screening of human cells (Kittler et al. 2004; Andersen and Tapon 2008).
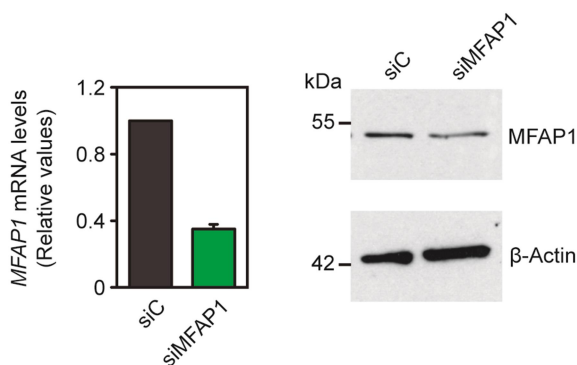
Figure R26. Analysis of siRNA-mediated MFAP1 depletion.
Relative *MFAP1* mRNA quantification measured by RT-qPCR in siRNA-transfected HeLa cells (left panel). *MFAP1* mRNA expression values were normalized respect to the expression of the *HPRT* housekeeping gene and means and SEM of three independent experiments are depicted. Western blot analysis showing MFAP1 expression in HeLa cells treated with control and MFAP1 siRNAs (right panel). β-actin protein was used as a loading control.
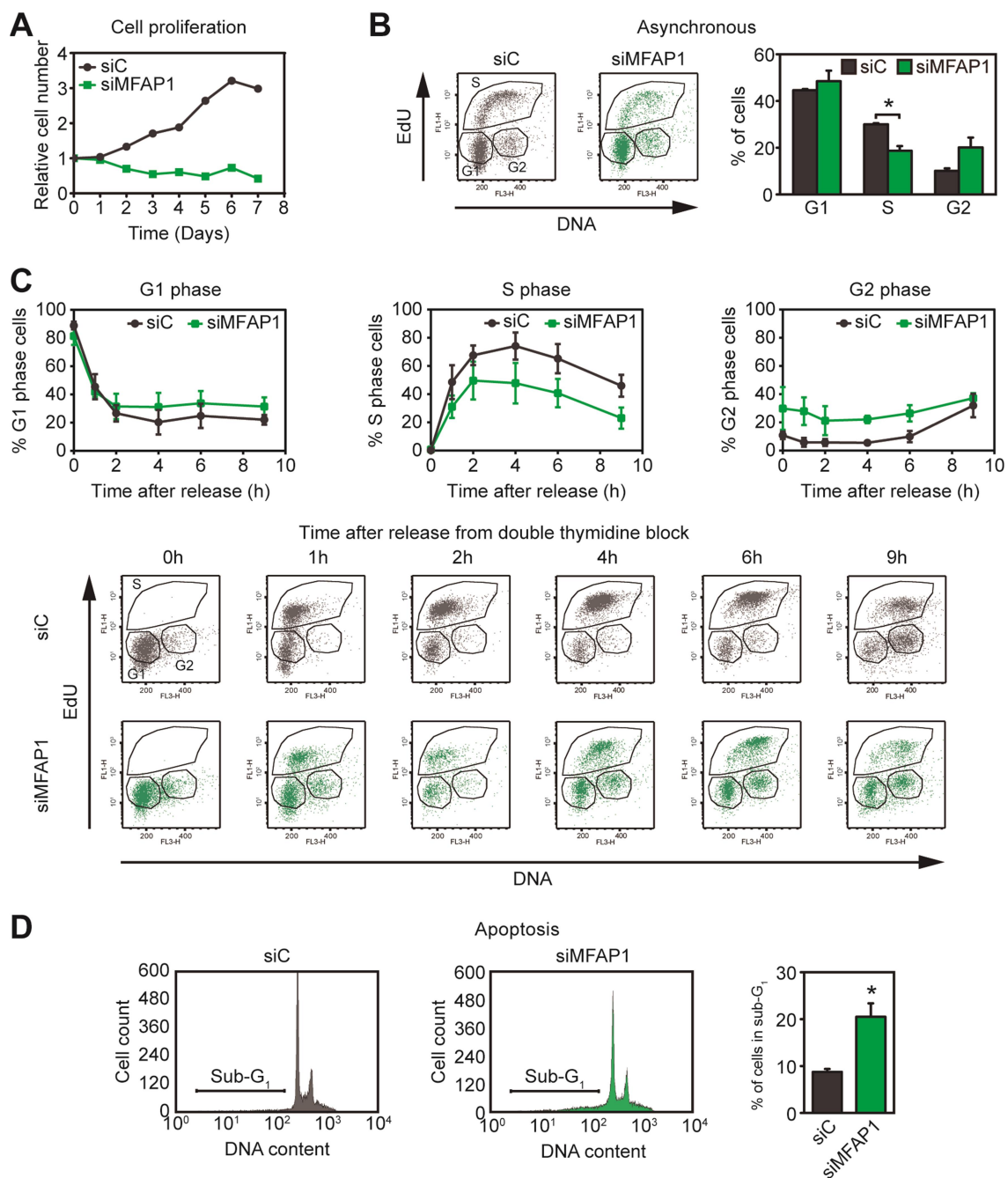
Figure R27. Defects in cell proliferation and cell cycle progression in MFAP1-depleted cells.
(A) Cell proliferation curves of HeLa cells after MFAP1 depletion. Cells were plated 48h after siRNA depletion (day 0). The relative number of living cells was determined by measuring the metabolic activity of mitochondrial dehydrogenases. Values were normalized to the day 0 value. A representative experiment is shown. **(B)** FACS analysis of asynchronously growing HeLa cells depleted of MFAP1 after 20 min of incubation with 20 µM EdU. Cells were separated based on their DNA content (7-AAD staining) and replication activity was analyzed by EdU incorporation. **(C)** FACS analysis of cell cycle progression in control and MFAP1-depleted cells after G1-phase synchronization by double thymidine block. Samples were collected at the indicated times (hours) after release and processed for flow cytometry. G1, S and G2 cell population were separated based on their DNA content and replication activity as in (B). The percentages of cells in G1, S and G2 phases are depicted (B, C). **(D)** Analysis of apoptosis in siC- and siMFAP1-transfected HeLa cells as determined by FACS analysis of cells displaying subG1-DNA content. Means and SEM of three independent experiments are depicted (B, C, D). See Materials and Methods 7-9.

3.3. Analysis of the recruitment to chromatin of RNAPII and different factors of the mRNA processing in MFAP1-depleted cells

To investigate more closely the relationship between THOC1 and MFAP1 we analyzed the recruitment of this factor to active chromatin. ChIP analysis in *β-actin* gene show that MFAP1 is recruited along transcribed genes (Figure R28A, B), as it has been reported for the THO complex (Luna et al. 2012).

Since it is known that THO/TREX complex is required for efficient transcription (Chavez et al. 2000; Dominguez-Sanchez et al. 2011a) we wondered whether MFAP1 depletion could also have an impact on transcription. ChIP analysis along *β-actin* gene showed similar levels of total RNAPII in control and MFAP1-depleted cells (Figure R29A, B), suggesting that transcription is not affected. Considering that MFAP1 is involved in mRNA metabolism we asked whether it could be important for the recruitment of other factors related to mRNA processing. For this purpose we performed ChIP analysis of THOC1, FIP1 and CPSF6 factors, which participate at different steps of mRNA biogenesis. FIP1 is part of the cleavage and polyadenylation specificity factor (CPSF) multisubunit complex (Kaufmann et al. 2004) and CPSF6 is a subunit of the Cleavage Factor I_m (CF I_m), which is a component of the 3′ end processing complex, that participates in the cleavage reaction and stimulates mRNA export (Ruepp et al. 2009). No significant differences were found in THOC1, FIP1 and CPSF6 signals in MFAP1-depleted cells, suggesting that the recruitment of these factors is not affected by MFAP1 depletion (Figure R29). Interestingly, a decrease in FIP1 recruitment at 3′end of the gene was observed in siMFAP1 cells, which could suggest a connection of MFAP1 with 3′end mRNA processing steps (Figure R29D).
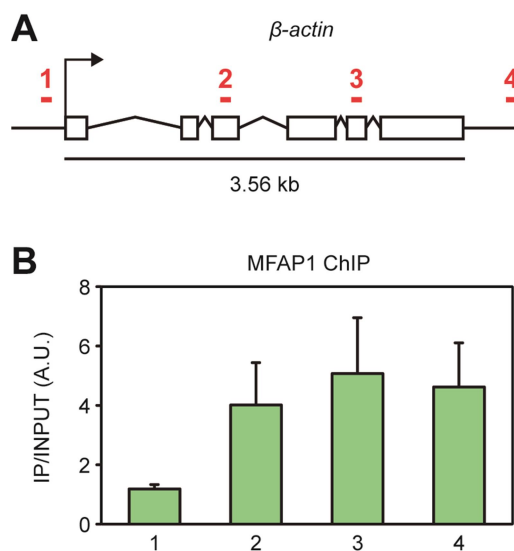
Figure R28. Recruitment of MFAP1 to chromatin along an actively transcribed gene.
(A) Schematic diagram of *β-actin* gene, with the exons depicted as open boxes. The arrow indicates the start of transcription. Numbered red lines indicate the regions where the ChIP analyses were performed. (B) ChIP analysis of MFAP1 within the *β-actin* gene in HeLa cells. Values represent the ratios of precipitated DNA (IP) to input DNA (INPUT). The positions of amplified regions are indicated on the x axis. Means and SEM of three independent experiments are depicted.
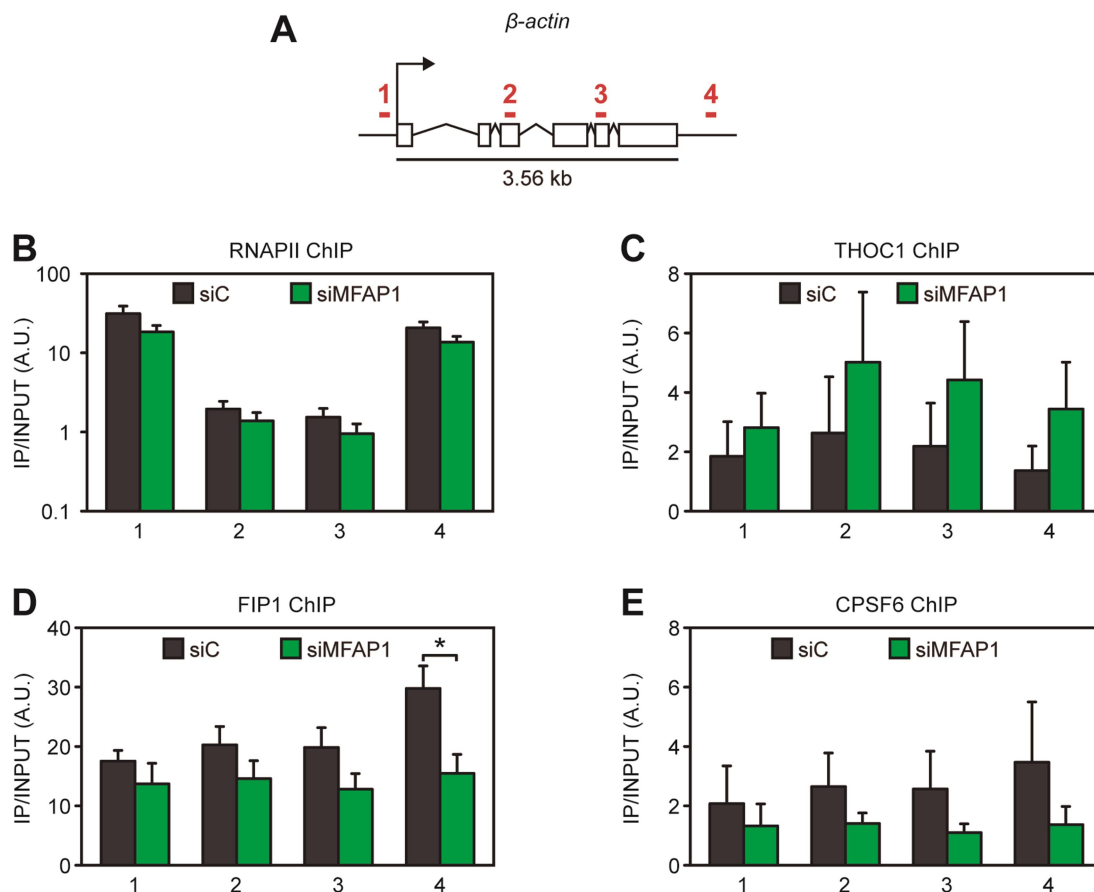
Figure R29. Analysis of the recruitment to chromatin of RNAPII and different factors of mRNA metabolism in MFAP1-depleted cells.
(A) Schematic diagram of *β-actin* gene. (B) Quantitative ChIPs analysis performed with antibody against total RNAPII in control and MFAP1-depleted HeLa cells within the *β-actin* gene. (C, D, E) ChIP analyses of THOC1, FIP1 and CPSF6 in siC- and siMFAP1-transfected HeLa cells within the *β-actin* gene, respectively. Means and SEM of three independent experiments are depicted. *, P < 0.05 (Student's t-test) (C). Other details as in Figure R28.

3.4. Genome instability in MFAP1-depleted cells

Next, to investigate whether MFAP1 has also a role in the maintenance of genome integrity we first determined γH2AX and 53BP1 foci, as markers of DNA damage/repair, in HeLa cells depleted of MFAP1. An increase in the percentage of cells with γH2AX foci was observed in siMFAP1 with respect to siC control cells (3.3 fold), however, no differences could be found in 53BP1 foci percentages after MFAP1 depletion (Figure R30A). Next, we performed more sensitive assays as single-cell electrophoresis that revealed a significant increase in tail moment in MFAP1 siRNA depleted cells (5.5 fold and 2.1 fold in alkaline or neutral single cell electrophoresis assays, respectively) indicating an accumulation of DNA breaks (Figure R30B, C).
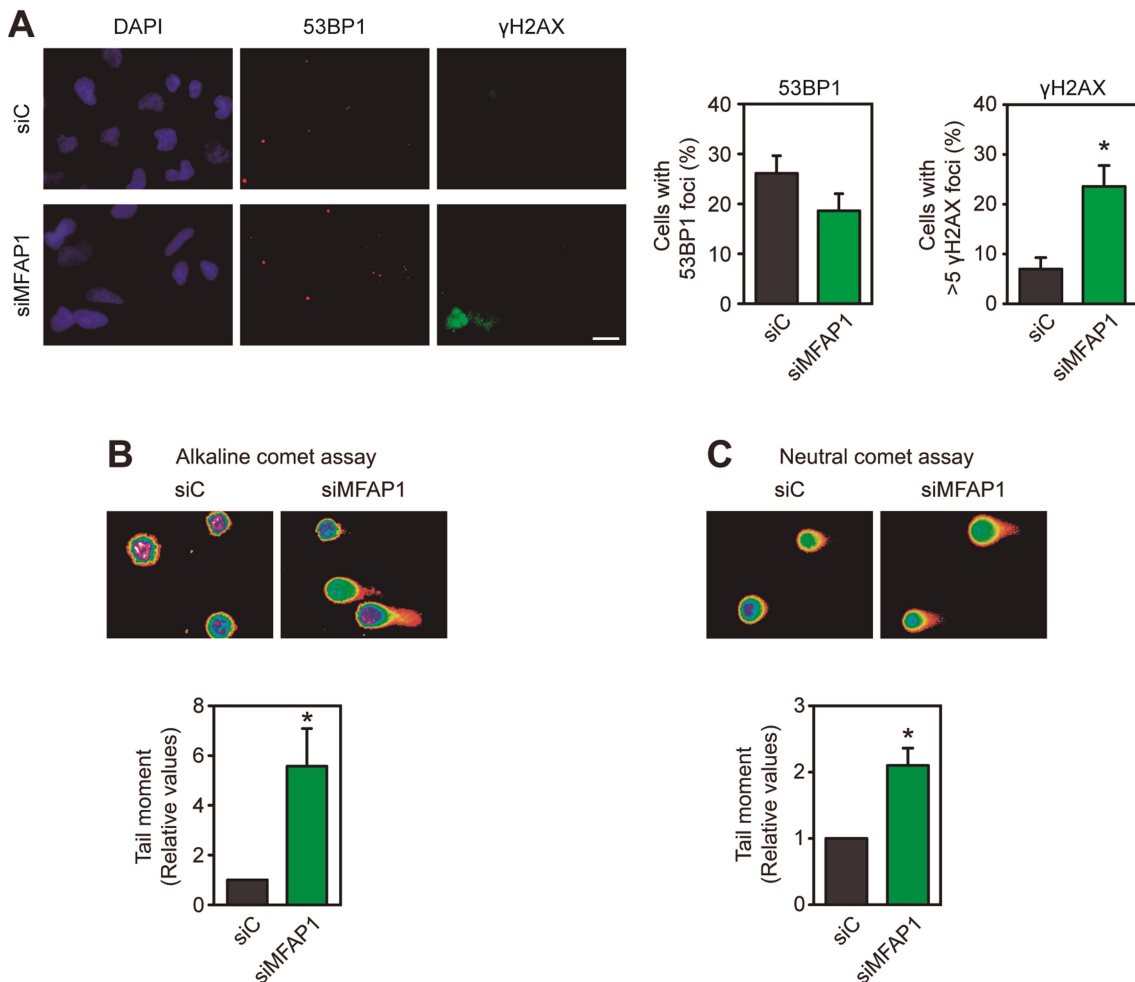
Figure R30. MFAP1 depletion effect on genome instability.
(A) Detection of 53BP1 and γH2AX foci by immunofluorescence in siC (control) and siMFAP1 transfected HeLa cells. The percentage of cells containing 53BP1 foci and cells with >5 γH2AX foci (n=3) is shown. Nuclei were stained with DAPI. Scale bar, 20 μm. **(B)** Alkaline single cell electrophoresis in control and MFAP1-depleted cells. **(C)** Neutral single cell electrophoresis of HeLa cells transfected with the indicated siRNAs. Relative comet-tail moments are plotted as means and SEM (n=3). *, $P < 0.05$ as determined by Student's t-test (A) or Mann-Whitney U test (B, C).

Then, we wondered whether these breaks could also be seen in the form of anaphase bridges, a well-known biomarker of chromosomal instability resulting from unrepaired DNA breaks or of interlinked sister-chromatid intermediates generated by replication stress. The percentage of anaphase cells containing bridges was significantly increased after MFAP1 depletion (Figure R31A). We also examined sister chromatin exchanges (SCEs) events, another sensitive indicator for genome instability. Although no differences in SCEs events were found in siMFAP1 cell (Figure R31B), we cannot rule out replication and recombination problems. From these experiments we conclude that MFAP1 is required for the maintenance of genome integrity.
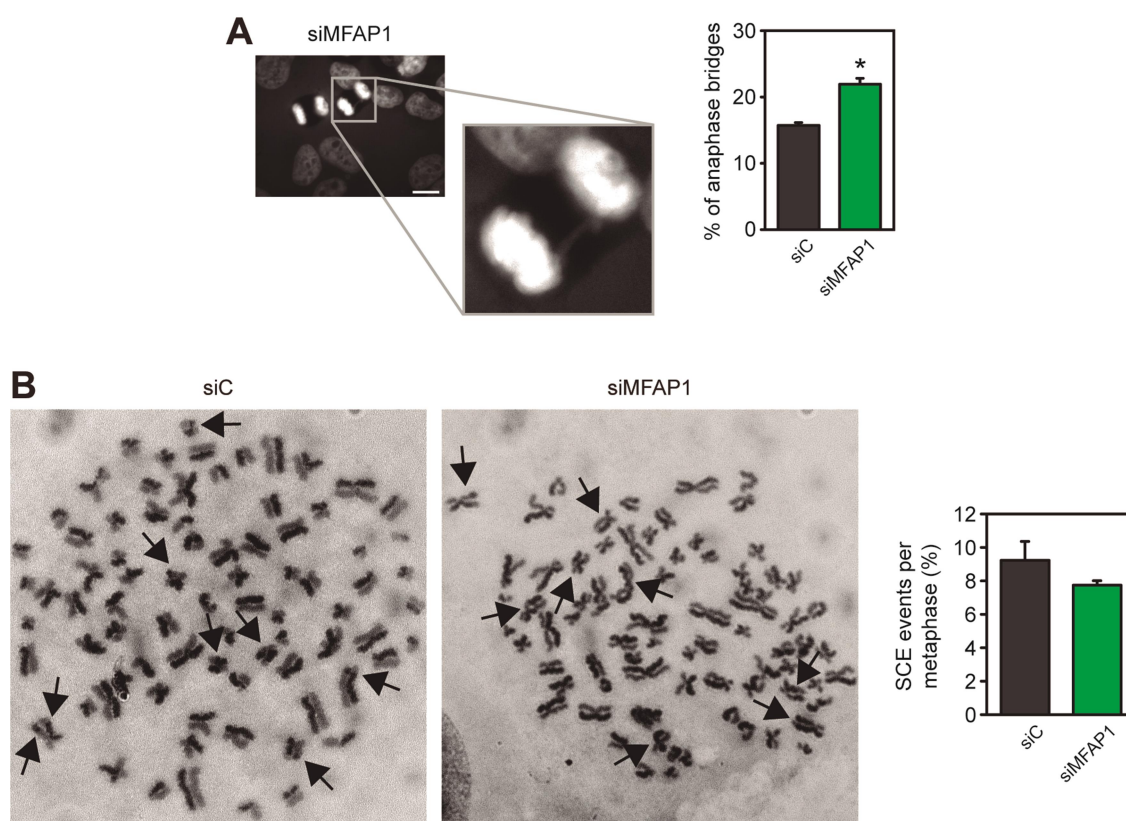
Figure R31. Analysis of anaphase bridges formation in MFAP1-depleted cells.
(A) Anaphase bridges in siC- and siMFAP1-transfected HeLa cells. Microphotography shows an example of anaphase bridge in MFAP1-depleted cells. The percentages of anaphase cells with anaphase bridges are plotted. Means and SEM of three independent experiments are depicted. Scale bar, 20 μm.
(B) Metaphase spreads of HeLa cells transfected with control and MFAP1 siRNAs. Sister chromatid exchange (SCE) events are indicated by arrows. The percentages of SCE events per metaphase are plotted. Means and SEM of three independent experiments are depicted. *, $P < 0.05$ (Student's t-test) (A).

3.5. Transcription and R loop-independent genome instability in MFAP1-depleted cells

We next investigated whether, as in the case of THO, genome instability observed after MFAP1 depletion is related to transcription and R loops. For this purpose we performed alkaline single-cell electrophoresis in the presence of the transcription inhibitor cordycepin. However, no significant reduction of DNA breaks levels was observed in cells depleted of MFAP1 factor (Figure R32A). Consistently, overexpression of RNase H1 did not suppress DNA breaks in MFAP1-depleted cells, as determined by γH2AX immunofluorescence, indicating that DNA breaks were not due to co-transcriptional RNA-DNA hybrids (Figure R32B). These results suggest that the effect of MFAP1 on genome instability is not directly mediated by its interaction with THO, since such genome instability is not transcription-dependent as it would be expected.
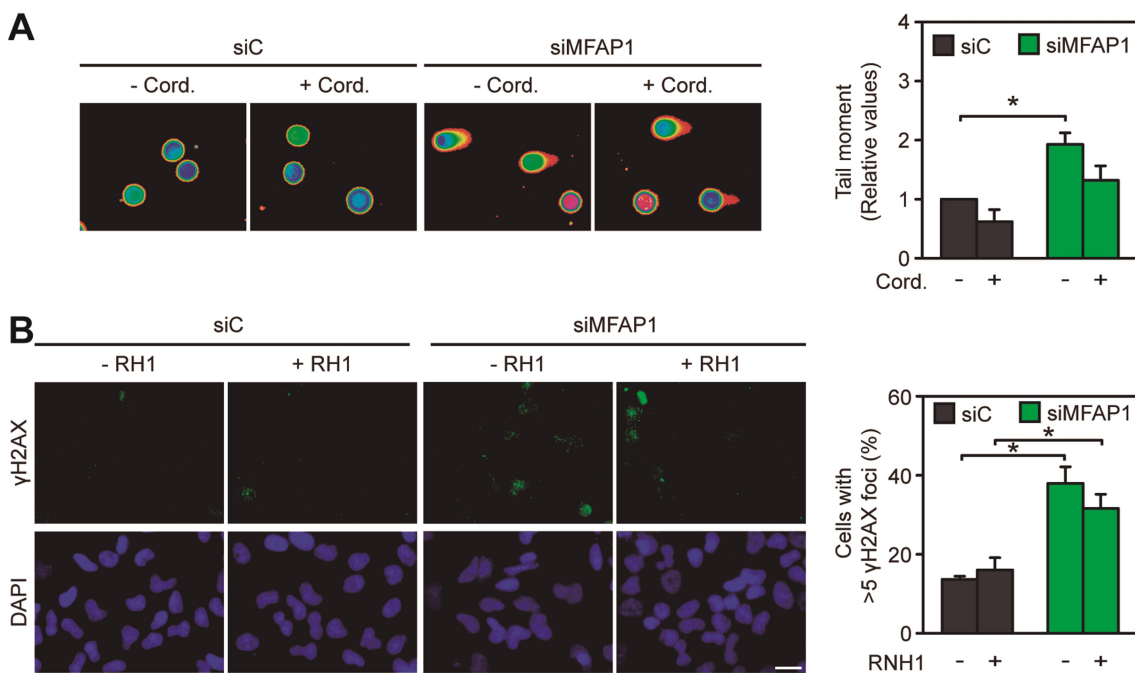
Figure R32. Dependency of the genome instability of MFAP1-depleted cells on transcription or R-loops.
(A) Alkaline comet assay in siC and siMFAP1 transfected HeLa cells untreated or treated with 50 μM cordycepin for 4 h. Relative comet-tail moments are plotted (n=3) as means and SEM. **(B)** Immunofluorescence of γH2AX after siC and siMFAP1 transfection with or without RNase H1 (RNH1) overexpression. Cells were transfected with empty pcDNA3 vector (-RNH1) or pcDNA3-RNaseH1 (+RNH1) during 24 h. Data are plotted as mean and SEM (n=3). *, P < 0.05 as determined by Mann-Whitney U test (A) or Student's t-test (B).

3.6. Microarray analyses of gene expression and alternative splicing in MFAP1-depleted cells

To understand how MFAP1 depletion leads to genome instability we next analyzed the global gene expression and alternative splicing profiles of control and MFAP1-depleted cells, using the Human Transcriptome Array (HTA) 2.0 from Affymetrix. This array contains >6.0 million distinct probes covering coding and non-coding transcript. 70% of the probes on this array cover exons for coding transcripts, and the remaining 30% of probes on the array cover exon-exon junctions, splice junctions and non-coding transcripts (for more details see Materials and Methods 15). We performed a microarray analysis with three different biological replicates of HeLa cells transfected with siC, siMFAP1 or siTHOC1 for 72 hours. From a total of 67528 genes analyzed, gene expression analysis of HTA 2.0 data identified 904 differentially expressed genes (P <0.05 and |linear fold change| >1.5), of which 359 (39.7%) were down-regulated and 545 (60.3%) were up-regulated in siMFAP1 cells (Figure R33A and Appendix 1). Regarding down-regulated transcripts, the majority were mRNAs (49%) or ncRNAs (36.2%), whereas among the up-regulated transcripts, the category of U snRNAs (40%) was significant represented (Figure R33B). Up-regulation of U snRNAs, which are part of spliceosome and participate in splicing, is consistent with a role of MFAP1 in pre-mRNA splicing. Gene ontology (GO) analyses were performed for those down- and up-regulated genes with assigned GO term (see Materials and Methods 15.1). A high variety of biological processes were found enriched in the down-regulated genes (P <0.05). Several of them were related to chromatin and gene silencing as shown in Table R3. However, down-regulated genes within these categories belong mostly to histone gene clusters and therefore were redundant. Interestingly, although not with the highest significance, down-regulated genes were enriched in genes involved in DNA repair such as *FANCD2*, a subunit of the Fanconi anemia complex involved in DNA repair and *RNASEH2A,* the catalytic subunit of the endonuclease RNaseH2, that participates in DNA replication and specifically degrades the RNA of RNA-DNA hybrids. GO analyses of the up-regulated genes revealed an enrichment of genes involved mainly in cellular response to hypoxia and regulation of apoptotic process as shown in Table R4. Up-regulation of apoptotic genes is consistent with our data (Figure R27D) and previous studies (Andersen and Tapon 2008).
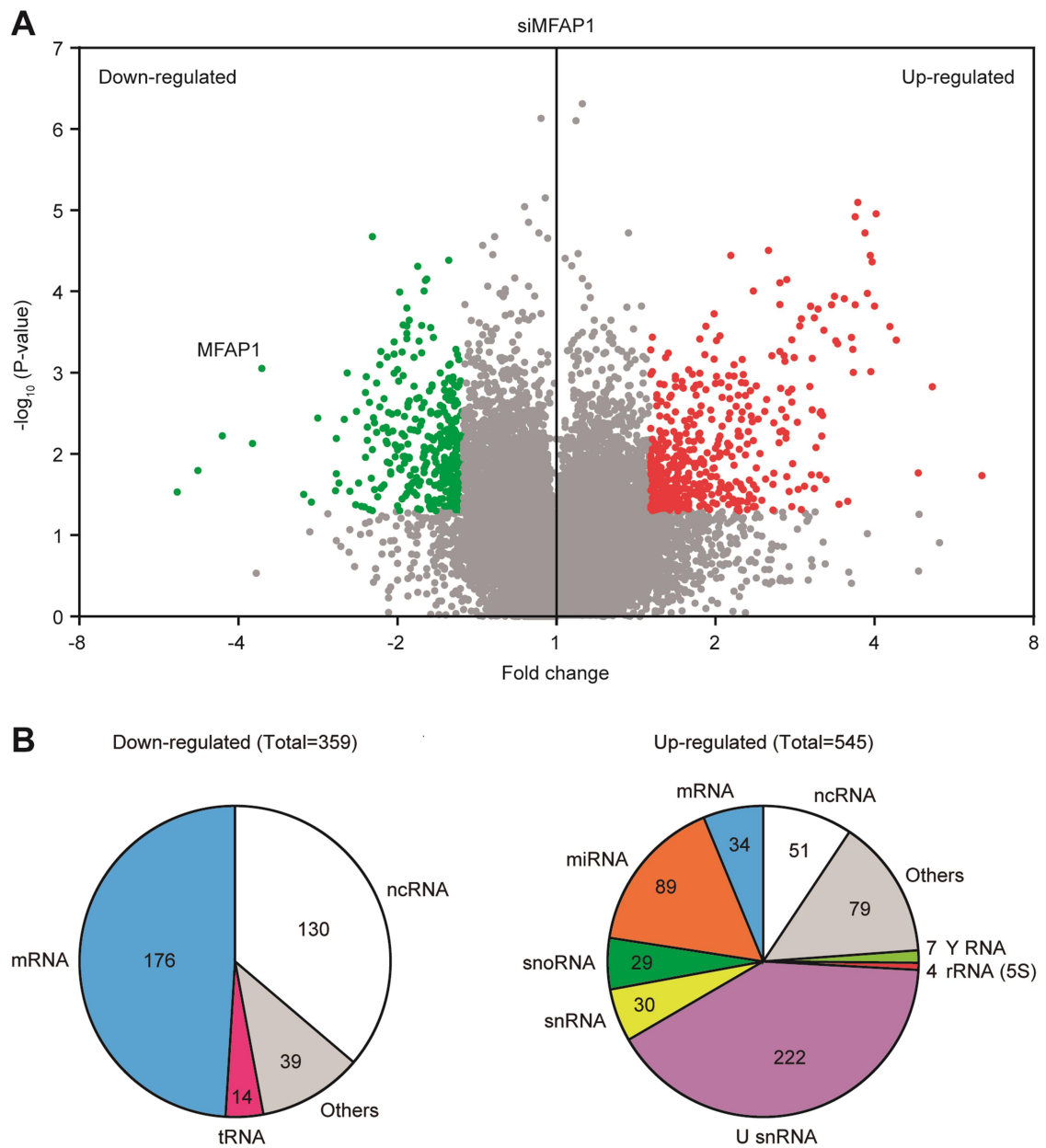
Figure R33. HTA 2.0 gene expression analysis of MFAP1-depleted cells.
(A) Volcano plot of differential gene expression in MFAP1-depleted cells compared to siC control cells. Green dots correspond to down-regulated genes and red dots correspond to up-regulated genes (P <0.05 and |linear fold change| >1.5) (see Appendix 1). **(B)** Pie charts showing the distribution of the different transcript types down-regulated (left) and up-regulated (right). Transcript numbers in each group are also indicated.

Table R3. Summary of the most representative GO terms enriched among down-regulated genes in siMFAP1 cells.

GO Biological Process	Count (Total 146)*	Genes	Fold enrichment	P-value	Q-value (Benjamini)
		Down-regulated			
GO:0006342 Chromatin silencing	22	HIST1H2AB, HIST2H3A, HIST2H2AA3, HIST2H2AA4, HIST1H4K, HIST1H2AG, HIST1H2AD, HIST2H4A, HIST2H3C, HIST2H4B, HIST2H3D, HIST1H4B, HIST1H3A, HIST1H2AI, HIST1H3B, HIST1H3D, H3F3B, HIST1H4D, HIST1H3F, HIST1H2AM, HIST1H3G, HIST1H3H	21.32	5.12E-22	1.34E-18
GO:0006306 DNA methylation	15	HIST2H3A, HIST1H4K, HIST2H4A, HIST2H3C, HIST2H4B, HIST2H3D, HIST1H4B, HIST1H3A, HIST1H3B, H3F3B, HIST1H3D, HIST1H4D, HIST1H3F, HIST1H3G, HIST1H3H	21.62	5.91E-15	1.03E-12
GO:0031047 Gene silencing by RNA	18	HIST2H3A, HIST1H4K, NUP88, IPO8, HIST2H4A, HIST2H3C, HIST2H4B, HIST2H3D, SEH1L, HIST1H4B, HIST1H3A, HIST1H3B, HIST1H3D, H3F3B, HIST1H4D, HIST1H3F, HIST1H3G, HIST1H3H	13.49	1.87E-14	2.87E-12
GO:0032200 Telomere organization	12	HIST1H4K, HIST1H4B, HIST1H3A, HIST1H3B, H3F3B, HIST1H3D, HIST1H4D, HIST1H3F, HIST2H4A, HIST1H3G, HIST1H3H, HIST2H4B	9.5	5.18E-08	4.84E-06
GO:0002230 Positive regulation of defense response to virus by host	7	HIST1H3A, HIST1H3B, PNPO, HIST1H3D, HIST1H3F, HIST1H3G, HIST1H3H	6.4	7.81E-04	0.02618
GO:0009267 Cellular response to starvation	10	SEH1L, IMPACT, HIST1H3A, HIST1H3B, PNPO, HIST1H3D, HIST1H3F, HIST1H3G, MYBBP1A, HIST1H3H	2.45	0.02054	0.34358
GO:0006281 DNA repair	10	EGFR, KDM1A, HIST1H4K, FANCD2, HIST1H4B, TRIM25, HIST1H4D, HIST2H4A, RNASEH2A, HIST2H4B	2.2	0.03808	0.47848

*146 genes out of 359 down-regulated genes were analyzed by Gene Ontology enrichment.

Table R4. Summary of the most representative GO terms enriched among up-regulated in siMFAP1 cells.

Up-regulated					
GO Biological Process	**Count (Total 26)***	**Genes**	**Fold enrichment**	**P-value**	**Q-value (Benjamini)**
GO:0071456 Cellular response to hypoxia	4	*EPAS1, HMOX1, MDM2, NDRG1*	18.98	0.00105	0.73225
GO:0042981 Regulation of apoptotic process	8	*CDKN1A, BTG2, TNFRSF10D, HMOX1, HIPK3, DAD1, MDM2, GDF15*	3.47	0.00489	0.42820
GO:0030330 DNA damage response, signal transduction by p53 class mediator	3	*CDKN1A, MDM2, NDRG1*	17.06	0.01229	0.62054
GO:0006397 mRNA processing	4	*DDX23, BTG2, PRPF3, SMNDC1*	5.66	0.02943	0.83200
GO:0007584 Response to nutrient	3	*HMOX1, DAD1, MDM2*	9.81	0.03457	0.82875

*26 genes out of 545 down-regulated genes were analyzed by Gene Ontology enrichment.

Next, we analyzed whether splicing patterns were altered in siMFAP1 cells. For that, Splicing Index (SI) was used to detect splicing changes between MFAP1-depleted cells and siC control cells. SI is calculated for each probe set on the array and represents the ratio of exon intensities between the two experimental conditions after normalization to the gene intensities in each condition, in this case, siMFAP1 versus siC. Thus, positive and negative values of SI indicate inclusion or skipping events of a probe set, respectively (see Materials and Methods 15.1). Alternative splicing analysis of HTA 2.0 data identified 5490 alternative splicing events (ASEs) or exon probe sets alternatively spliced (P <0.01 and |Splicing Index| >2), which belonged to 3689 genes (Appendix 2). 1712 probe sets (31.2%) showed a SI <-2, whereas 3778 probe sets

(68.8%) showed a SI >2. Analysis of the alternative splicing type revealed that cassette exon (49%) was the most common type, while no case of mutually exclusive exons was found (Figure R34A, B). Importantly, these alterations in splicing patterns in siMFAP1 cells seem not to be due to transcription problems, since no relationship was found between gene expression levels and splicing indexes (Figure R34C).
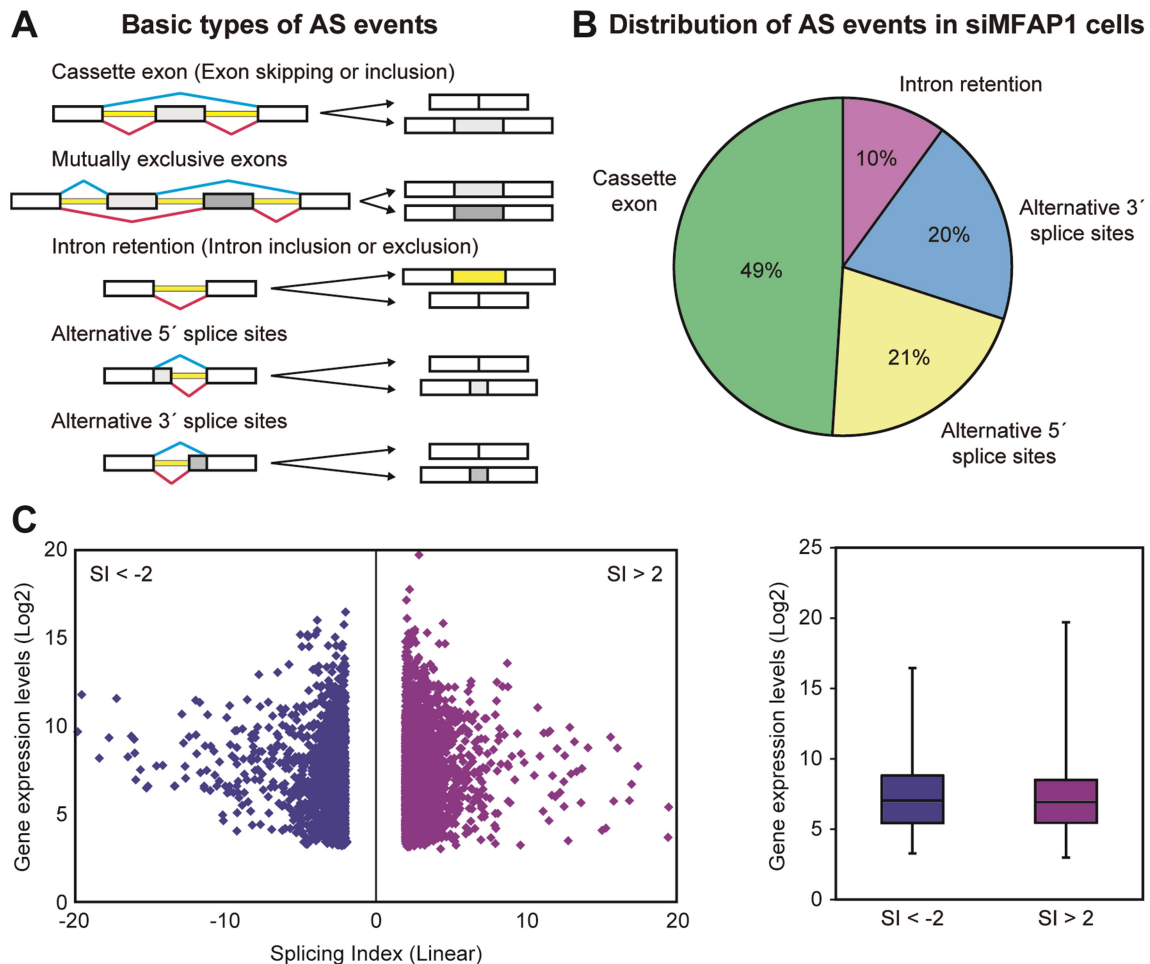
Figure R34. HTA 2.0 alternative splicing analysis in MFAP1-depleted cells.
(A) Scheme of the five basic modes of alternative splicing including: cassette exon (exon skipping or inclusion), mutually exclusive exons, intron retention (intron inclusion or exclusion), alternative 5´splice sites and alternative 3´splice sites. Exons are depicted as white boxes and alternative exons are indicated as grey boxes. Introns are represented as yellow lines and splicing events are indicated with diagonal lines. **(B)** Pie chart showing the alternative splicing type distribution in siMFAP1 cells. Alternative splicing events in siMFAP1 cells with P <0.01 and |Splicing Index| >2 were considered. This distribution was calculated from those alternative splicing events with annotations for the type of alternative splicing in Ensembl database (n=2491). **(C)** Analysis of the gene expression levels of probe sets with splicing indexes -2 > SI > 2 (P <0.01) in siMFAP1 cells (see Appendix 2).

GO analysis identified many biological processes enriched on the differentially spliced genes in siMFAP1 cells (Appendix 3). Three GO term clusters, with the highest

enrichment (enrichment score >8) stood out over the rest and were related to DNA repair, cell cycle regulation and chromatin organization and modification (Table R5).

Table R5. Top set of GO term enriched among differentially spliced genes in MFAP1-depleted cells.

Differentially spliced genes				
GO Biological Process	**Count (Total 2470)***	**Fold enrichment**	**P-value**	**Q-value (Benjamini)**
Related to DNA repair (Enrichment score: 10.96)				
GO:0006974 Cellular response to DNA damage stimulus	193	1.62	2.22E-12	1.49E-09
GO:0006259 DNA metabolic process	224	1.52	2.18E-11	7.29E-09
GO:0006281 DNA repair	136	1.75	2.69E-11	8.39E-09
Related to cell cycle (Enrichment score: 8.73)				
GO:0022402 Cell cycle process	307	1.54	6.31E-16	1.56E-12
GO:0007049 Cell cycle	361	1.46	3.61E-15	6.65E-12
GO:1903047 Mitotic cell cycle process	215	1.65	1.83E-14	2.45E-11
GO:0000278 Mitotic cell cycle	227	1.6	9.33E-14	8.73E-11
GO:0044770 Cell cycle phase transition	136	1.71	1.60E-10	4.54E-08
GO:0044772 Mitotic cell cycle phase transition	127	1.69	1.31E-09	3.14E-07
GO:0000086 G2/M transition of mitotic cell cycle	55	2.05	1.65E-07	2.57E-05
GO:0051726 Regulation of cell cycle	204	1.4	1.94E-07	2.93E-05
GO:0044839 cell cycle G2/M phase transition	58	2	1.94E-07	2.89E-05
GO:0010564 Regulation of cell cycle process	127	1.46	7.48E-06	6.20E-04
GO:1901987 Regulation of cell cycle phase transition	77	1.58	3.65E-05	0.002322995
GO:0007346 Regulation of mitotic cell cycle	105	1.47	3.98E-05	0.002468098
GO:1901990 Regulation of mitotic cell cycle phase transition	72	1.58	6.33E-05	0.003697803
Related to chromatin organization and modification (Enrichment score: 8.23)				
GO:0051276 Chromosome organization	257	1.47	2.42E-11	7.81E-09
GO:0018205 Peptidyl-lysine modification	108	1.8	3.86E-10	1.06E-07
GO:0016570 Histone modification	112	1.73	3.73E-09	8.31E-07
GO:0016569 Covalent chromatin modification	129	1.599872189	3.20E-08	6.11E-06
GO:0006325 Chromatin organization	158	1.398208713	6.68E-06	5.63E-04

*From an initial list of 3689 differentially spliced genes, GO analysis was performed from 2470 genes resulting from deleting those that had no GO term annotations or were redundant.

Furthermore, we identified a set of genes (119 genes) that had both altered gene expression level (P <0.05 and |linear fold change| >1.5) and splicing pattern (P <0.01 and |Splicing Index| >2). Many of these genes were related to cell cycle regulation and protein modifications (Appendix 4).

Taken together, these data suggest that MFAP1 depletion has a major impact on splicing than on gene expression. The fact that in our splicing analyses we have observed a GO term enrichment of genes related to DNA repair and cell cycle, could partially explain the genome instability phenotype of siMFAP1 cells.

We also carried out a gene expression analysis of HeLa cells depleted with siTHOC1 using the same Human Transcriptome Array (HTA) 2.0 from Affymetrix. 84 genes out of 67528 genes analyzed were de-regulated in siTHOC1 cells (P <0.05 and |linear fold change| >1.5). Among them, 52 genes (61.9%) were down-regulated (6 mRNA, 2 lncRNA, 1 miRNA, 25 ncRNA, 18 others) and 32 genes (38.1) were up-regulated (5 mRNA, 6 U snRNA, 2 miRNA, 6 snoRNA, 1 snRNA, 7 ncRNA, 5 others) (Appendix 5). GO analysis of the de-regulated genes revealed that no biological process was significantly enriched among these genes. Next we compared the list of de-regulated genes in siMFAP1 and siTHOC1cells. As shown in Figure R35, the overlap between down- and up-regulated genes upon MFAP1 or THOC1 depletion was significant in both cases (P < 0.0001). Importantly, gene expression analysis of siMFAP1 and siTHOC1 separately or in common showed that a large proportion of U snRNA but also snoRNAs were up-regulated (Appendix 6-7).

Gene expression changes

| Down-regulated | Up-regulated |
| siMFAP1 (359)   siTHOC1 (52) | siMFAP1 (545)   siTHOC1 (32) |

346   13   39

530   15   17
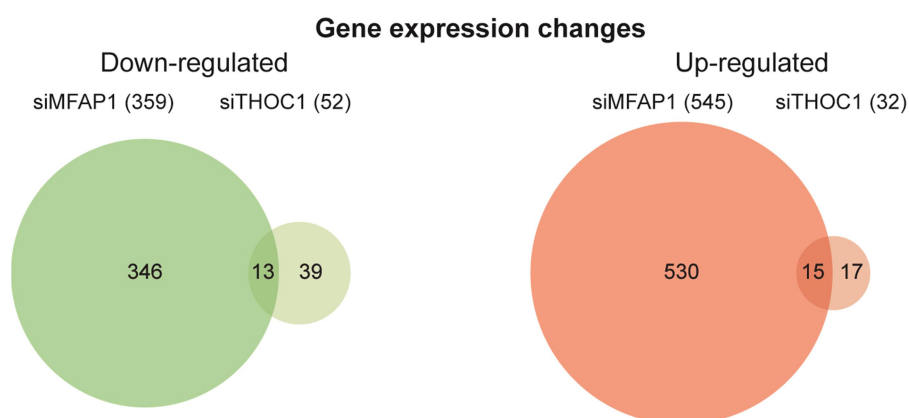
Figure R35. Common gene expression changes between MFAP1- and THOC1-depleted cells.
Venn diagrams showing the overlap between the lists of genes down-regulated (left panel) and up-regulated (right panel) in MFAP1- and THOC1-depleted cells, as provided by Pangloss (see Appendix 6-7). Expression levels (P <0.05; |linear fold change| >1.5). P-value < 0.0001 in all cases (Fisher´s exact test, two-tailed).

Given the physical interaction of THOC1 with MFAP1 and other splicing-associated factors (Rappsilber et al. 2002; Hegele et al. 2012), we wondered whether the depletion of THOC1 could lead to changes in splicing. Alternative splicing analysis of HTA 2.0 data revealed that 921 genes (1022 ASESs) were differentially spliced in THOC1-depleted cells (P <0.01 and |Splicing Index| >2) (Appendix 8). Analysis of AS type of genes with assigned annotations (n=372) showed that cassette exon was the most common type, and that AS event distribution was similar in siTHOC1 cells and in siMFAP1 cells (Figure R36A and Figure R34B). Interestingly, GO analysis of genes with AS changes in siTHOC1 cells showed an enrichment in the same biological processes (DNA repair and cell cycle) to those identified when MFAP1 is depleted (Table R6; Appendix 9 and Table R5). In fact, a 32.46% (299 out of 921) of genes with altered splicing patterns in siTHOC1 cells overlaps with the differentially spliced genes in MFAP1-depleted cells (Figure R36B and Appendix 10) Taken together, these observations could indicate that THO may affect alternative splicing through interactions with MFAP1 and maybe other splicing factors. Moreover, it is also possible that in part, splicing changes could occur in response to genome instability caused by THO depletion.

A Distribution of AS events in siTHOC1 cells

B Alternative splicing changes
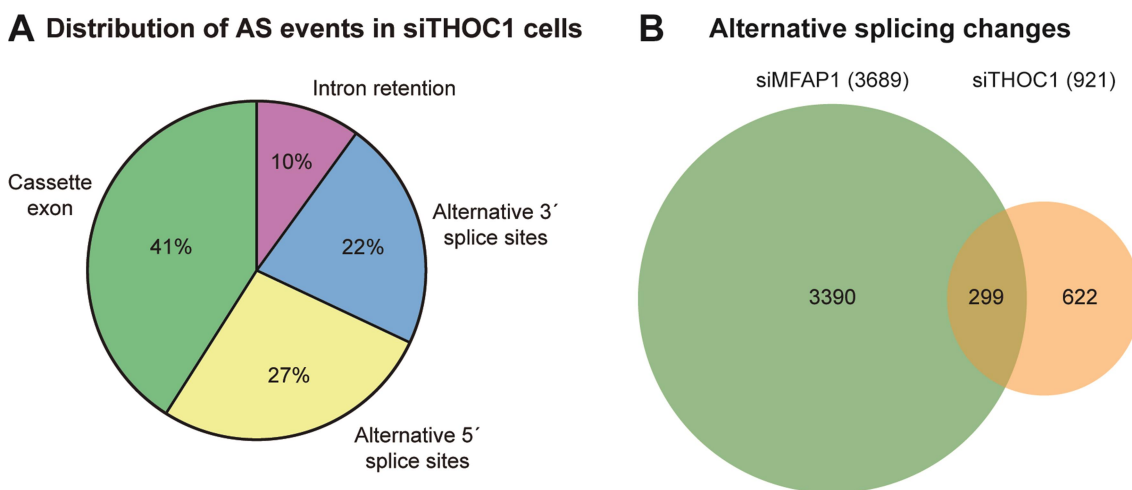
Figure R36. Alternative splicing changes in THOC1-depleted cells.
(A) Pie chart showing the alternative splicing type distribution in siTHOC1 cells. Alternative splicing events in siTHOC1 cells with P <0.01 and |Splicing Index| >2 were considered. This distribution was calculated from those alternative splicing events with annotations for the type of alternative splicing in Ensembl database (n=372). **(B)** Venn diagrams showing the overlap between differentially spliced genes in MFAP1- and THOC1-depleted cells (see Appendix 10). Venn diagrams were provided by Pangloss. Alternative splicing (P <0.01; |Splicing Index| >2). P-value < 0.0001 in all cases (Fisher´s exact test, two-tailed).

118

Table R6. Top set of GO term enriched among differentially spliced genes in THOC1-depleted cells.

Differentially spliced genes				
GO Biological Process	Count (Total 715)*	Fold enrichment	P-value	Q-value (Benjamini)
Related to organelle organization (Enrichment score: 4.24)				
GO:0007017 Microtubule-based process	51	1.94	9.82E-06	0.015055221
GO:0007010 Cytoskeleton organization	76	1.56	1.13E-04	0.052323667
GO:1902589 Single-organism organelle organization	99	1.44	1.71E-04	0.064090158
Related to DNA repair (Enrichment score: 3.24)				
GO:0006974 Cellular response to DNA damage stimulus	61	1.78	1.78E-05	0.018167641
GO:0006259 DNA metabolic process	65	1.53	5.82E-04	0.150897088
GO:0006281 DNA repair	34	1.51	0.018808449	0.519838373
Related to cell cycle (Enrichment score: 3)				
GO:0022402 Cell cycle process	94	1.635	2.62E-06	0.005379546
GO:0000278 Mitotic cell cycle	69	1.67	3.00E-05	0.026151011
GO:0007049 Cell cycle	105	1.47	4.23E-05	0.028662154
GO:1903047 Mitotic cell cycle process	62	1.63	1.60E-04	0.063751314
GO:0044770 Cell cycle phase transition	41	1.76	6.16E-04	0.146759537
GO:0044772 Mitotic cell cycle phase transition	39	1.78	6.81E-04	0.155063149
GO:0010564 Regulation of cell cycle process	42	1.66	0.001651155	0.230447136
GO:0007346 Regulation of mitotic cell cycle	34	1.63	0.006500088	0.388379223
GO:1901987 Regulation of cell cycle phase transition	25	1.75	0.009361909	0.425232602
GO:0000086 G2/M transition of mitotic cell cycle	16	2.05	0.011556337	0.464656094
GO:0051726 Regulation of cell cycle	57	1.37	0.013587172	0.494410542
GO:1901990 Regulation of mitotic cell cycle phase transition	23	1.73	0.014552589	0.504664704
GO:0044839 Cell cycle G2/M phase transition	16	1.91	0.021269934	0.54453913

*From an initial list of 921 differentially spliced genes, GO analysis was performed from 715 genes resulting from deleting those that had no GO term annotations or were redundant.

To get further inside in this comparative analysis we focused our analysis on those genes that showed AS changes in both MFAP1- and THOC1-depleted cells. GO and clustering analyses indicated that these genes participated in diverse biological processes such as cell morphogenesis, protein modification or protein location among others (Figure R37 and Appendix 11). Importantly, many of these genes were involved in DNA repair, RNA export and cell cycle and were listed in Figure R37. Interestingly, some genes involved in RNA transport and closely related to the THO complex (*THOC1/hHPR1, PCID2/hTHP1, TPR/hMLP1, DDX39B/hSUB2*) were differentially spliced in both siMFAP1 and siTHOC1 cells. All these data support a physical and functional interaction of THOC1 and MFAP1 at the interface of transcription and genome stability.

Cell morphogenesis
Regulation of catalytic activity
Protein modification (ubiquitination, proteasome...)
Microtubule-movement
DNA REPAIR
Protein location
RNA EXPORT
Development and morphogenesis
Heat response
CELL CYCLE

MMS19, PHLPP1, BACH1, TMX3, DNAJC10, MCM10, HIST2H4B, ZBTB38, ATF2, VPS13C, HMOX1, ATG7, MAPKAP1, TLK2, TPR, PMS1, EIF2B4, DDX39B, UBE4B, POLE, ATR, CENPJ, PPARGC1B, ITPR1, DEPDC5, TMEM67, SPAG9, ATXN3, CDKN1A, RIF1, BRE, UGGT2, RBM14, SETD2, MAP3K13, MELK, USP45, THOC1, DBF4B, NAP1L1, CTNNB1, PRIM2, TNKS, RBMS1, NOL8, NFIB

HNRNPA1L2, PCID2, FYTTD1, DDX39B, SRSF11, TNKS, TPR, SETD2, THOC1, CAV2, TSG101, TMEM67, CDKN1A, RBM14

BACH1, CAV2, DBF4B, TSG101, MCM10, SYCP2, AZI2, ATF2, CTNNB1, APP, ANK3, PRIM2, TNKS, TLK2, TPR, TAF1, LAMTOR4, CEP192, PCID2, DDX39B, POLE, SKP2, NEK10, ATR, CENPJ, SEPT10, PSMB8, PSMA2, TMEM67, CDKN1A, RAB11FIP3, RIF1, BRE, DIS3L2, RBM14, MELK, THOC1, ACVR1

Figure R37. GO term cluster enriched among common genes differentially spliced in MFAP1- and THOC1-depleted cells.
Top ten GO term clusters enriched among common genes with AS changes in MFAP1- and THOC1-depleted cells (left panel). Genes within DNA repair, RNA export and cell cycle clusters are shown (right panel). See Appendix 11.

4. ANALYSIS OF GENOME INSTABILITY IN YEAST MUTANT OF Sin3A COMPLEX AND *MFAP1* HOMOLOG GENES

We have identified two novel partners of human THO complex, the Sin3A complex and MFAP1 protein, whose depletion leads to genome instability. Given that the Sin3 complex is a conserved eukaryotic complex, we performed a functional analysis in the yeast model organism in order to see whether genome maintenance functions are conserved and yeast *sin3* mutants show a similar phenotype as those observed in siSAP130 and siSIN3 human cells. In the case of MFAP1 we carried out the same approach with the putative orthologue gene.

4.1. Genome instability in yeast Sin3-Rpd3 mutants

Sin3 is an evolutionarily conserved corepressor complex involved in transcription regulation. Most of its subunits are well conserved from yeast to human, as is the case of the scaffold SIN3 protein or the histone deacetylase HDAC/Rpd3. Nevertheless, SAP130 seems to have no homolog in yeast. Two different Sin3-Rpd3 complexes have been identified in yeast, large (Rpd3L) that has been implicated in deacetylation predominantly at the promoter region, and small (Rpd3S) that appears to be involved in deacetylation predominantly at coding regions (Grzenda et al. 2009). Both complexes contain the core components (Rpd3, Sin3 and Ume1) but different additional subunits as shown in Figure R38.
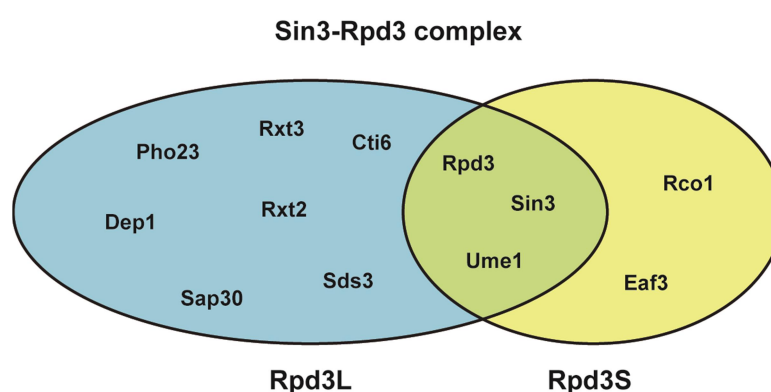
Figure R38. Yeast Sin3-Rpd3 complex.
Schematic representation of the two different yeast Sin3-Rpd3 complexes (Rpd3L and Rpd3S). Sin3, Rpd3 and Ume1 are core components (green), shared by the two complexes (Kadamb et al. 2013). Blue corresponds to the specific subunits in Rpd3L and yellow corresponds to the specific subunits in Rpd3S.

In addition to its role in transcription repression, Sin3-Rpd3 mutants have been related with genome instability (Silverstein and Ekwall 2005), but a direct connection between transcription and genome instability was not established when we started this work.

Taking into account the physical and functional relationship between human Sin3A and THO complexes we decided to get insight into the role of yeast Sin3-Rpd3 in the maintenance of genome stability. We carried out the analysis with *sin3* and *rpd3* mutants and mutants of representative subunits of Rpd3L and Rpd3S complexes (Figure R38). First, we analyzed the sensitivity of the Sin3-Rpd3 mutants to different genotoxic agents or irradiation by serial dilutions in media containing: 4-nitroquinoline N-oxide (4-NQO), a UV-mimetic that generates guanine and adenine adducts and oxidative damage; Ultraviolet light (UV) irradiation, which leads to bulky adducts; Methyl Methanesulfonate (MMS), an alkylating agent; Hydroxyurea (HU), which inhibits replication by depleting the amount of deoxyribonucleotide triphosphate (dNTPs); Camptothecin (CPT), a cytotoxic quinoline alkaloid which inhibits the DNA enzyme topoisomerase I (topo I). Among the mutants analyzed, only two, showed sensitivity to genotoxic agent, at least at the conditions assayed: *sin3Δ* displayed sensitivity to UV and HU, and *eaf3Δ* displayed sensitivity to 4-NQO, UV and HU (Figure R39). These results are in agreement with a key role of Sin3 as scaffold of the Sin3-Rpd3 complex. The stronger sensitivity of *eaf3* mutant can be explained as Eaf3 being a part of two different corepressor complexes, Sin3-Rpd3 and NuA4 (Reid et al. 2004).

Figure R39. Analysis of sensitivity of Sin3-Rpd3 mutants to genotoxic agents.
Sensitivity to 4-NQO, UV, MMS, HU and CPT of WT (BY4741), *sin3Δ* (YOL004W), *sap30Δ* (YMR263W), *rco1Δ* (YMR075W), *ume1Δ* (YPL139C), *rpd3Δ* (YNL330C), *sds3Δ* (YIL084C), *pho23Δ* (YNL097C), *eaf3Δ* (YPR023C). 10-fold serial dilutions of exponentially growing cultures are shown. Experiment performed in collaboration with A. Panek.

Next, we analyzed spontaneous recombination as an indirect measure of DNA breaks. For this, we used different plasmid-born recombination systems (L and LYΔNS) based on direct truncated repeats of 0.6-kb of the *LEU2* gene that are transcribed from the *LEU2* promoter. These systems differ in the length of the transcribed intervening sequence (31 bp for L, and 3.7 kb for LYΔNS) (see Figure M4) which allow the analysis of transcription-associated recombination. Recombination frequencies were measured as the frequency of Leu$^+$ colonies. The analysis with L and LYΔNS recombination systems revealed an increase in recombination frequency when a long sequence is transcribed (LYΔNS) in WT and the different mutants of Sin3-Rpd3 (Figure R40A). We found the stronger hyper-recombination phenotype in *rpd3Δ* and *rco1Δ* mutants, being the increase in recombination lower in *sin3Δ*, *ume1Δ*, *sap30Δ* and *pho23Δ* as compared with that of WT cells (Figure R40A). To confirm whether mutation in Sin3-Rpd3 complex lead to a transcription-dependent recombination, we determined the effect of three representative mutants (*sin3Δ*, *sap301Δ* and *rco1Δ*) in the L-*lacZ* and GL-*lacZ* recombination systems in which the G+C-rich *lacZ* gene is located between the *leu2* repeats and transcription is regulated by the constitutive LEU2 promoter (P_{LEU2}) (L-*lacZ* system) or by the inducible GAL1,10 promoter (P_{GAL}) (GL-*lacZ* system) (Figure R40B). Transformants were grown under conditions of low (P_{GAL} promoter in 2% glucose), medium (P_{LEU} promoter in 2% glucose) and high levels of transcription (P_{GAL} promoter in 2% galactose) of the recombination system used in each case. As can be seen in Figure R40B, recombination frequencies are higher in the mutants than in WT cells in low transcription conditions, moreover, comparison between low and medium transcription levels indicate that the higher the strength of transcription of the recombination system used, the stronger the increase in recombination. At high transcription conditions recombination frequencies were lower than expected, possibly due to poor growth of transformants on galactose media (Figure R40B). The data suggest that Sin3-Rpd3 mutants cause more recombination events and consequently more DNA breaks. We tested this prediction by quantifying the number of Rad52 foci, which globally marks recombination repair centers (Lisby et al. 2001). As can be seen in Figure R40C almost all the mutants lead to a significant increase in the percentage of cells with Rad52 foci. Importantly, Rad52-foci accumulation was fully suppressed by RNase H1 (RNH1) overexpression (Figure R40C), indicating that part of the genetic instability of Sin3-deficient cells is due to R loops, which is consistent with our data in human cells.
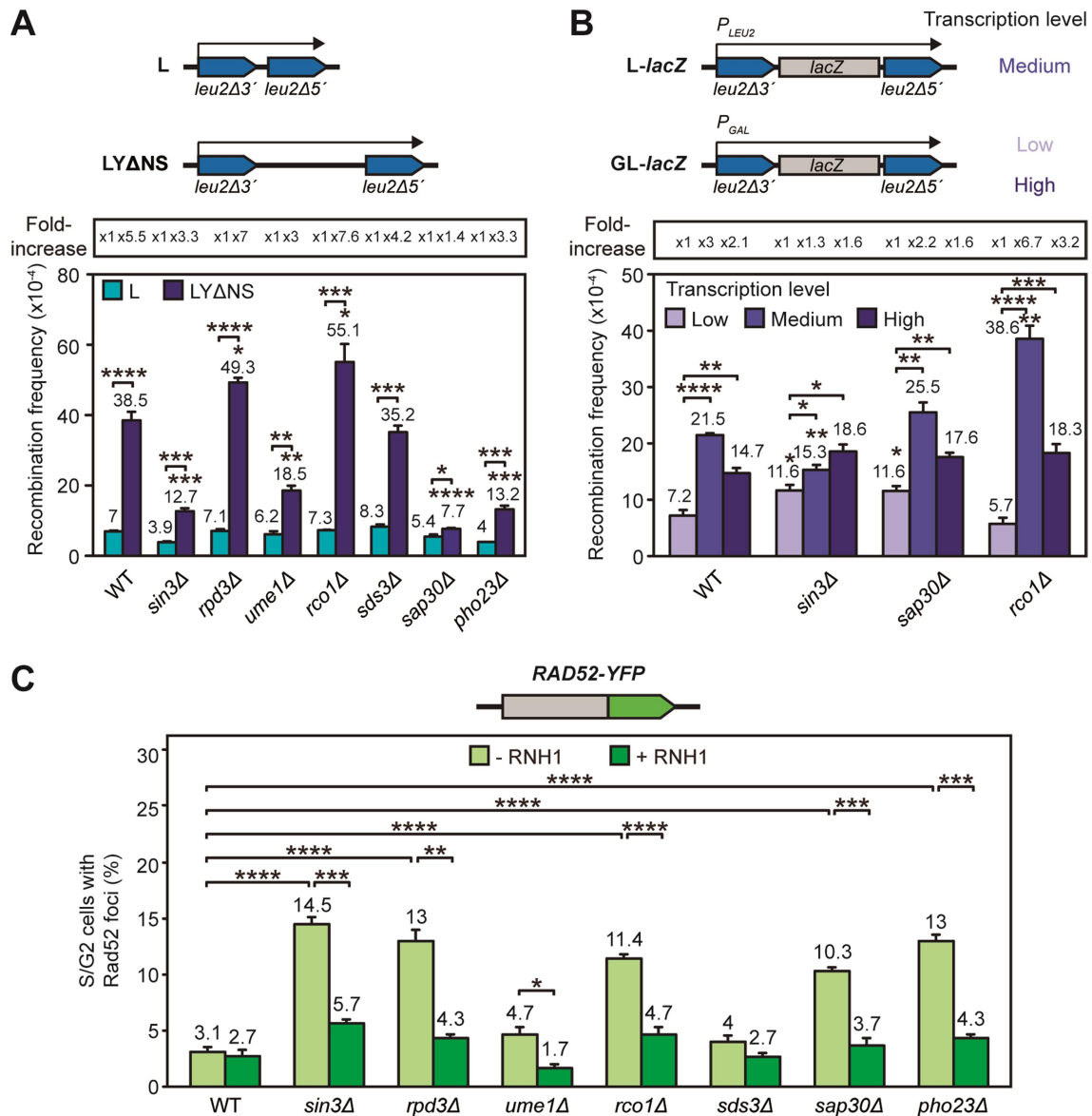
Figure R40. Transcription and R loop dependency of genome instability in yeast Sin3-Rpd3.
(A) Recombination analysis of WT (BY4741), *sin3Δ* (YOL004W), *rpd3Δ* (YNL330C), *ume1Δ* (YPL139C), *rco1Δ* (YMR075W), *sds3Δ* (YIL084C), *sap30Δ* (YMR263W) and *pho23Δ* (YNL097C) cells carrying L and LYΔNS plasmid system. **(B)** Recombination analysis of WT (BY4741), *sin3Δ* (YOL004W), *sap30Δ* (YMR263W) and *rco1Δ* (YMR075W) cells carrying the L-lacZ system, expressed under the control of *LEU* promoter (medium) or GL-lacZ system in glucose or galactose, expressed under the control of *GAL* promoter (low and high transcription levels, respectively). Means and SEM of three independent experiments are depicted (A, B). **(C)** Percentage of spontaneous Rad52-YFP foci formation in the strains detailed in (A) with or without RNase H1 (RNH1) overexpression from the pCM189-RNH1 plasmid. Means and SEM of three independent experiments are depicted. A scheme of recombination system (A, B) or Rad52-YFP plasmid (C) is shown on top of each panel. ****, P < 0.0001; ***, P < 0.001, **, P < 0.01, *, P < 0.05 (Student´s *t*-test) (A, B, C). Experiments performed in collaboration with A. Panek.

Next, we analyzed the effect of the absence of Sin3 in cells lacking the THO-subunit Hpr1, using the chromosomal *leu2-k::ADE2-URA3::leu2-k* system to assay recombination in *sin3Δ*, *rpd3Δ* and *hpr1Δ* single mutants and in *sin3Δhpr1Δ* and *rpd3Δhpr1Δ* double mutants (Figure R41A, B). This system is located in chromosome III and contains two 2.16-kb long directs repeats of the *leu2-k* allele separated by the *ADE2* and *URA3* genes. Deletions caused by recombination between two repeats give rise to the loss of the *URA3* marker, which can be scored in 5-FOA-containing medium (see Materials and Methods 16.2). Recombination levels in *hpr1Δ* were significantly higher than in wild-type cells (3084-fold), as previously described (Aguilera and Klein 1990; Huertas et al. 2006), and a 4- and 6.8-fold increase in recombination was observed in *sin3Δ* and *rpd3Δ* respectively (Figure R41A, B). Interestingly, the absence of Sin3 in *hpr1Δ* (*sin3Δhpr1Δ*) leads to a 8-fold reduction in *hpr1Δ* recombination levels. However, no reduction was observed in *rpd3Δhpr1Δ*, suggesting that the lack of Rpd3 could have a pleiotropic effect. In addition to the recombination analysis, the *sin3Δhpr1Δ* double mutant also showed a reduction in Rad52-foci accumulation compared with single mutants (Figure R41C). This partial suppression was not due to a reduction in transcription, since no significant differences in mRNA levels between single and double mutants were observed, as determined by RT-qPCR (Figure R41D). Altogether, these results are in agreement with data obtained in human cells and support a model in which yeast Sin3-Rpd3 and THO complexes work together in the maintenance of genome integrity.
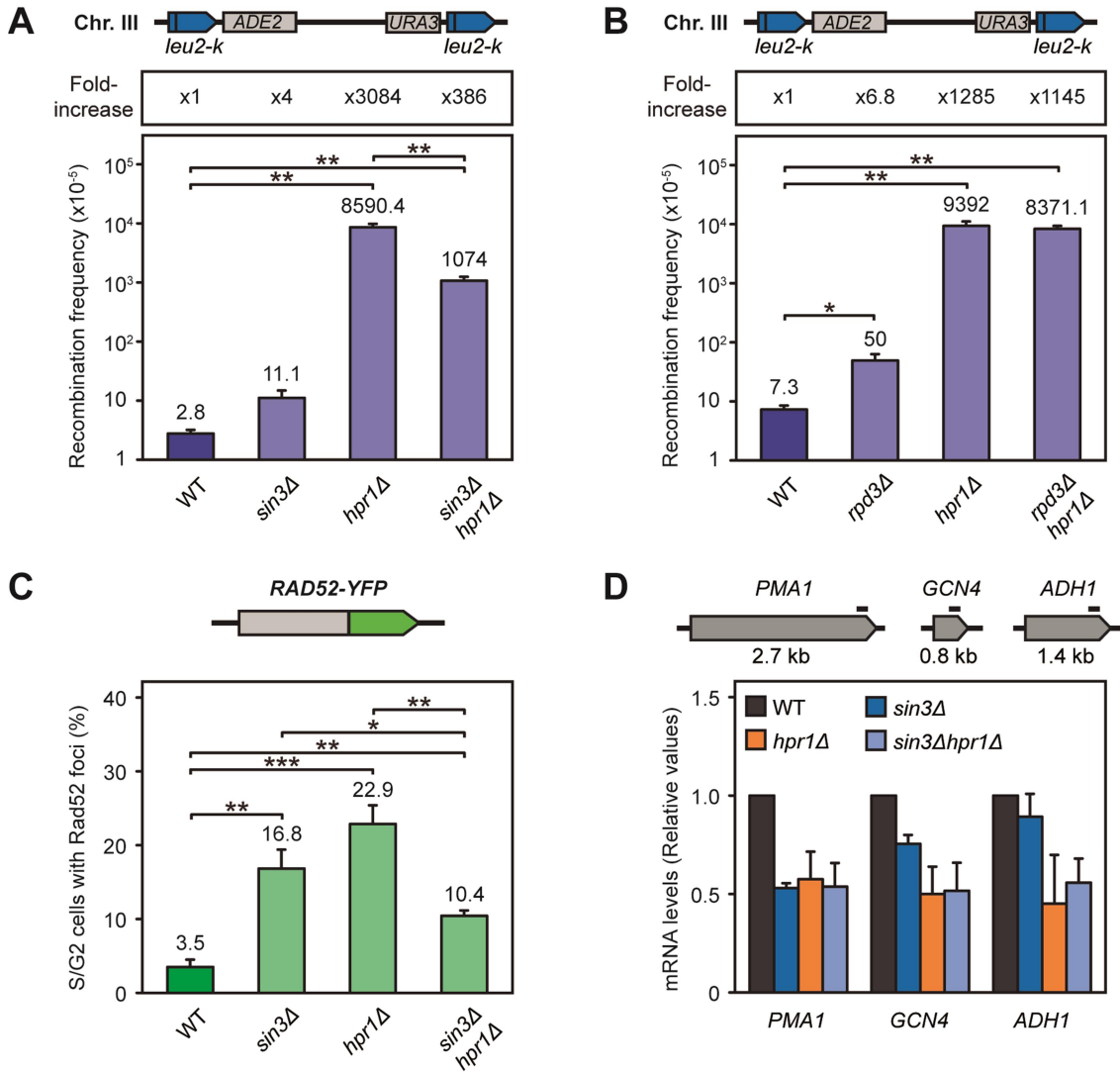
Figure R41. Suppression of the *hpr1Δ* genome instability phenotype by *sin3Δ*.
(A) Recombination analysis of WT (LKWT), *sin3Δ* (LKSIN3), *hpr1Δ* (LKH1) and *hpr1Δ sin3Δ* (LKH1SIN3) strains carrying the chromosomal direct-repeat system *leu2-k::ADE2-URA3::leu2-k*. **(B)** Recombination analysis of WT (LKWT), *rpd3Δ* (LKRPD3), *hpr1Δ* (LKH1) and *hpr1Δ rpd3Δ* (LKH1RPD3) strains in the same chromosomal recombination system. **(C)** Percentage of spontaneous Rad52-YFP foci formation in the strains detailed in (A). Means and SEM of three independent experiments are depicted (A, B, C). **(D)** Relative mRNA quantification of *PMA1*, *GCN4* and *ADH1* genes by RT-qPCR in the strains analyzed in (A). mRNA expression values were normalized to the expression of the *SCR1* gene. Means and SEM of two independent experiments are shown. A scheme of recombination system (A, B), Rad52-YFP plasmid (C) and regions of genes where RT-qPCRs were performed (D) is shown on top of each panel. ****, $P < 0.0001$; ***, $P < 0.001$, **, $P < 0.01$, *, $P < 0.05$ (Student´s *t*-test) (A, B, C). Experiments performed in collaboration with JM. Santos-Pereira.

4.2. ***SPP381* the putative ortholog of *MFAP1* and its role in the maintenance of genome stability**

Human *MFAP1* is highly conserved in *Caenorhabditis elegans* and *Drosophila melanogaster* (Andersen and Tapon 2008; Ma et al. 2012). Moreover, it has been also reported a homologue gene in *Schizosaccharomyces pombe* called Saf3 (Splicing associated factor 3) (Ren et al. 2011), but no ortholog has been identified in *Saccharomyces cerevisiae*. To determine the putative *MFAP1* homolog in budding yeast we compared the amino acid sequence of MFAP1 protein (Human) and Saf3 protein (*S. pombe*) with the *S. cerevisiae* proteome by BLAST analysis (Basic Local Alignment Search Tool). As a result, we found one candidate, Spp381 protein (*S. cerevisiae*) from BLAST analysis using the amino acid sequence of Saf3 as query sequence. The percentages of identity and similarity were only about 12-13% and 22-23% respectively, as can be observed in Table R7 and in the multiple sequence alignment between these proteins (Figure R42). Despite the low sequence homology the fact that Spp381, Saf3 and MFAP1 proteins are pre-mRNA splicing factors involved in the same stages of the splicing process and are linked to the NineTeen complex (NTC), through their interaction with Prp38, lead us to contemplate the possibility of *SPP381* as the putative ortholog of *SAF3* and *MFAP1* genes (Lybarger et al. 1999; Ren et al. 2011; Hegele et al. 2012).

Table R7. Percentage of identity and similarity between Spp381, Saf3 and MFAP1 protein sequences

	Saf3 (*S. pombe*)		MFAP1 (*H. sapiens*)	
Spp381 (*S. cerevisiae*)	12.87*	23.51**	13.33*	22.37**
Saf3 (*S. pombe*)			25.17*	40.40**

* % of identity
** % of similarity

```
Spp381(S.cerevisiae) ------------------------------------------------------------- 0
Saf3    (S.pombe)    MPIDTPLIPH--------------------KPISRYRSGRLEGLESESSSE---SDYEEVSDS------------- 40
MFAP1 (H. sapiens)   MSVPSALMKQPPIQSTAGAVPVRNEKGEISMEKVKVKRYVSGKRPDYAPMESSDEEDEEFQFIKKAKEQEAEPEEQEEDS 80

Spp381(S.cerevisiae) ---MSFRFFKRRIDTSSADESSSA-DEEHP---DQNVSLTEKSA--------SLSHSDIGG-EILNGTGK--NRTPNDG 61
Saf3    (S.pombe)    --------------HDQENSISSSRHVITPVFEKEGISETKTSS--------------NFQNINP-----VQTIDNS 84
MFAP1 (H. sapiens)   SSDPRLRRFLQNRISEDVEERLARHRKIVEPEVVGESDSEVEGDAWRMEREDSSEEEEEEIIDEEIERRRGMMRQRAQERK 60

Spp381(S.cerevisiae) QESNESDGS-PESDESPESEESSDNSDS-SDSDDMRPLPRPLEMKKKANNLQKATKIDQPWNTQDDARVLQTKK-----E 134
Saf3    (S.pombe)    ASFEYETDASSAEGGSN-SAASSSEEDSSDSEYEMELRRRILLLPPRFTSKVIKNRA---KANEEDTEVLKK-------- 152
MFAP1 (H. sapiens)   NEEMEVMEVEDEGRSGEGESESESEYEEYTDSEDEMEPRLKPVFIRKKDRV---------TVQEREAEALKQKELEQEAK 230

Spp381(S.cerevisiae) NMIKNIDKANQVAKNYETMKLRILN-TNYSTNEELIKQCMLLDDNDEVDSEKERQKWFERQNERKQKHRRIQLAKQRESEE 213
Saf3    (S.pombe)    -------VTSQKI-LEETIKREILLKETKNNNELL-N--DIDDTDGIDPQSEYELWKLRHILRKKRDKEKSLELEREKMA 221
MFAP1 (H. sapiens)   RMAEERRKYILKI-VEEETKKEI-----EENKRSL-AALDALNTDLENDEEYEAWKVRELKRIKRDREDREALEKEKAE 303

Spp381(S.cerevisiae) YEAKRFEAM---------QKGKD-----GNTKYLVIL---------LKEKERLDHKKQRSAEKVEKSHNNN-RYKITRTKN 270
Saf3    (S.pombe)    IEERRLMNSEEREACDLKDAEASRRGKKKSSMQFLQKYYHKGAFYQNED-IVV-SKRDYSEATEGEVLNKDLLPKPMQIRG 299
MFAP1 (H. sapiens)   IERMRNLTEEERRAELRANGKVITNKAVKGKYKFLQKYYHRGAFFMDEDEBV-YKRDFSAPTLEDHFNKTILPKVMQVR- 381

Spp381(S.cerevisiae) VEFGDLGKNSRDYEETEYSVI-------------------------------------------------------- 291
Saf3    (S.pombe)    DLFAKAGQT-------RWTHLANEDTTKEGSAWYDPKNPILQKNLHRLGGLHS--ISPLSKRKRT- 355
MFAP1 (H. sapiens)   -NFGRSGRT-------KYTHLVDQDTTSFDSAWGQESAQNTKFFKQKAAGVRDVFERPSAKKRKTT 439
```

Figure R42. Alignment of *Saccharomyces cerevisiae* Spp381, *Schizosaccharomyces pombe* Saf3 and *Homo sapiens* MFAP1 proteins.
Multiple sequence alignment of Spp381 (*S.cerevisiae*), Saf3 (*S. pombe*) and MFAP1 (*H. sapiens*) proteins generated by Clustal Omega. The dashes within the aligned sequences, indicate the gap that were inserted in order to optimize the alignment. Gray background indicates similarity, and black background indicates identity.

In order to confirm whether *SPP381* was the functional homologue of MFAP1 we carried out an analysis of yeast mutants. First, we analyzed whether the *spp381-ts* thermosensitive mutant was sensitive to genotoxic agents leading to replication impairment and/or DSBs by serial dilutions in media containing 4-NQO, UV irradiation, MMS, HU or CPT. We found that *spp381-ts* cells were more sensitive to 4-NQO, UV and HU than the wild type, suggesting an accumulation of DNA damage in Spp381-deficient cells (Figure R43). Next, we determined whether this phenotype also results in an increase of spontaneous recombination. Indeed, recombination levels using the L and LYΔNS or L-*lacZ* and GL-*lacZ* recombination systems were increased in *spp381-ts* with respect to WT levels, except when transcription rate was high (P_{GAL} in galactose), due to growth problems in galactose media (Figure R44A, B). In accordance with the hyperrecombinant phenotype, high levels of recombinogenic breaks were inferred by determining the frequency of Rad52-YFP foci in *spp381-ts* and *spp381Δ* mutants (Figure R44C, D). Importantly, such increase in Rad52-YFP foci percentage was not suppressed after RNH1 overexpression in both mutants, indicating that R-loops are not responsible for the genetic instability of Spp381-deficient cells (Figure R44C, D). These data are in agreement with previous studies showing that *spp381* mutant and other splicing factors lead to CIN phenotype while RNA-DNA hybrid levels appear to be not affected (Stirling et al. 2011; Chan et al. 2014). In addition, although complement analysis in yeast would be required to confirm the homology between

MFAP1 and Spp381 proteins, the similar genome-instability phenotype of *spp381* mutant and MFAP1-depleted cells, indicates a conserved function during evolution.
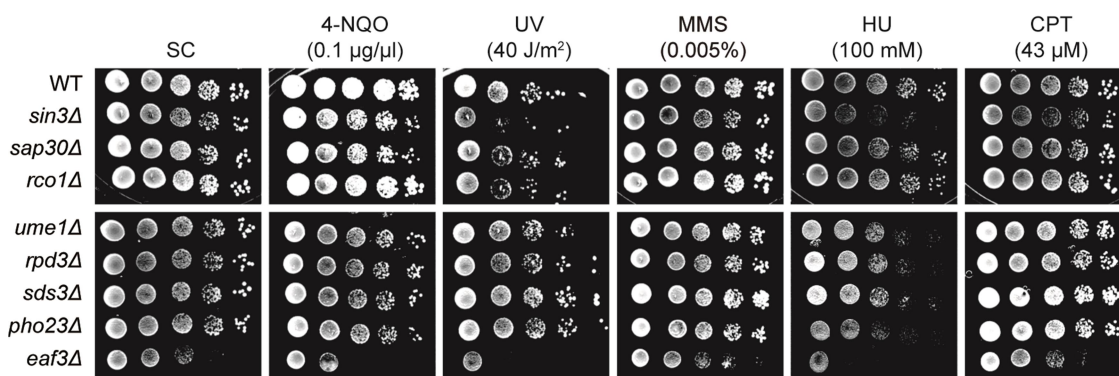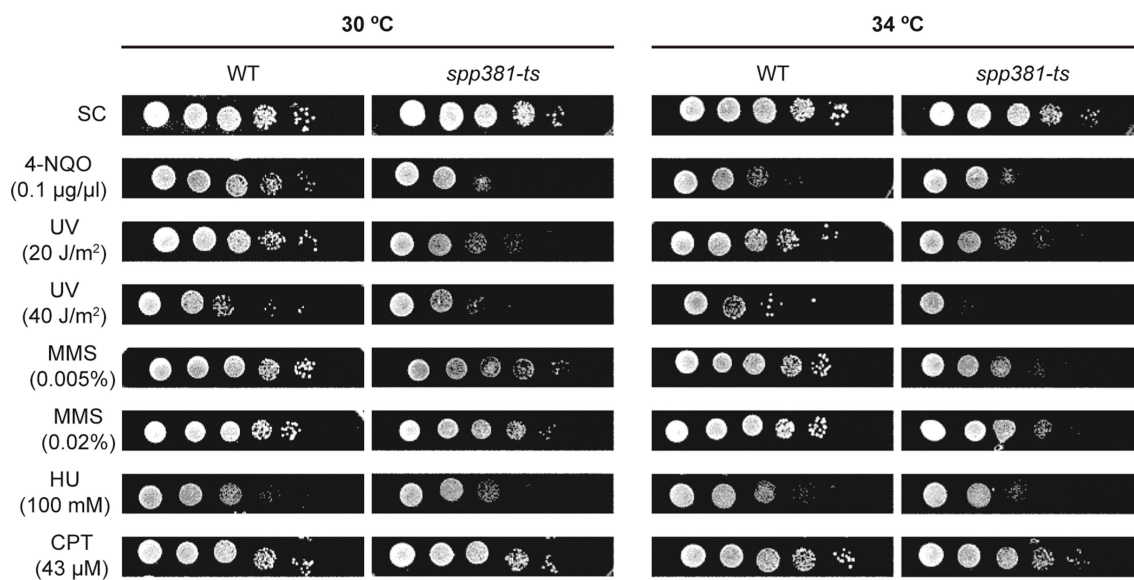

Figure R43. Analysis of sensitivity of *spp381-ts* mutant to genotoxic agents.
Sensitivity to 4-NQO, UV, MMS, HU and CPT of WT (BY4741), *spp381-ts* (YPH2622). 10-fold serial dilutions of exponentially growing cultures are shown. Experiment performed in collaboration with A. Panek.
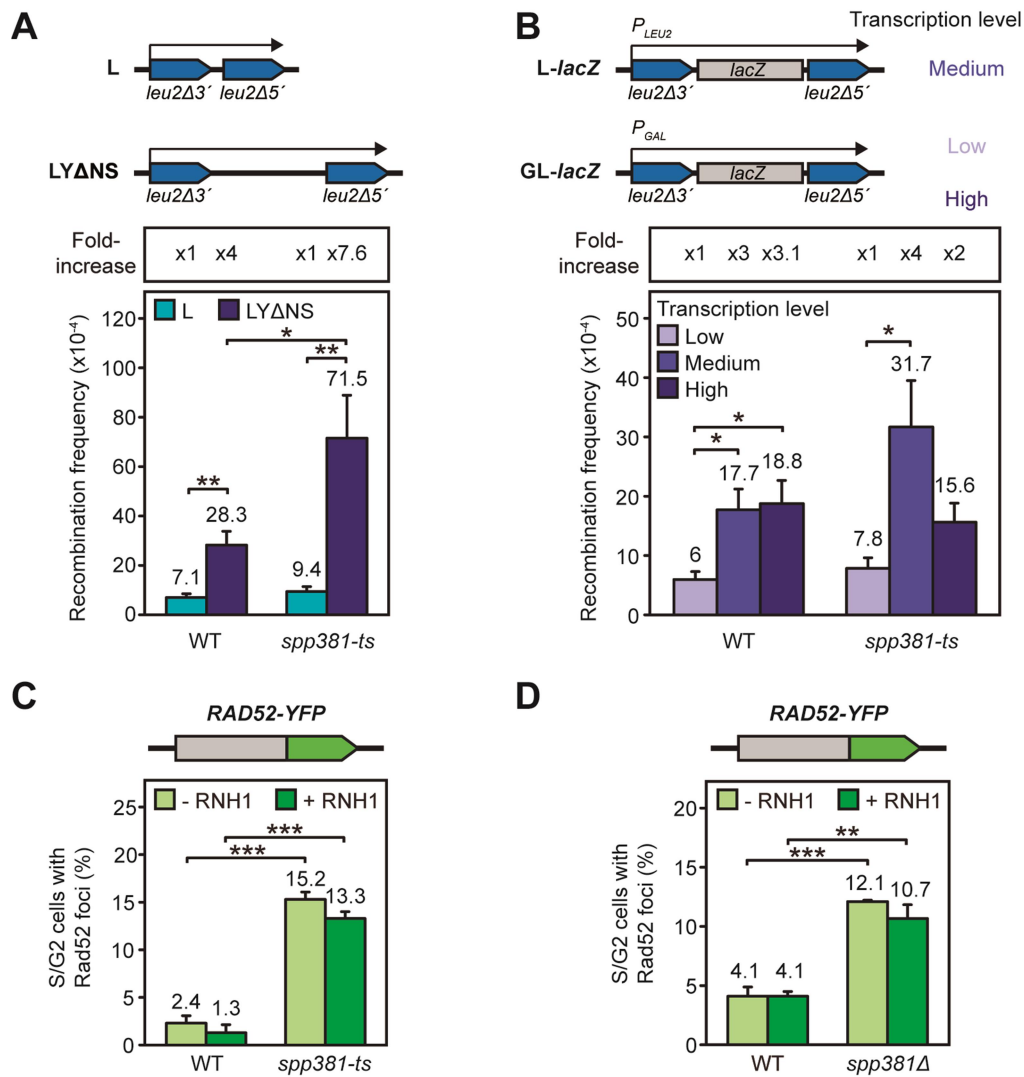
Figure R44. Genome instability and R-loop independency by lack of Spp381 in *S. cerevisiae*.
(**A**) Recombination analysis of WT (YDR007W), and *spp381-ts* (YPH2622-2D-4B) cells carrying L and LYΔNS plasmid system. (**B**) Recombination analysis of WT (YDR007W), and *spp381-ts* (YPH2622-2D-4B) cells carrying the L-*lacZ* system, expressed under the control of *LEU* promoter (medium) or GL-*lacZ* system in glucose or galactose, expressed under the control of *GAL* promoter (low and high transcription levels, respectively). Means and SEM of three independent experiments are depicted (A, B). (**C**) Percentage of spontaneous Rad52-YFP foci formation in the strains detailed in (A) with or without RNase H1 (RNH1) overexpression from the pCM184-RNH1 plasmid. (**D**) Percentage of spontaneous Rad52-YFP foci formation in WT (CEN.PK) and *spp381Δ* (CEN.RO20-3A) strains with or without RNase H1 (RNH1) overexpression from the pCM189-RNH1 plasmid. Means and SEM of three independent experiments are depicted (A, B, C, D). A scheme of recombination system (A, B) or Rad52-YFP plasmid (C, D) is shown on top of each panel. ***, P < 0.001, **, P < 0.01, *, P < 0.05 (Student´s *t*-test) (A, B, C, D). Experiments performed in collaboration with A. Panek (A, B, C).

DISCUSSION

134

In this thesis we have identified new partners of human THO/TREX and we have analyzed the impact on genome stability of two of them. We show that THOC1, a subunit of the THO/TREX complex, interacts with the Sin3A histone deacetylase complex, and with MFAP1, a spliceosome-associated factor. Depletion of the Sin3A complex leads to R-loop-mediated genome instability, similar to that conferred by THO mutants. Importantly, we provide evidence that inhibition of histone deacetylation results in R-loop accumulation. In addition we show that histone acetylation inhibition suppresses DNA breaks and R loops in THOC1-depleted cells, suggesting that both THO and Sin3A complexes work together to prevent R loops. By contrast, depletion of MFAP1 leads to DNA breaks and genome instability that are not dependent on transcription neither on RNA-DNA hybrids. Gene expression and splicing analyses of MFAP1-depleted cells point that the impact of this protein on genome integrity is mainly due to its role in splicing.

1.   NEW PARTNERS OF HUMAN THO/TREX

THO/TREX is an RNA-binding factor with an important role in mRNP assembly. It connects transcription with mRNA export and is involved in the maintenance of genome integrity (Chavez et al. 2000; Rondon et al. 2010). This complex is an example of the relevance of mRNP biogenesis in genetic stability (Aguilera 2005). The current view is that THO/TREX, as well other RNA binding and processing factors, are necessary for the formation of a correct mRNP preventing the formation of RNA-DNA hybrids that are partially responsible for the transcription impairment and the genetic instability phenotypes observed in these mutants (Aguilera and Garcia-Muse 2013). Nevertheless, whether this is sufficient to explain how specific RNA binding factors control co-transcriptional accumulation of harmful R loops is unclear.

During mRNP biogenesis, THO is recruited to active chromatin and interacts with a variety of proteins that either contribute to THO binding to the pre-mRNA or use THO as a landing platform subsequently binding to nascent pre-mRNA (Luna et al. 2012). THO/TREX has been shown to interact with a variety of factors involved in the different stages of gene expression that includes interactions with transcription machinery, mRNA processing factors (capping, splicing and 3´ end processing factors) and mRNA export factors among others (Zhou et al. 2002; Chen et al. 2007; Luna et al.

2012; Meinel et al. 2013). Identification of new interactors of human THO/TREX could help to better understand how mRNP biogenesis contributes to the maintenance of genome stability.

By using the yeast two-hybrid system we identified several putative interactors of human THO/TREX (Figure R2 and Table R2). Among them, we have found different partners that have been shown to interact with components of THO/TREX. Thus, using UAP56 as a bait we found that this helicase interacts with its paralog URH49, in agreement with a previous high-throughput two-hybrid screening in which URH49 was detected as one frequent interactor of UAP56 (Lehner et al. 2004). Given the high homology shared by UAP56 and URH49 (90% amino acid sequence identity) it is possible that these helicases could form heterodimers and homodimers, as suggested by our results and previous analysis (Lehner et al. 2004). In this two-hybrid high-throughput analysis UAP56 was also found to interact with itself and with CIP29, an hnRNP homologous to yeast protein Tho1 identified as suppressor of THO mutants (Jimeno et al. 2006). We have not found neither of these interactors using UAP56 as a bait, in contrast, our screening with URH49 reveals that only this factor interacts with CIP29 (Table R2). Our data is consistent with previous biochemical studies and functional analysis that propose that URH49 and UAP56 have different partners and different impact on cell division (Dufu et al. 2010; Yamazaki et al. 2010), the former interacting with CIP29 to form the AREX complex (alternative mRNA export complex), the latter with Yra1/Aly to form TREX (revised in (Luna et al. 2012)). Thus, our results are in agreement with previous data showing the functional specificity of these helicases, and importantly, can also serve us as validation of our screening and procedures.

A new putative protein partner, ZBTB44 was found to interact with ALY (Figure R2 and Table R2). Although ZBTB44 is an uncharacterized protein, proteins that contain ZBTB domains (Zinc finger and Broad complex, Tramtrack, and Bric-à-brac) have been shown to bind to regulatory regions in target genes and are usually related to transcription repression by recruiting corepressors and histone modification enzymes (Beaulieu and Sant'Angelo 2011; Lee and Maeda 2012). In this sense, it is worth noting that ALY interacts with the transcription factor E2F2 and modulates the expression of a large number of E2F-responsible genes (Osinalde et al. 2013). This suggests that ALY in addition to its role in mRNA export (Katahira 2012)

136

could have a role in transcription regulation through its interaction with different transcription factors. It would be important, first to validate this new interaction between ALY and ZBTB44, and in that case, further investigate this relationship and whether or not THO/TREX as a complex is required for this crosstalk.

Nineteen proteins were identified as new putative interactors of THOC1 (Figure R2 and Table R2). Although THO is formed by a core of different interacting subunits that associates with ALY and UAP56 forming TREX (Strasser et al. 2002; Masuda et al. 2005), we have not identified any of THO subunits among these THOC1-interactors. One possible explanation for this is the fact that, in some cases, *in vivo* interactions between subunits of a complex are stable only when all the components are present. In fact, THO behaves as a structural and functional unit (Chavez et al. 2000; Garcia-Rubio et al. 2008) and lack of any subunit of yeast THO affects the stability of the complex (Huertas et al. 2006). New THOC1-interacting proteins identified are involved in many different processes (Table R2). In this thesis we have focused our studies on two factors, SAP130, a subunit of the histone deacetylase complex Sin3A (Fleischer et al. 2003) and the splicing factor MFAP1 (Makarov et al. 2002; Hegele et al. 2012). Before to discuss the functional relevance of these interactions, we will describe briefly some other putative nuclear THOC1-interactors that could be interesting to study, despite they have not yet been validated, at least at the condition assayed.

In our yeast two-hybrid screening using THOC1 as a bait we have found different proteins involved in protein degradation mediated by ubiquitination such as CUL1, NUB1 and ZSWIM2 (see Table R2 for more details and references). These putative interactions are consistent with THOC1 being a substrate of ubiquitination machinery (Gwizdek et al. 2005; Song et al. 2013). In addition other putative interactors such CDC23, SPATA22 and QRICH1, play a role in cell cycle regulation, cell differentiation and/or apoptosis (see Table R2 for more details and references), which it is compatible with the described role of THOC1 in these processes. Thus, THOC1, was initially described as a human nuclear matrix protein that binds to the retinoblastoma tumor suppressor protein pRb (Durfee et al. 1994), and different reports connect this factor with cell proliferation, apoptosis and cancer (Wang et al. 2009a; Dominguez-Sanchez et al. 2011b; Heath et al. 2016). Finally, regarding DNA metabolism, it is worth to mention two new putative THOC1-interactors that participates in chromatin segregation, replication and genome instability: SMC2 (a subunit of condensin

complex) and REV1 (a translesion synthesis DNA polymerase) (Nelson et al. 1996; Losada and Hirano 2005).

In summary, although these interactions remains to be validated, they are in agreement with an scenario with THO beeing a factor that plays a relevant role in a wide range of cellular processes as gene expression, cell differentiation and in the maintenance of genome integrity.

2. ROLE OF SIN3A COMPLEX IN THE MAINTENANCE OF GENOME STABILITY

In this Thesis we have provided evidence for a physical and functional interaction between THO and Sin3A-HDAC complexes. The crosstalk between THO and a chromatin modifier as the Sin3A complex extends the understanding of the mode of action of THO to connect nuclear functions, necessary for proper mRNP formation and genome integrity.

2.1. The THO/TREX RNA-processing factor talks to the Sin3A corepressor complex

THO-Sin3A physical interaction has been validated by different approaches as two-hybrid, co-immunoprecipitation assays and also with PLA analysis that detect *in situ* association of proteins (Figure R3-R4). Importantly, we show that THOC1 interacts with SAP130, as determined in the two-hybrid analysis, but also with SIN3, the core component that acts as the scaffold of the Sin3A complex (Figure R10). Moreover, the observation that UAP56 associates with SIN3 supports that THO/TREX and Sin3A complexes physically interact *in vivo* (Figure R11).

The Sin3A co-repressor complex through its histone deacetylase activity and its interaction with a large number of DNA-transcription factors regulates a wide variety of genes involved in different cellular processes in eukaryotic cells (Cowley et al. 2005; David et al. 2008; Pellegrino et al. 2012). Accordingly, proliferation of siSAP130 cells was impaired (Figure R6). Although a repressive role for the Sin3A complex in

regulating gene expression has been established, accumulating evidence has revealed that this complex also localizes inside of actively transcribed genes (Pile and Wassarman 2000; Wang et al. 2009b; Jelinic et al. 2011; Kadamb et al. 2013). This localization of the Sin3A complex inside genes, which we confirmed by ChIP-qPCR (Figure R7), is consistent with our findings of a new physical and functional interaction between THO and Sin3A-HDAC complexes.

The Sin3A complex has been shown to play different functions beyond cell cycle and proliferation, including DNA methylation, rDNA silencing, control of replication timing, centromeric heterochromatin maintenance or DNA damage repair (revised in (Silverstein and Ekwall 2005)). Here, we show that Sin3A complex is necessary for maintenance of genome integrity. Depletion by siRNA of different representative subunit of Sin3A complex leads to DNA breaks that are suppressed by transcription inhibitors and RNase H1 overexpression (Figure R12-R13), an R-loop-mediated genome instability phenotype similar to that conferred by THO mutants (Huertas and Aguilera 2003). Consistently, R loops are accumulated in Sin3A-depleted cells along a gene from 5' to 3', and not only at promoter proximal regions, as would be expected according its predominant binding around TSS (Figure R15-R16). Along this same line, although Sin3A repress transcription at promoters, accumulating evidence indicates that Sin3A also localizes inside actively transcribed genes (Figure R7) (Pile and Wassarman 2000; Jelinic et al. 2011; Kadamb et al. 2013).

Interestingly, a role of SIN3 in the maintenance of genome integrity has also been observed in yeast *sin3Δ* mutants, as observed by increased Rad52 foci and gross chromosomal rearrangements, such increases being suppressed by RNase H1 overexpression (Wahba et al. 2011). Moreover, it has been reported that *sds3Δ* mutant shows chromosome instability and elevated hybrids levels (Stirling et al. 2011; Chan et al. 2014). This is consistent with our data using mutants of the different subunits of yeast Sin3-Rpd3. We show that *rpd3Δ* and *rco1Δ* mutants increase recombination frequency, as an indirect measure of genome instability, and most of the mutants (*sin3Δ*, *rpd3Δ*, *rco1Δ*, *sap30Δ* and *pho23Δ*) show Rad52-foci accumulation which can be fully suppressed by RNase H1 overexpression (Figure R40). Altogether indicate that the role of Sin3 in the maintenance of genome integrity seems to be conserved, but importantly we show data indicating that Sin3A and THO complexes work together to avoid genome instability (discussed below).

2.2. Histone deacetylation prevents co-transcriptional R-loop formation

Given that Sin3A regulates transcription through deacetylation of nucleosomes by histone deacetylases (HDACs) and recruitment of chromatin remodelers (Silverstein and Ekwall 2005), our results suggest that histones possibly need to be deacetylated after transcription to prevent R loops. The acetylation status of chromatin can be modified by the action of histone deacetylases (HDACs) and histone acetyltransferases (HATs) which in turn have relevant consequences for gene expression, regulating the accessibility of transcription factors (Shahbazian and Grunstein 2007; Venkatesh and Workman 2015), but also for genome stability (Lahue and Frizzell 2012). Importantly, we showed here that treatment with two different HDAC inhibitors, TSA and SAHA, leads to an accumulation of R loops (Figure R18). These results are not an indirect consequence of changes in transcription levels, and support a direct role of the activity Sin3A complex in preventing R loops, as these HDAC inhibitors have been previously shown to affect the recruitment of Sin3A to chromatin (Smith et al. 2010). Histone acetylation and deacetylation alternate in a dynamic manner along the transcription process. Consistently, active genes are enriched in both HATs and HDACs, whose levels correlate with gene expression and histone acetylation (Wang et al. 2009b). The coordination of different histone modifications (acetylation, ubiquitination, methylation, etc.), and the action of chromatin remodelers is necessary for transcription (Venkatesh and Workman 2015), and our study with HDACi treatments and Sin3A depletion suggests that one objective of chromatin resetting after the passage of the RNAPII is to prevent the formation of R loops (Figure D1A), which have been shown to have a negative effect on transcription elongation indeed (Tous and Aguilera 2007). Reassembly of ubiquitylated H2A-H2B dimers mediated by the nucleosome-reorganizing complex FACT is also necessary to reset chromatin and ensure effective transcription (Belotserkovskaya et al. 2003; Li and Reinberg 2011; Venkatesh and Workman 2015). Interestingly, transcription- and R-loop-dependent genome instability has been shown in yeast and human cells deficient in the FACT reorganizing complex (Herrera-Moyano et al. 2014), consistent with our findings that co-transcriptionally mediated chromatin remodeling or chromatin modifications are crucial to prevent R-loop-mediated genome instability.

Interestingly, recent evidence reported a connection between R loops and increased chromatin accessibility. Co-transcriptional R-loop are formed along gene

body and enriched at promoters, especially CpG island promoters, and terminator regions (Santos-Pereira and Aguilera 2015; Chedin 2016). The analysis of R-loop profile together with chromatin state identified an association between R loops and open chromatin. Promoter and gene body regions with R loops especially showed hyper-accessibility, measured as increased DNase hypersensitivity and higher formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) signal. Analysis of different open and active chromatin marks showed hyper-acetylation (H3K9 and H3K27 acetylation), hyper-methylation of H3K4 (H3K4me2 and H3K4me3) and depletion of the heterochromatic mark H3K9me3 at promoter regions (Sollier and Cimprich 2015; Chedin 2016). Moreover, multiple chromatin binding factors, including the Sin3A complex, were found similarly enriched at these regions (Ginno et al. 2012; Sanz et al. 2016). This link between a more open and accessible chromatin and the fact that promoters are hotspots of R-loop formation could suggest that this chromatin state could facilitates R-loop formation.

2.3.   Functional implication of THO and Sin3A complex association in R-loop prevention

The physical interaction between THO and Sin3A and the similar R-loop genome instability phenotype upon depletion of these factors suggest that THOC1 and Sin3A could act together to prevent R-loop-mediated genome instability. The observation that RNA-DNA levels are higher in THOC1-Sin3A double-depleted cells than in the single depletion of each factor supports this idea (Figure R17). We propose that THO does not only contribute to maintaining genome stability due to its role in mRNP biogenesis but also because it could be important for proper deacetylation during transcription. We reasoned that THO could interact with Sin3A to transiently promote histone deacetylation as a way to transiently close chromatin, thus preventing the nascent RNA to hybridize with the DNA template. Although THO may not be necessary for the recruitment of Sin3A to chromatin (Figure R19A, B), it could directly or indirectly promote the co-transcriptional action of the Sin3A complex. ChIP analysis and western blot show a slight increase of global histone acetylation in siTHOC1 cells (Figure R19C, D). Recent experiments in our lab using specific antibodies raised against acetylation of individual lysine residues on histones have allowed to better appreciate differences in histone acetylation levels. Western blot analysis of histone H3 lysine 14

141

(H3K14) acetylation, a histone mark associated with active transcription (Agalioti et al. 2002), showed a high increase in Sin3A-depleted cells but also in THOC1-depleted cells (Aleix Bayona-Feliu personal communication). Importantly, we show that the HAT inhibitor anacardic acid suppresses the genome instability and the R-loop accumulation phenotype associated with THO depletion (Figure R20-R21). Therefore, these data together with the R-loop-dependent genome instability associated with Sin3A depletion, supports a model in which THO suppresses genome instability by promoting the transient action of the Sin3A complex after the passage of the RNAPII, preventing the hybridization of the nascent RNA with the template DNA (Figure D1B). In the future, it would be interesting to study whether Sin3A depletion leads to replication defects similar to those described in the absence of THO by DNA combing analysis (Wellinger et al. 2006; Dominguez-Sanchez et al. 2011a). It is possible that changes in histone acetylation would help explain, at least in part, the replication defects caused by THOC1 depletion. The observation that R loop-mediated transcription-replication collisions and genome instability are increased in yeast and human cells deficient in the FACT chromatin-reorganizing complex (Herrera-Moyano et al. 2014), supports the view that chromatin plays a crucial role preventing R loop-mediated genome instability. It would certainly be interesting to identify the mechanism by which THO, whether directly or indirectly promote the co-transcriptional action of the Sin3A complex or whether other histone deacetylases complex could contribute in a similar manner to genome stability.

In summary, our data reveals a new scenario in which an mRNP factor, such as THO, which plays a role in transcription elongation, RNA processing and export, talks to a chromatin repressive modifier to link mRNP biogenesis, chromatin and genome instability (Figure D1B). It would be interesting to address whether deficiency in others chromatin remodelers also lead to a genome instability phenotype mediated by R-loop. Moreover, it would certainly be interesting to know whether additional proteins are required for this crosstalk and whether other RNA binding factors such as UAP56, SRSF1 (Li and Manley 2005), SETX (Skourti-Stathaki et al. 2011) among others, prevent R loops via a similar ability to control chromatin remodeling. Importantly, our results suggest that specific co-transcriptional RNA binding factors may play a novel role as a transcriptional and RNA processing checkpoint to dictate post-transcription chromatin closing to prevent harmful RNA-DNA interactions.

142
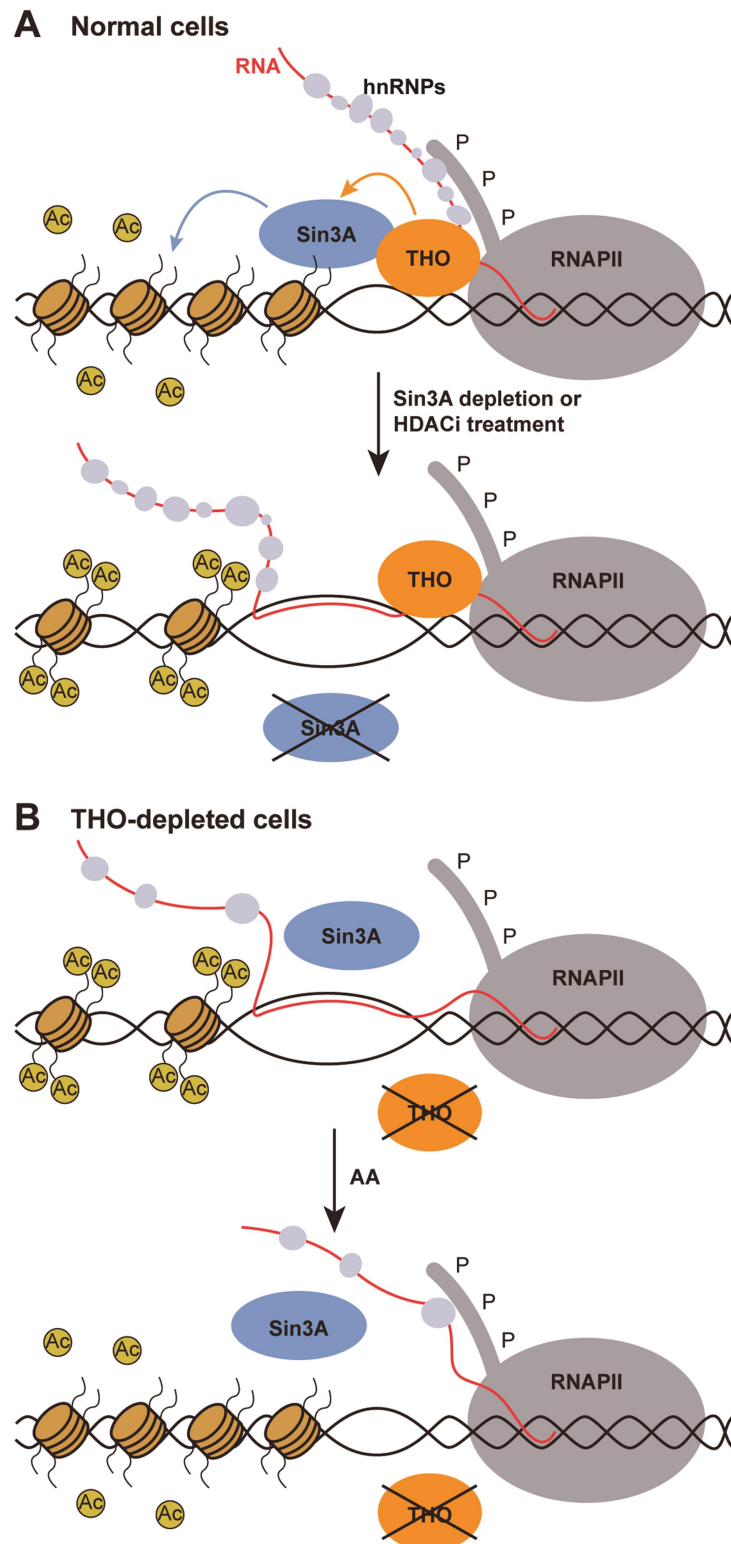
Figure D1. Model to explain the role of histone deacetylation in preventing R-loop-dependent genome instability.
(A) During transcription, both deacetylation of histones mediated by the Sin3A complex, and formation of an optimal mRNP particle mediated by THO, help prevent R-loop formation. Depletion of Sin3A or HDACi treatment lead to a hyper-acetylated and more open chromatin, thus allowing R-loop formation.
(B) Depletion of THO would lead to the formation of a suboptimal mRNP, and a lower co-transcriptional action of Sin3A, thus contributing to R-loop accumulation. However, this phenotype is suppressed by hypo-acetylation of chromatin achieved by HATi treatment. AA, anacardic acid.

143

2.4. Implication of Sin3A complex in R loop-induced epigenetic and structural changes in chromatin

Different works during last years have shown that there is a link between R loops and chromatin modulation. A set of histone modifications has been shown to correlate with R loops and its positive roles in cells, as is the case of transcription activation and termination (Santos-Pereira and Aguilera 2015). R loops formed in human CpG island promoters could promote transcription activation in association with histone marks such as H3K4me2, H3K4me3 and H3 acetylation, which are characteristic of active transcription (Ginno et al. 2012; Chen et al. 2015). R loops formed at terminator region seems to be important for termination mechanism of numerous yeast and human genes, and in human it correlates with the repressive heterochromatin mark H3K9me2 (Ginno et al. 2013; Skourti-Stathaki et al. 2014). Therefore, R loops could alter local chromatin and modulate its accessibility to carry out its regulatory functions. However these chromatin changes can also occur when R loops are formed as aberrant byproducts of transcription. H3S10-P and H3K9me2, marks of chromatin condensation and heterochromatinization respectively, are two main chromatin modifications associated with aberrant R loops (Castellano-Pozo et al. 2013; Skourti-Stathaki et al. 2014). The absence of THO in yeast, *Caenorhabditis elegans* and human cells leads to an increase of H3S10-P, which is suppressed by RNase H1 overexpression (Castellano-Pozo et al. 2013). Increased levels of H3K9me2 have been observed in *C. elegans thoc-2* mutants and at triplet repeat expansions in which R loops accumulate (Castellano-Pozo et al. 2013; Groh et al. 2014). Although transient formation of repressive chromatin may be critical for stabilizing the damaged chromatin and for remodeling the chromatin to enable efficient DNA repair (Ayrapetov et al. 2014), the aberrant accumulation of local compacted chromatin could be an obstacle for transcription and/or replication thus leading to genome instability. Therefore, the mechanism by which R loops trigger chromatin compaction and its functional relevance remains unclear. Our data indicate that although human Sin3A-depleted cells accumulate R loops (Figure R14-R16), they show no differences in H3S10-P or H3K9me2 levels (Figure R22-R23). Instead, siTHOC1-depleted cells show an increase in H3S10-P which has been reported to be R-loop-dependent (Castellano-Pozo et al. 2013) and in H3K9me2 (Figure R22-R23). Importantly, depletion of Sin3A complex is able to suppress the increase in H3S10-P and H3K9me2 caused by THOC1 depletion (Figure R22-R23). This suggests that the activity of Sin3A complex could be required in R-loop-induced chromatin compaction.

In fact, deacetylation of H4K14 by yeast Hst2 (human SirT2p) is an intermediary step in the cascade of histone modification to promote chromatin condensation in mitosis (Perrod et al. 2001; Wilkins et al. 2014). On the other hand, suppression of chromatin condensation in *hpr1Δ* yeast mutant in combination with specific histone H3 point mutation at serine 10 (S10) that cannot be phosphorylated (H3S10A) or that mimic constitutive phosphorylation (H3S10D) lead to a decrease of genome instability, measured as a decrease in recombination frequency and in the percentage of Rad52 foci (Castellano-Pozo et al. 2013). Our data show that Sin3A depletion reduces the high levels of γH2AX foci caused by THOC1 depletion in human cells. Moreover, the absence of Sin3 in yeast *hpr1Δ* mutant greatly reduces its high levels of recombination and Rad52 foci. These results suggest that Sin3 complex is required for the genome instability caused by R loops and support the idea that R loop-dependent genome instability may be due in part to the chromatin compaction they trigger (Figure D2).
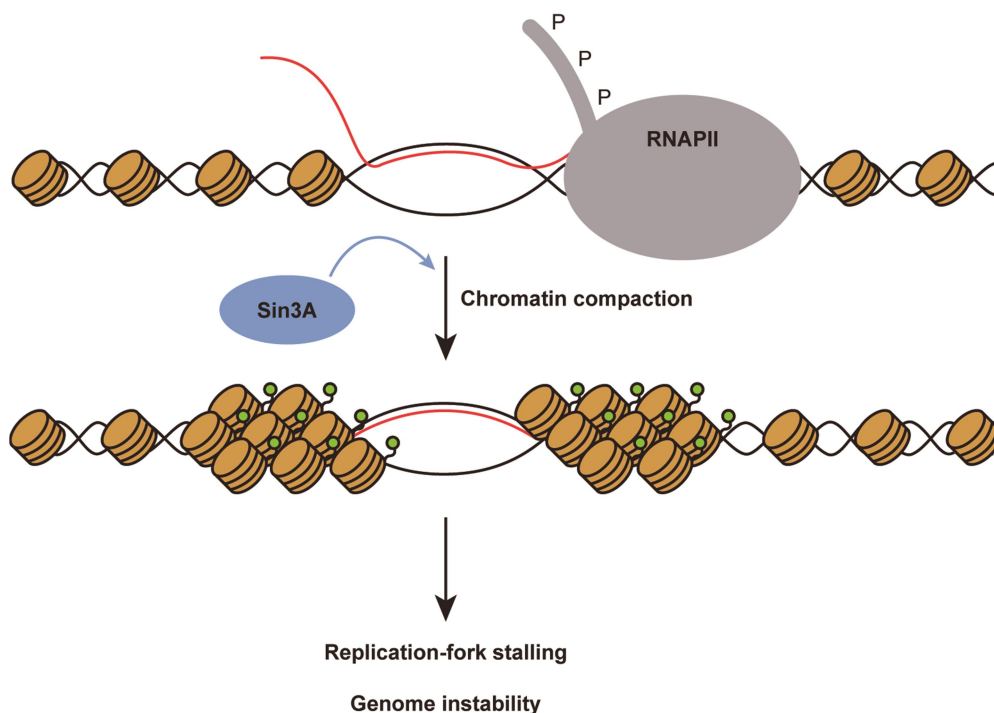
Figure D2. A model to explain a possible role of Sin3A complex in R-loop-induced chromatin compaction.
R loops trigger chromatin condensation and heterochromatin formation. These more compacted areas within the chromatin can hamper the progression of replication, which in turn, leads to genome instability. The action of the Sin3A complex may be required in the cascade to compact the chromatin. Model based in (Santos-Pereira and Aguilera 2015).

3.   ROLE OF MFAP1 IN THE MAINTENANCE OF GENOME STABILITY

We have reported a novel interaction between the splicing factor MFAP1 and THOC1. A number of studies have determined that THO/TREX plays an important role connecting the different steps of mRNP biogenesis (Aguilera 2005; Rondon et al. 2010). The main goal of our study has been to try to elucidate the physiological relevance of the interaction between MFAP1 and THOC1 at the interface of mRNP biogenesis and genome instability.

3.1.   The relationship between THO/TREX, splicing and mRNP biogenesis

THOC1 and MFAP1 interaction, validated in human cells by different approaches (Figure R3-R4), is in agreement with previous studies showing a relationship between THO/TREX and the splicing machinery. In fact, THO/TREX subunits were first identified in proteomic analysis as protein associated to the spliceosome (Rappsilber et al. 2002; Zhou et al. 2002; Chen et al. 2007). Moreover, THO/TREX complex was found to colocalize with splicing factors in nuclear speckle domains in vivo (Zhou et al. 2000; Masuda et al. 2005; Kota et al. 2008), as we have also shown for MFAP1 (Figure R25).

According this interaction, analysis of data from proteomic spliceosome studies reveals that THOC1 and MFAP1 have common protein interactors as is the case of RED/IK and hPRP22/DHX8 (Hegele et al. 2012). IK is a component of B complex of human spliceosome and helps recruit other factors, and DHX8 is present at C complex and participates in the release of spliced mRNAs from spliceosome (Ohno and Shimura 1996; Hegele et al. 2012). Interestingly, THO/TREX proteins are important for efficient export of spliced mRNA from the nuclear speckles to the cytoplasm (Dias et al. 2010), consistently with the view of THO/TREX as a complex linking different steps of mRNP biogenesis (Aguilera 2005). Whether MFAP1 has a role in mRNA export and if MFAP1 and THOC1 interaction have a physiological relevance on the export of spliced mRNAs is a question that will require further investigation.

The link between splicing and THO/TREX seems to be especially important in metazoans, where most of the genes contain introns. In yeast, THO/TREX is directly recruited co-transcriptionally by the transcription machinery whereas in metazoans,

146

THO/TREX is recruited to mRNA in a splicing- and capping-dependent manner (Zhou et al. 2000; Custodio et al. 2004; Masuda et al. 2005; Reed and Cheng 2005). The splicing complex Prp19 complex or also called NTC plays a relevant role in the recruitment of TREX to elongating RNAPII, and in the coupling of transcription elongation, splicing and mRNA export (Chanarat et al. 2011). Since both THO/TREX and numerous splicing factors, including MFAP1, closely associate to the Prp19 complex one possibility was that MFAP1 could contribute to the loading of THO/TREX. However, THOC1 recruitment to the chromatin was not impaired in siMFAP1 depleted cells as determined by ChIP analysis (Figure R29C).

Splicing has an impact on the regulation of multiple steps in mRNP biogenesis, thus coupling connections between splicing and transcription elongation 3′ end formation, RNA stability, and mRNA export have been reported (Kornblihtt et al. 2004; Braunschweig et al. 2013; Saldi et al. 2016). We have explored the impact of MFAP1 depletion in transcription and 3′ end mRNA processing by ChIP analysis (Figure R29). The levels of total RNAPII in siMFAP1 cells were found to be similar to those found in control cells, suggesting that depletion of this factor does not have a direct role on transcription. However, genome-wide analysis of microarray showed that depletion of MFAP1 has an impact on gene expression (discussed below). In turn, the levels of FIP1 and CPSF6 factors (components of 3′ end processing machinery) are lower in MFAP1-depleted cells than in siC control cells and significantly lower in the case of FIP1 at 3′end of the gene (Figure R29D, E). Since these factors are involved in the cleavage and polyadenylation of pre-mRNA, this data suggests that MFAP1 could be important in later pre-mRNA processing steps. In this same line, TREX plays a key role in 3′ end processing. In yeast, Yra1 is involved in the recruitment of 3′ end processing factors (Johnson et al. 2009). Moreover, the absence of yeast THO leads to the attachment of the 3′ end region of several genes to the nuclear pore, which can block RNAPII progression, and inefficient polyadenylation (Rougemaille et al. 2008; Saguez et al. 2008). This is mediated in part by selective degradation of Fip1, a key component of the polyadenylation machinery (Preker et al. 1995; Saguez et al. 2008). Importantly, human THOC5 has been reported to control 3′ end processing by recruiting cleavage and polyadenylation factors including CPSF100 and CFIm68 and therefore the connection between THO/TREX and 3′ end processing seems to be conserved in humans (Katahira et al. 2013; Tran et al. 2014).

3.2. Proliferation and cell cycle defects in MFAP1-depleted cells

Several studies have evidenced defects in the cell cycle progression in the absence of MFAP1. A siRNA-based screening in human cells identified a group of splicing factors, including MFAP1, as essential factors for cell division (Kittler et al. 2004). In this screening it was observed that MFAP1-depletion caused mitotic arrest, cytokinesis defect and abnormal spindle. A more recent screening to look for new mitotic regulator in human cells reported that lack of functional MFAP1 leads to abnormal nuclear morphology, mitotic arrest and defects in the segregations of chromosomes during cells division (Sundaramoorthy et al. 2014). Similarly, reduced levels of *Drosophila* homologue of MFAP1 was reported to increase the proportion of cells arrested in G_2/M phases (Andersen and Tapon 2008). Our data corroborate these defects in cell cycle progression. We show that depletion of MFAP1 seriously affects cells proliferation and FACS analyses clearly show an arrest in G_2/M, which may explain the increase in apoptotic cells (Figure R27).

There are many examples of splicing factors that affect cell cycle progression. *S. cerevisiae* Cef1 protein is involved in pre-mRNA processing and its loss causes arrest in G_2. The removal of one intron of *α-tubulin* (*TUB1*) mRNA is able to reestablish the cell cycle of *cef1-13* thermosensitive mutant, suggesting that cell cycle arrest is an indirect effect of defective splicing (Burns et al. 2002). In other cases, the splicing factors could have a direct role in cell cycle regulation. One example of this is the SR protein B52, a serine-rich protein that regulates splicing and levels of dE2F2, a transcriptional factor relevant for cell cycle progression in *Drosophila* (Rasheva et al. 2006). However, it is unclear whether MFAP1 or other splicing factors affect cell cycle due to a direct role in cell cycle regulation or as an indirect consequence of altering the expression and/or the splicing of some related genes. In *Drosophila*, the arrest in G_2/M in MFAP1 or Prp38-depleted cells is associated with a reduction of *stg/cdc25* mRNA levels, which encodes a phosphatase essential for G_2/M progression (Andersen and Tapon 2008). It is unknown whether *stg* downregulation and the consequent cell cycle arrest corresponds to defects in its splicing or whether defective splicing could activates specifically a mitotic checkpoint that regulates stg. Our gene expression and alternative splicing analyses in MFAP1-depleted cells provide clues about the direct or indirect role of this factor in cell cycle and other processes (see below).

3.3. Genome instability associated to MFAP1-depletion

Our data show that lack of MFAP1 in human cells compromises genome stability, since depletion of this factor lead to an increase in DNA damage, as measure by γH2AX foci immunofluorescence, alkaline and neutral single-cell electrophoresis and anaphase bridges formation (Figure R30-R31). Interestingly, MFAP1 associated genome instability seems not be associated to R-loop formation, since overexpression of RNase H1 does not suppress the increase in γH2AX foci caused by MFAP1 depletion (Figure R32B). Our functional analysis in *S. cerevisiae* with mutants of *SPP381* gene, the putative MFAP1 orthologue in yeast, as determined by homologue sequence, reveals a similar genome instability phenotype with high increase in DNA damage that is not R-loop dependent (Figure R44C, D). These results are in agreement with previous studies showing that *spp381* mutant and other splicing factors lead to chromosome instability (CIN) phenotype while RNA-DNA hybrid levels appear to be not increased (Stirling et al. 2011; Chan et al. 2014). Altogether indicates that MFAP1 role in the maintenance of genome is conserved along the evolution.

Several splicing and other RNA processing factors have been identified in different global screenings in yeast and humans as factors involved in the maintenance of genome instability (Paulsen et al. 2009; Wahba et al. 2011; Stirling et al. 2012; Chan et al. 2014). Provided the functional proximity of splicing factors to THO/TREX and SRSF1 (Li and Manley 2005), both suppressors of R loops and R-loop-mediated genome instability, it is sometimes assumed that genome instability promoted by deficiency in most RNA binding proteins and splicing factors are of the same nature. However, our study argues against this idea. Since genome instability caused by MFAP1 depletion cannot be suppressed by transcription inhibition or by overexpression of RNase H1 (Figure R32), it is possible that deficiencies in a large number of splicing factors could lead to genome instability by a different mechanism.

Therefore, it is likely that mutations or depletion of many splicing and processing factors cause genome instability as an indirect effect, as it seems to be the case of MFAP1, which has been shown to affect mRNA splicing and sister chromatid cohesion likely due to the putative impact that they may have in alternative splicing and expression of other genes (Andersen and Tapon 2008; Sundaramoorthy et al. 2014; Papasaikas et al. 2015). As a consequence, this result strengthens the idea that only

specific RNA processing and binding proteins, such as THO/TREX or SRSF1, has a direct role in the control of genome stability.

3.3.1. Gene expression and alternative splicing are affected in MFAP1-depleted cells

To shed light on whether MFAP1 could influence the stability of the genome in a direct or indirect manner, we performed global gene expression and alternative splicing analyses in siMFAP1 cells. We found that MFAP1 depletion affects the expression of 904 genes (P < 0.05 and |linear fold change| > 1.5), of which 359 (39.7%) are down-regulated and 545 (60.3) are up-regulated (Figure R33A). Down-regulated genes are enriched in mRNAs (49%) and ncRNAs (36.2%) (Figure R33B, left panel). GO analysis of these genes identified numerous biological processes that were redundant since down-regulated genes that belong to these categories were histone genes in all cases (Table R3). Histone genes are found in clusters and its transcription is replication-dependent (Marzluff et al. 2002). The levels of histone mRNAs are tightly regulated reaching the maximum levels at S-phase to be coupled to DNA synthesis process (Harris et al. 1991; Nelson et al. 2002). Moreover, it has been shown that G1 checkpoint activation after DNA damage induced by ionizing radiation not only inhibits DNA synthesis but also inhibit histone genes transcription (Su et al. 2004). Since MFAP1 depletion seems not to arrest cell cycle in G1 or S-phase, a lower detection of histone mRNAs could be due to the reduction in S-phase population observed in siMFAP1 cells (Figure R27B). In addition to histone genes, several genes related to DNA repair were down-regulated and some of them, including *FANCD2*, *RNASEH2* and *EGFR* were also found differentially spliced suggesting that MFAP1 could be mediating in the regulation of DNA repair genes (Table R3 and Appendix 4). Up-regulated genes are enriched in U snRNA, which is consistent with a role of MFAP1 in pre-mRNA splicing (Figure R33B, right panel). GO analysis identified several genes related to hypoxia, apoptotic process or DNA damage response among others, which are probably up-regulated in response to the different stress conditions in siMFAP1 cells (Table R4).

Analysis of splicing-sensitive microarray data identified that MFAP1 has a large impact on splicing since its depletion leads to splicing changes in 3689 genes (5490 ASEs) (P < 0.01 and |Splicing Index| > 2). Until now, there are some studies in different organisms showing that the absence of MFAP1 leads to changes in splicing

and suggesting that the observed phenotypes are possibly an indirect consequence of the alteration of the splicing of certain genes (Andersen and Tapon 2008; Ma et al. 2012; Sundaramoorthy et al. 2014). Reduced levels of *Drosophila* MFAP1 causes cells to arrest in G_2/M due to direct or indirect down-regulation of mitotic string/cdc25 phosphatase, which is necessary for G_2/M transition (Andersen and Tapon 2008). In human cells, MFAP1 depletion leads to defects in sister chromatid cohesion as a consequence of impaired splicing of Sororin (*CDCA5*), a protein required for cohesion and for stable association of cohesion with chromatin (Sundaramoorthy et al. 2014). Based on our analysis, MFAP1 depletion does not seem to affect the mRNA expression levels of any of the three human Cdc25 phosphatases (*CDC25A*, *CDC25B*, *CDC25C*) and neither of Sororin (*CDCA5*). However, both all *CDC25* phosphatases and *CDCA5* are differentially spliced in siMFAP1 cells (Appendix 12), which could be interfering with the correct functioning of these proteins.

Importantly, AS analysis in MFAP1-depleted cells reveals an enrichment of AS changes in genes involved in important processes for DDR such as DNA repair, cell cycle and chromatin organization and modification (Table R5). It is clear that DNA damage and the activation of DDR triggers a variety of global changes aimed at preserving the integrity of the genome. Depending on the type of DNA lesion, different signal transduction cascades and downstream pathways are regulated to coordinate cell cycle, DNA repair and apoptosis, thus ensuring the stability of the genome (Jackson and Bartek 2009; Ciccia and Elledge 2010). All this implies the participation of multitude of factors as well as post-translational modifications and changes not only in gene expression but also splicing pattern. There is increasing evidence that splicing plays very important roles at different steps in response to DNA damage, however the mechanisms are unclear.

On one hand, depletion of many splicing factors leads to DNA damage, which in many cases is mediated by R-loop formation (Figure D3A) (Li and Manley 2005; Paulsen et al. 2009; Wan et al. 2015). Moreover, recent evidences suggest that modulation of spliceosome is important to sense the DNA damage and trigger the DDR (Shkreta and Chabot 2015). It has been shown that blocked transcription after DNA damage promotes selective chromatin displacement of spliceosome, especially late-state spliceosomes, thus favoring R-loop formation, which in turn activates DDR (Figure D3A) (Sordet et al. 2009; Tresini et al. 2015). However, this does not seem to be the

case of MFAP1 since it participates at early steps of spliceosome assembly (B complex activation) and also does not lead to R-loop-dependent DNA damage (Figure R32).

On the other hand, initial stages in response to DNA damage are based on post-translational modification (acetylation, ubiquitylation, sumoylation, phosphorylation, and others) of a variety of factors to regulate their localization, stability and activity. Many of them are spliceosomal proteins and splicing regulators (Figure D3B) (Shkreta and Chabot 2015). It would be interesting to address in the future whether MFAP1 is one of these factors, in fact, the protein sequence contains multiple phosphorylation sites (Olsen et al. 2006); or whether its expression or localization changes after DNA damage.

Finally, all these changes in activity, localization and post-translational modifications have the purpose of producing an effective response to fight the DNA damage. Genes involved in DNA repair, cell-cycle control and apoptosis are the main responsible for carrying out this response through changes in their expression levels but also in their alternative splicing (Figure D3C) (Lenzken et al. 2013; Shkreta and Chabot 2015). In some cases, alternative splice variant could be more efficient; however in many cases the functional impact is still unknown (Shkreta and Chabot 2015). Interestingly, the analysis of six studies in different cell lines that used camptothecin, UV irradiation and sodium arsenate to induce DNA damage reported AS changes in genes associated with DNA repair, cell-cycle control, apoptosis and chromatin organization and modification (Shkreta and Chabot 2015). Although DNA damage and cell cycle arrest observed in MFAP1 depleted cells could be indirect, the fact that GO and cluster analysis show an enrichment of AS changes in genes involved in DNA repair, cell cycle and chromatin organization and modification (Table R5), which is important for DNA repair process, suggests that MFAP1 could play a more direct role in the maintenance of genome integrity than previously thought. MFAP1 could influences splicing response with the aim of counteracting DNA damage. The specific contribution of MFAP1 to spliceosome dynamics and splicing decisions is a question that needs to be addressed in the future. In fact, a recent study has reported the regulatory potential of spliceosomal components since depending on the order and duration of their recruitment influences on alternative splice site selection of target genes (Papasaikas et al. 2015).
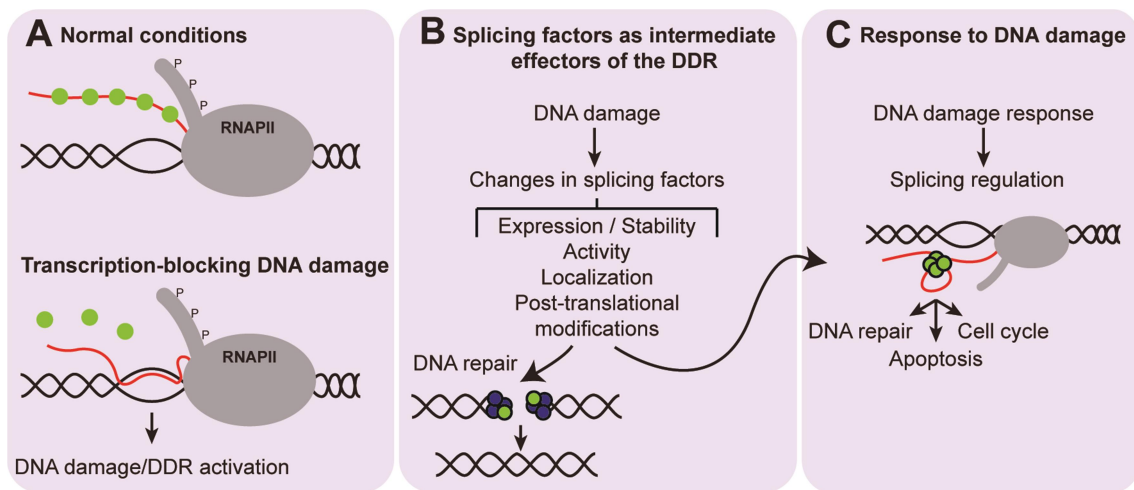
Figure D3. Role of splicing factors in genome stability and in the DDR.
(A) In normal conditions, many splicing-related proteins contributes to protect the nascent RNA and therefore to prevent R-loop formation. Under transcription-blocking DNA damage conditions, the spliceosome displacement and the consequent R-loop formation could serve as a mechanism to amplify the DDR signaling pathway. **(B)** Intermediate stages of DDR require the action of many splicing factors to coordinate late steps of the DDR. For this, splicing factors undergo changes in their expression/stability, activity or localization, which is largely mediated by post-translational modifications. **(C)** Finally, changes in expression levels and splicing of genes involved in DNA repair, cell cycle and apoptosis allow the cell to successfully respond to DNA damage. Splicing factors (green circles); DNA repair proteins (blue circles).

3.3.2. THOC1-MFAP1 interaction links alternative splicing and genome instability

Given the physical interaction, to further explore the functional association of THOC1 with MFAP1 and with splicing we also carried out global gene expression and alternative splicing analyses in THOC1-depleted cells. Gene expression of 84 genes ($P < 0.05$ and |linear fold change| > 1.5) are affected in siTHOC1 cells, of which 52 (61.9%) are down-regulated and 32 (38.1%) are up-regulated (Appendix 5). These data suggest that THOC1 depletion does not seem to have a large impact on gene expression, as previously observed in a stable cell line expressing an inducible shTHOC1 (Dominguez-Sanchez et al. 2011a). Importantly, gene expression analysis of siTHOC1 and siMFAP1 indicate up-regulation in a large proportion of U snRNA but also small nucleolar RNAs (snoRNAs) (Appendix 5, Figure R33 and Figure R35). U snRNAs are an essential component of the spliceosome. The main function of snoRNAs is to guide some RNAs, as is the case of rRNA, for post-transcriptional modifications. Moreover, recent studies suggest that snoRNAs and shorter RNA form derived from them may have other functions, including regulation of splicing (Khanna and Stamm 2010; Kishore et al. 2010; Williams and Farzaneh 2012). Interestingly, it has been shown that

yeast THO complex, through its physical and functional connections with the TRAMP polyadenylation complex, negatively controls the expression of snoRNAs (Larochelle et al. 2012). Consistently, yeast THO mutants result in increased levels of mature snoRNAs, which is in agreement with our data.

AS analysis reveals that 921 genes (1022 ASEs) show AS changes in THOC1-depleted cells (P < 0.01 and |Splicing Index| > 2) (Appendix 8). Remarkably, an enrichment of these AS changes occurs in genes involved in DNA repair and cell cycle, as in the case of siMFAP1 cells (Table R6). In fact, many genes differentially spliced are common in siTHOC1 and siMFAP1 cells. Among them, it is important to notice the presence of genes involved in DNA repair and cell cycle but also RNA transport, supporting a physical and functional association of THOC1 and MFAP1 (Figure R37). In agreement with this data, gene expression profiling in cells depleted of THO components in *Drosophila* revealed a notably up-regulation of genes involved in DNA repair (Rehwinkel et al. 2004). Given the direct role of THO in the maintenance of genome stability, it is probable that part of splicing changes observed occur indirectly in response to transcription and R-loop-dependent genome instability caused by THOC1 depletion. However, our results suggest that THO may also affect alternative splicing through interactions with MFAP1 and maybe other factors (Figure D4). In fact, different genome-wide analyses performed in order to identify alternative splicing regulators have identified not only canonic splicing factors but also proteins involved in other steps of mRNA processing. Depletion of several subunits of the THO/TREX complex were found to affect AS of *Fas/CD95*, which is important for apoptosis signaling pathway (Tejedor et al. 2015). Many others mRNA processing factor were identified as regulators of *Bcl-x* AS, also involved in apoptosis (Moore et al. 2010). Other examples are RBMX, an hnRNP with a role in splice site selection regulation or the RNA helicases DDX5 and DDX17, which had been shown to regulate AS of several DNA- and chromatin-binding factors (Heinrich et al. 2009; Dardenne et al. 2012). All this highlights the importance of pre-mRNA processing in the maintenance of genome stability. Further understanding of how THO influences splicing dynamics would contribute to shed light on the molecular mechanisms of THO to help maintain genome integrity.
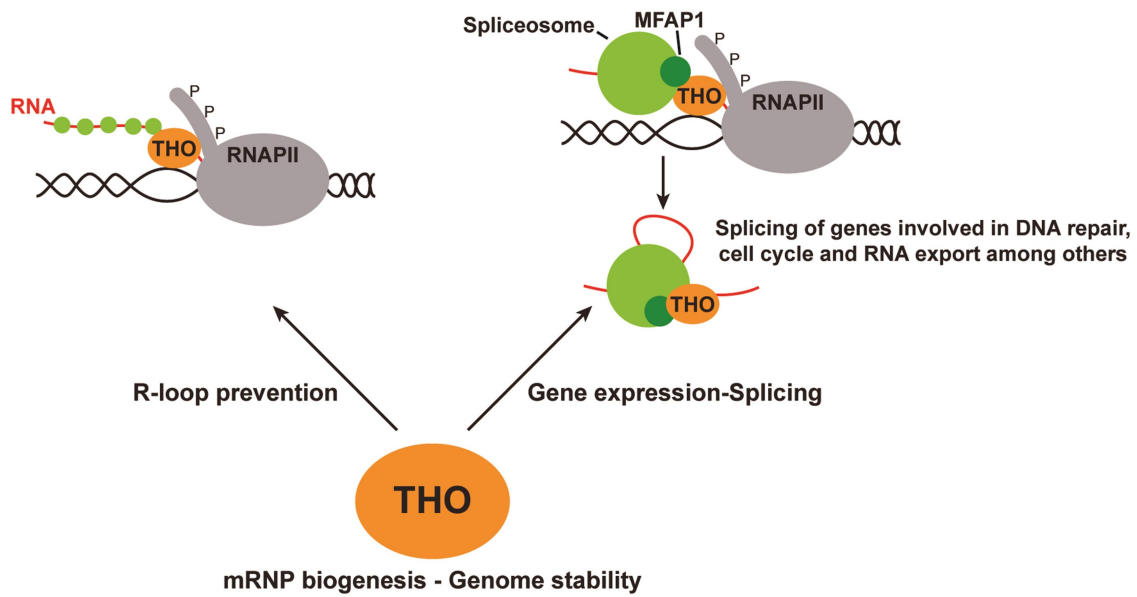
Figure D4. A dual role of THO in maintaining genome integrity.
The THO complex helps prevent genome instability by different mechanisms through its role in mRNP biogenesis. THO binds co-transcriptionally to the nascent mRNA and contributes to the formation of an export-competent mRNP, thus protecting RNA from degradation and from R-loop formation and preparing it to be exported. Moreover, THO could help maintain genome stability by modulating the splicing of genes involved in response to DNA damage through its association with MFAP1 and maybe other splicing-related proteins. Splicing factors (green circles).

CONCLUSIONS / CONCLUSIONES

CONCLUSSIONS

1. THOC1 interacts with the Sin3A histone deacetylase complex and with the spliceosome-associated factor MFAP1 in human cells.

2. Depletion of the Sin3A complex causes transcription and R loop-mediated genome instability, as determined by the accumulation of DNA breaks.

3. Co-transcriptional histone deacetylation is important to prevent R-loop formation, as supported by the fact that histone hyper-acetylation lead to high R-loop levels detected by S9.6 immunofluorescence and DRIP analysis in cells depleted of Sin3A or treated with histone deacetylase inhibitors.

4. Genome instability and R-loop accumulation in THOC1-depleted cells is mediated by improper histone deacetylation, caused by a defective crosstalk between THO and Sin3A since inhibition of histone acetyltransferases suppresses genome instability and R-loop accumulation in THOC1-depleted cells.

5. Sin3A depletion suppresses the increase of histone marks associated with R loops such as H3S10-P and H3K9me2.

6. Depletion of MFAP1 leads to proliferation and cell cycle defects and to an increase in DNA damage that is not R-loop dependent.

7. Depletion of MFAP1 has a wide effect in splicing that is mainly observed at alternative splicing of genes involved in DNA repair, cell cycle and chromatin organization and modification, suggesting that MFAP1 role in genome integrity is indirect, presumably mediated by the genes it regulates.

CONCLUSIONES

1. THOC1 interacciona con el complejo histona desacetilasa Sin3A y con MFAP1, un factor asociado al madurosoma, en células humanas.

2. El silenciamiento de componentes del complejo Sin3A causa inestabilidad genética mediada por transcripción y bucles R (*R loops*), determinada por la acumulación de roturas en el ADN.

3. La desacetilación de histonas durante la transcripción es importante para prevenir la formación de *R loops*. Esto se ve apoyado por el hecho de que un estado hiperacetilado de las histonas conduce a altos niveles de *R loops*, como se detecta mediante inmunofluorescencia con el anticuerpo S9.6, así como mediante análisis de inmunoprecipitación de híbridos de ADN-ARN (DRIP), en células silenciadas para Sin3A o tratadas con inhibidores de desacetilasas de histonas.

4. La inestabilidad genética y la acumulación de *R loops* en ausencia de THOC1 están mediadas por una inapropiada desacetilación de histonas, causada por una falta de interacción funcional entre THOC1 y Sin3A. De acuerdo con esta conclusión, la inhibición de acetiltransferasas de histonas suprime la inestabilidad genética y la acumulación de *R loops* causadas por la ausencia de THOC1.

5. El silenciamiento de componentes del complejo Sin3A suprime el incremento de marcas de histonas asociadas con *R loops* tales como H3S10-P y H3K9me2.

6. El silenciamiento de *MFAP1* produce defectos en la proliferación y ciclo celular, además de un incremento de daños en el ADN, que no es dependiente de la formación de *R loops*.

7. El silenciamiento de *MFAP1* tiene un amplio efecto en la maduración de intrones (*splicing*), principalmente en el *splicing* alternativo de genes implicados en la reparación del ADN, en el ciclo celular y en la organización y modificación de la cromatina. Esto sugiere que el papel de MFAP1 en la integridad del genoma es indirecto, presumiblemente mediado por los genes que este factor regula.

APPENDIX

Appendix 1. Down- and up-regulated genes in MFAP1-depleted cells.

http://bit.ly/2kxMZ3U

Appendix 2. Differentially spliced genes in MFAP1-depleted cells.

http://bit.ly/2lTuXte

Appendix 3. GO analysis of differentially spliced genes in MFAP1-depleted cells.

http://bit.ly/2ls41Oh

Appendix 4. Genes with altered expression levels and splicing patterns in MFAP1-depleted cells.
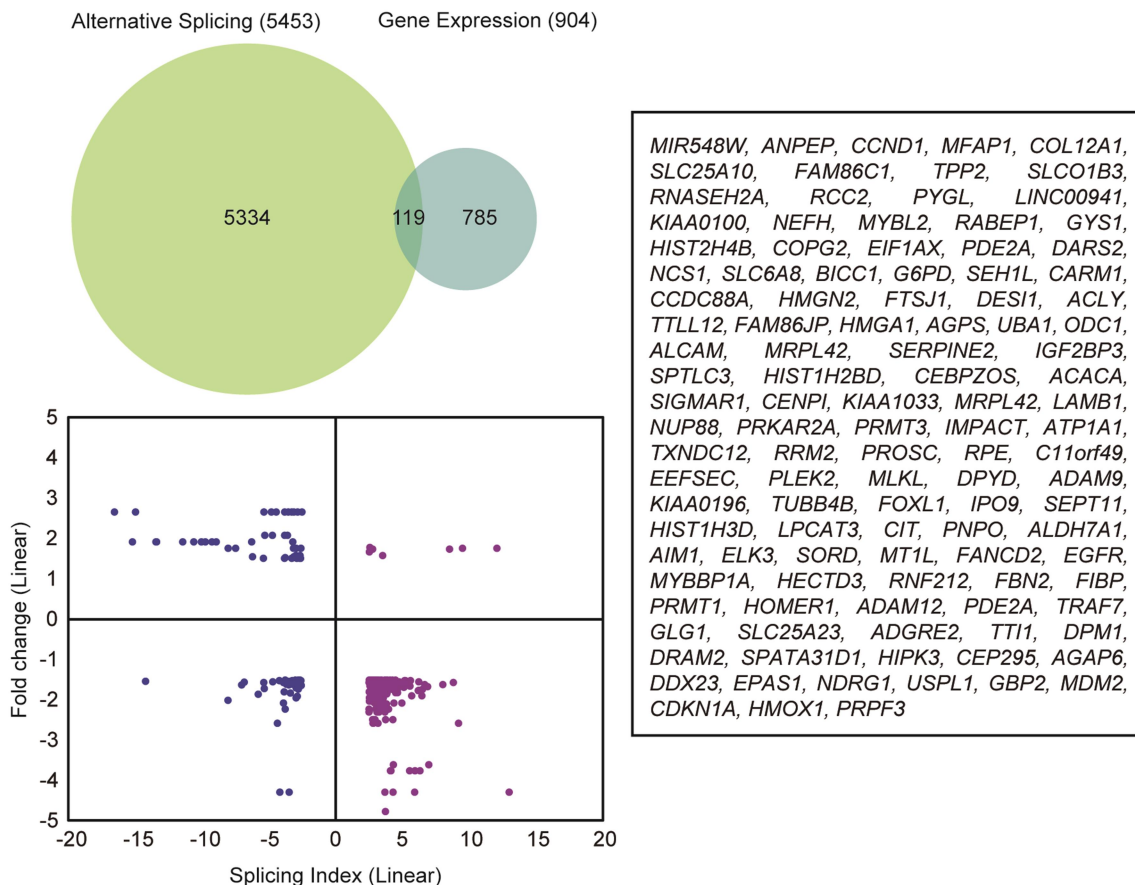
Venn diagram showing the overlap between the list of genes with abnormal splicing patterns and the list of genes with altered expression levels as provided by Pangloss (top, left panel). Graph showing the genes significantly altered in splicing (P <0.01; |Splicing Index| >2) and expression levels (P <0.05; |linear fold change| >1.5) (bottom, left panel). List of genes with alterations in splicing and gene expression in MFAP1-depleted cells (right panel).

Appendix 5. Down- and up-regulated genes in THOC1-depleted cells.

Gene symbol (or accession number)	Description	Fold Change (Linear)	ANOVA P-value
DQ586720	piRNA	-6.12	0.005
NAP1L1	mRNAlike lncRNA nucleosome assembly protein 1-like 1	-5.02	0.001
THOC1	THO complex 1	-3.93	0.001
DQ576853	piRNA	-2.74	0.031
DQ586951	piRNA	-2.6	0.030
G3BP1	mRNAlike lncRNA GTPase activating protein (SH3 domain) binding protein 1	-2.45	0.005
TC14001853.hg.1*	lncRNA	-2.36	0.008
CHORDC1	mRNAlike lncRNA cysteine and histidine rich domain containing 1	-2.33	0.029
DQ579470	piRNA	-2.24	0.008
DQ572964	piRNA	-2.03	0.025
DQ577389	piRNA	-2.03	0.027
DQ596092	piRNA	-2.03	0.046
PKI55	DKFZp434H1419	-1.98	0.023
AK095091	mRNAlike lncRNA	-1.92	0.007
YY1	mRNAlike lncRNA YY1 transcription factor	-1.88	0.003
DQ584981	piRNA	-1.87	0.032
TC05002264.hg.1*	lncRNA	-1.86	0.027
TC02004518.hg.1*	lncRNA	-1.85	0.008
PDE10A	mRNAlike lncRNA phosphodiesterase 10A	-1.85	0.030
DQ593983	piRNA	-1.8	0.046
DQ593983	piRNA	-1.8	0.046
DQ590616	piRNA	-1.8	0.046
DQ590616	piRNA	-1.8	0.046
LINC00657	long intergenic non-protein coding RNA 657	-1.8	0.033
TC02001272.hg.1*	lncRNA	-1.73	0.023
RBM26	mRNAlike lncRNA RNA binding motif protein 26	-1.73	0.011
REEP5	receptor accessory protein 5	-1.72	0.001
TC17000111.hg.1*	-	-1.71	0.002
OR5V1	olfactory receptor, family 5, subfamily V, member 1	-1.7	0.016
CR608583	mRNAlike lncRNA	-1.67	0.038
HNRNPH2	heterogeneous nuclear ribonucleoprotein H2 (H)	-1.67	0.003
SIPA1L1	mRNAlike lncRNA signal-induced proliferation-associated 1 like 1	-1.67	0.037
TC06002070.hg.1*	-	-1.64	0.021
SBNO1	mRNAlike lncRNA strawberry notch homolog 1 (Drosophila)	-1.64	0.014
AK055438	mRNAlike lncRNA	-1.63	0.009
HELZ	mRNAlike lncRNA helicase with zinc finger	-1.61	0.031
ARIH2	mRNAlike lncRNA ariadne RBR E3 ubiquitin protein ligase 2	-1.6	0.030
PKI55	DKFZp434H1419	-1.59	0.018

Gene symbol (or accession number)	Description	Fold Change (Linear)	ANOVA P-value
HIST2H2BE	histone cluster 2, H2be	-1.57	0.002
DQ588775	piRNA	-1.57	0.023
KDSR	mRNAlike lncRNA 3-ketodihydrosphingosine reductase	-1.57	0.020
TC20001195.hg.1*	lincRNA	-1.57	0.039
TC03002702.hg.1*	lincRNA	-1.56	0.007
AK057721	mRNAlike lncRNA	-1.56	0.013
C16orf72	chromosome 16 open reading frame 72	-1.56	0.021
DQ584119	piRNA	-1.55	0.005
FAM210B	mRNAlike lncRNA family with sequence similarity 210, member B	-1.55	0.047
MIR4653	microRNA 4653	-1.54	0.045
TC02003815.hg.1*	lincRNA	-1.52	0.014
TC01006059.hg.1*	lncRNA	-1.51	0.008
TC09002110.hg.1*	lncRNA	-1.51	0.014
DQ578136	piRNA	-1.51	0.026
SNORD82	small nucleolar RNA, C/D box 82	1.51	0.016
TC06000154.hg.1*	U6 small nuclear 987, pseudogene	1.51	0.002
OR4M2	olfactory receptor, family 4, subfamily M, member 2	1.51	0.015
SNORD3C	small nucleolar RNA, C/D box 3C	1.51	0.018
DDX39A	lncRNA DEAD (Asp-Glu-Ala-Asp) box polypeptide 39A	1.51	0.000
TC01003213.hg.1*	U6 small nuclear 1309, pseudogene	1.52	0.005
TC10002372.hg.1	lncRNA	1.52	0.018
DQ600737	piRNA	1.52	0.018
SAA3P	serum amyloid A3 pseudogene	1.53	0.026
TC12000580.hg.1*	U6 small nuclear 166, pseudogene	1.53	0.034
SNORD3D	small nucleolar RNA, C/D box 3D	1.53	0.019
RNU5E-1; RNU5D-1	RNA, U5E small nuclear 1; RNA, U5D small nuclear 1	1.54	0.020
TC02001559.hg.1*	ncrna:snoRNA	1.56	0.021
PHLDA3	mRNAlike lncRNA pleckstrin homology-like domain, family A, member 3	1.58	0.049
GDF15	growth differentiation factor 15	1.6	0.036
RNU6-30P	RNA, U6 small nuclear 30, pseudogene	1.61	0.023
DQ601004	piRNA	1.61	0.048
SUPT6H	mRNAlike lncRNA SPT6 homolog, histone chaperone	1.61	0.022
DQ576856	piRNA	1.62	0.006
TC11001451.hg.1*	ncrna:snRNA	1.62	0.002
LIF	leukemia inhibitory factor	1.62	0.007
PAPPA	pregnancy-associated plasma protein A, pappalysin 1	1.65	0.001
TC18000218.hg.1*	U6 small nuclear 116, pseudogene	1.67	0.000
MIR3661	microRNA 3661	1.69	0.027
BC014579	mRNAlike lncRNA	1.69	0.041
NR_037852_8	lncRNA	1.71	0.011
SNORD3B-1	small nucleolar RNA, C/D box 3B-1	1.76	0.004

Gene symbol (or accession number)	Description	Fold Change (Linear)	ANOVA P-value
SNORD3B-2	small nucleolar RNA, C/D box 3B-2	1.76	0.004
STAMBPL1	STAM binding protein-like 1	1.84	0.026
MIR4434	microRNA 4434	2.19	0.013
HMOX1	heme oxygenase 1	2.6	0.012
HMOX1	mRNAlike lncRNA heme oxygenase 1	2.87	0.041

* In case of neither gene symbol nor accession number, the transcript cluster ID has been provided.

Appendix 6. Common down-regulated genes in MFAP1- and THOC1-depleted cells.

Gene Symbol (or accession number)	Description	Fold Change (Linear)		ANOVA P-value	
		siMFAP1	siTHOC1	siMFAP1	siTHOC1
DQ586720	piRNA	-3.01	-6.12	0.0314	0.0046
DQ579470	piRNA	-2.61	-2.24	0.0064	0.0082
TC09002110.hg.1*	lncRNA	-2.37	-1.51	0.0226	0.0138
G3BP1	mRNAlike lncRNA GTPase activating protein (SH3 domain) binding protein 1	-2.29	-2.45	0.0011	0.0051
NAP1L1	mRNAlike lncRNA nucleosome assembly protein 1-like 1	-2.26	-5.02	0.0023	0.0008
CR608583	mRNAlike lncRNA	-1.94	-1.67	0.0106	0.0375
TC14001853.hg.1*	lncRNA	-1.9	-2.36	0.0002	0.0084
KDSR	mRNAlike lncRNA 3-ketodihydrosphingosine reductase	-1.75	-1.57	0.0338	0.0197
HIST2H2BE	histone cluster 2, H2be	-1.74	-1.57	0.0218	0.0020
TC02004518.hg.1*	lncRNA	-1.65	-1.85	0.0067	0.0080
LINC00657	long intergenic non-protein coding RNA 657	-1.59	-1.8	0.0225	0.0326
AK057721	mRNAlike lncRNA	-1.53	-1.56	0.0129	0.0125
DQ577389	piRNA	-1.53	-2.03	0.0081	0.0266

* In case of neither gene symbol nor accession number, the transcript cluster ID has been provided.

Appendix 7. Common up-regulated genes in MFAP1- and THOC1-depleted cells.

Gene Symbol (or accession number)	Description	Fold Change (Linear)		ANOVA P-value	
		siMFAP1	siTHOC1	siMFAP1	siTHOC1
RNU5E-1; RNU5D-1	RNA, U5E small nuclear 1; RNA, U5D small nuclear 1	1.96	1.54	0.0028	0.0195
HMOX1	heme oxygenase (decycling) 1	2.08	2.6	0.0257	0.0119
RNU6-166P	RNA, U6 small nuclear 166, pseudogene	2.12	1.53	0.0417	0.0335
BC014579	mRNAlike lncRNA	2.18	1.69	0.0389	0.0413
RNU6-1309P	RNA, U6 small nuclear 1309, pseudogene	2.28	1.52	0.0013	0.0048
HMOX1	mRNAlike lncRNA heme oxygenase (decycling) 1	2.32	2.87	0.0209	0.0409
SNORD3C	small nucleolar RNA, C/D box 3C	2.72	1.51	0.0035	0.0184
SNORD82	small nucleolar RNA, C/D box 82	2.73	1.51	0.0190	0.0158
SNORD3D	small nucleolar RNA, C/D box 3D	3.16	1.53	0.0031	0.0189
TC11001451.hg.1*	ncrna:snRNA	3.32	1.62	0.0001	0.0024
RNU6-30P	RNA, U6 small nuclear 30, pseudogene	3.64	1.61	0.0005	0.0228
SNORD3B-1	small nucleolar RNA, C/D box 3B-1	3.65	1.76	0.0009	0.0039
SNORD3B-1; SNORD3B-2	small nucleolar RNA, C/D box 3B-1; small nucleolar RNA, C/D box 3B-2	3.65	1.76	0.0009	0.0039
RNU6-116P	RNA, U6 small nuclear 116, pseudogene	4.03	1.67	0.0000	0.0003
GDF15	growth differentiation factor 15	5.15	1.6	0.0014	0.0364

* In case of neither gene symbol nor accession number, the transcript cluster ID has been provided.

Appendix 8. Differentially spliced genes in THOC1-depleted cells.

http://bit.ly/2kS058i

Appendix 9. GO analysis of differentially spliced genes in THOC1-depleted cells.

http://bit.ly/2ml3zAv

Appendix 10. Common differentially spliced genes in MFAP1- and THOC1-depleted cells.

http://bit.ly/2mESuLq

Appendix 11. GO analysis of common differentially spliced genes in MFAP1- and THOC1-depleted cells.

http://bit.ly/2lgDLJF

Appendix 12. AS changes of the *CDC25* phosphatases and *CDCA5* genes in siMFAP1-depleted cells.

	Gene expression		Alternative splicing	
	Fold Change (Linear)	P-value	Splicing Index (Linear)	P-value
CDC25A	-1.09	0.08	2.14	0.016
CDC25B	1	0.37	2.03	0.007
CDC25C	-1.01	0.3	2.18	0.018
CDCA5	1.09	0.04	2.31	0.006
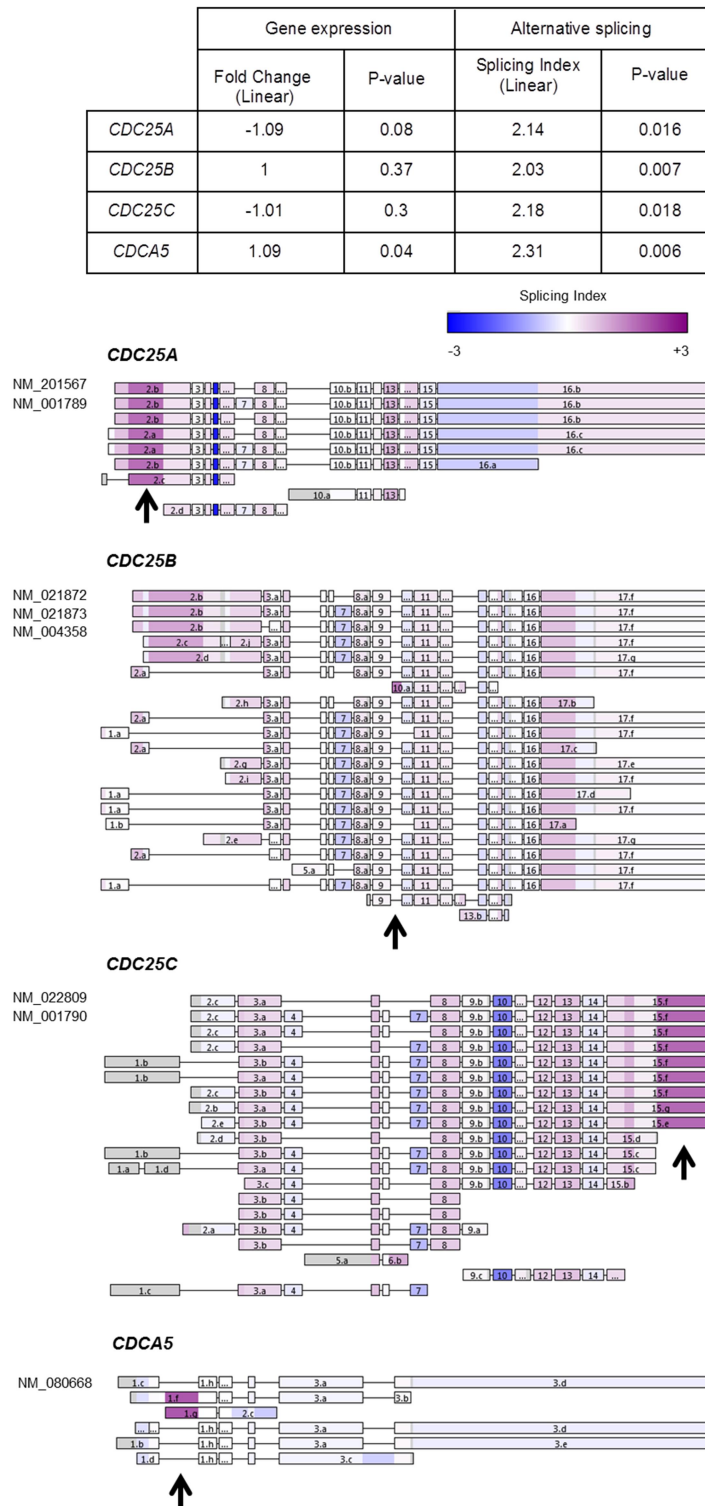
Table showing the gene expression (Fold change) and alternative splicing (Splicing Index) data in the *CDC25A*, *CDC25B*, *CDC25C* and *CDCA5* genes in MFAP1-depleted cells (upper panel). Schematic representation of *CDC25A*, *CDC25B*, *CDC25C* and *CDCA5* splice variants (lower panel). Microarray probe sets with negative splicing indexes (lower inclusion) are depicted in blue, and probe sets with positive splicing indexes (higher inclusion) are depicted in purple. Bold arrows indicate the probe sets that showed significant AS changes in MFAP1-depleted cells (P <0.01; |Splicing Index| >2). Accession numbers are provided on the left.

168

MATERIALS AND METHODS

1. GROWTH MEDIA AND CONDITIONS

1.1. Bacteria cell culture

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

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

1.2. *Saccharomyces cerevisiae* cell culture

Yeast cells were cultured at 30 °C and experiment were performed in SC or YPAD medium unless otherwise noted. Liquid cultures were incubated on horizontal orbital shakers at 200 rpm. Sporulation of diploids cells was induced in SPO medium during 3-4 days.

The following culture media were used for yeast:

- *YPAD* (rich medium): 1% yeast extract, 2% bacto-peptone, 2% glucose, 20 mg/ml adenine.
- SD (Synthetic Dextrose), minimal medium: 0.17% yeast nitrogen base (YNB) without amino acids, 0.5% ammonium sulfate, 2% glucose.
- *SC* (Synthetic Complete) medium: SD medium supplemented with adenine, aspartic acid, arginine, histidine, leucine, lysine, methionine, tryptophan and uracil at concentrations described in (Sherman et al. 1986). When some of the components are absent it is specified. For example SC-Trp means SC without tryptophan.
- *SC/+X-α-Gal/+AbA*: SC medium supplemented with 40 µg/ml X-α-Gal and 125 ng/ml aureobasidin (AbA). X-α-Gal and AbA were added after autoclaving.
- *SPO* (Sporulation) medium: 1% Potassium Acetate, 0.1% yeast extract and 0.005% glucose. Supplements were added at the levels of 25% of those used for SC medium.

Solid media were prepared by adding 2% agar before autoclaving.

1.3. Human cell culture

HeLa and HEK293T cells were cultured in DMEM (Gibco, USA) supplemented with 10% heat inactivated fetal bovine serum (FBS), 2 mM L-glutamine, 100 µg/ml streptomycin, 60 µg/ml penicillin and 0.25 µg/ml amphotericin B. Cells were maintained at 37 °C 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 *Escherichia coli*. Used for plasmid selection in *E. coli*. (Use: 100 µg/ml).
- *Kanamycin*, Kan (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 interferes with fungal membrane permeability (Use: 0.25 µg/ml). Used to prevent growth of bacteria, yeast and fungi in human cell culture.

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.
- *Chymostatin* (SIGMA): inhibitor of many proteases, including chymotrypsins and cathepsins.
- *Phenylmethanesulfonyl fluride*, (PMSF) (SIGMA): Inhibitor of serine and cysteine proteases. (Use: 1 mM).

- *Phosphatase inhibitor cocktail 2* (SIGMA): a mixture of inhibitors of acid and alkaline phosphatase as well as tyrosine protein phosphatases. Contains sodium vanadate, sodium molybdate, sodium tartrate, and imidazole. (Use: 1/100).

- *Dethyl pirocarbonate* (DEPC) (SIGMA): inhibitor of RNAses.

- *Thymidine* (SIGMA): is a pyrimidine deoxynucleoside, an inhibitor of DNA synthesis. It is used to synchronize the cells in G1/early S phase. (Use: 2 mM).

- *Nocodazole* (SIGMA): an inhibitor of microtubule polymerization. It is used to synchronize the cells in G2/M phase. (Use: 50 ng/ml).

- *KaryoMAX colcemid solution* (Invitrogen): N-desacetyl-N-methylocolchicine (Colcemid) prevents spindle formation during mitosis, causing metaphase arrest. (Use: 0.1 μg/ml)

- *Cordycepin* (SIGMA): Adenosine antagonist 3' deoxyadenosine, inhibitor of RNA chain elongation. (Use: 50 μM).

- *5,6-dichloro-1-β-D-ribofurosylbenzimidazole* (DRB) (SIGMA): Adenosine analogue, inhibitor of RNA polymerase II elongation. (Use: 100 μM).

- *Trichostatin A* (TSA) (SIGMA): a specific inhibitor of histone deacetylase (HDAC) from *Streptomyces hygroscopicus*. (Use: 100, 250 nM).

- *Suberoylanilide hydroxamic acid* (SAHA) (SIGMA): A potent, reversible pan-histone deacetylase (HDAC) inhibitor. It inhibits both class I and class II HDACs. (Use: 5, 7.5 μM).

- *Anacardic acid* (AA) (Millipore): An inhibitor of histone acetyltransferase (HAT) activity. (Use: 30 μM).

- 5-fluoroorotic acid (FOA) (USB): Toxic analog of the uracil that kills URA3 cells but not ura3 mutants (Boeke et al. 1984). (Use: 500 mg/l).

2.3. Enzymes and antibodies

- *Alkaline phosphatase* (Roche): hydrolyzes 5'-monophosphate groups from DNA impeding the religation of double-stranded DNA (dsDNA) ends after an enzymatic cut.

- *T4 DNA ligase* (Roche): catalyzes the formation of phosphodi-ester bonds between neighboring 3´-hydroxyl- and 5´-phosphate ends in dsDNA.

- *DNA polymerase Expand High Fidelity PCR system* (Roche): mixture of Taq and Pwo polymerases. Used in PCRs that required high fidelity.

- *Phusion High-Fidelity DNA polymerase* (Finnzymes): *Pyrococcus*-like DNA polymerase fused with a processivity-enhancing domain. Used in PCRs for long or difficult amplicons that required high fidelity and processivity.

- *Go-Taq Flexi DNA polymerase* (Promega): used for PCRs to generate probes and check.

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

- *Proteinase K* (Roche): very efficient serine protease from *Pichia pastoris* with no pronounced cleavage specificity.

- *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 the Table M1 and Table M2 below.

Table M1. Primary antibodies used in this study.

Antibody	Source	Epitope	Reference	Use
c-Myc (Clone 9E10)	Mouse	Synthetic peptide corresponding to residues 408-439 of the human p62-c-Myc protein	631206 (Clontech)	WB (1:500) TBS-0.2% T 5% milk
GFP	Mouse	A mixture of two monoclonal antibody (7.1 and 13.1 clones) that recognizes both wild type and mutant forms of GFP	11814460001 (Roche)	WB (1:1000) TBS-T 5% milk or BB IP (2 μg)

Antibody	Source	Epitope	Reference	Use
GFP	Rabbit	Recombinant full length GFP made in *Escherichia coli*	ab6556 (Abcam)	PLA (1:250)
β-Actin	Rabbit	Synthetic peptide derived from within residues 1-100 of Human beta Actin	ab8227 (Abcam)	WB (1:1000) TBS-T 5% milk
Nuclear Matrix Protein p84 [E10] antibody (THOC1)	Mouse	Fusion protein containing amino acids 15-374 of human p84	ab487 (Abcam)	WB (1:2500) BB or odyssey buffer IF (1:500) PLA (1:250) ChIP (3 μg)
MFAP1	Rabbit	Synthetic peptide within Human MFAP1 amino acids 28-58 (N terminal)	ab175508 (Abcam)	WB (1:500) TBS-T 5% milk IF (1:250) ChIP (15 μl)
SC35 (phospho) Nuclear speckle marker	Mouse	Antibody that recognizes a phospho-epitope of the non-snRNP (small nuclear ribonucleoprotein particles) factor SC35	ab11826 (Abcam)	IF (1:500)
SAP130	Rabbit	Recombinant protein fragment containing a sequence corresponding to a region within amino acids 771 and 1005 of Human SAP130	ab111739 (Abcam)	PLA (1:50)
SAP130	Rabbit	Synthetic peptide corresponding to a region between amino acids 998-1048 of Human SAP130	ab114978 (Abcam)	WB (1:1000) odyssey buffer ChIP (5 μg)
SMC2	Rabbit	Synthetic peptide, which represented a portion of Human SMC2 within exon 27	ab10412 (Abcam)	WB (1:2500) BB IP (5 μg)
mSin3A	Rabbit	Synthetic peptide corresponding to amino acids 1-19 of Mouse mSin3A	ab3479	WB (1:2000) BB or odyssey buffer PLA (1:250) ChIP (5 μg)
mSin3A (G-11)	Mouse	Epitope mapping within the PA$_{H2}$ region of mSin3A of mouse origin	sc-5299 (Santa Cruz Biotechnology)	IF (1:50) PLA (1:50)
Phospho-Histone H2A.X (Ser139), clone JBW301	Mouse	Synthetic peptide corresponding to amino acids 134-142 of human histone H2A.X	05-636 (Millipore)	IF (1:500)
53BP1	Rabbit	Amino acids 350 and 400 of Human 53BP1	NB100-304 (Novus Biologicals)	IF (1:500)

175

Antibody	Source	Epitope	Reference	Use
S9.6	Mouse	Antibody that detects DNA-RNA hybrids	Hybridoma cell line HB-8730	DRIP (10 μg) IF (1:200)
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)
UAP56	Rabbit	Synthetic peptide conjugated to KLH derived from within residues 300-400 of Human UAP56	ab47955 (Abcam)	IF (1:200) PLA (1:200)
ALY	Rabbit	Synthetic peptide within Human Aly amino acids 200 to the C-terminus	ab202894 (Abcam)	IF (1:250) PLA (1:250)
THOC2	Rabbit	Synthetic peptide conjugated to KLH derived from within residues 100-200 of Human THOC2	ab46685 (Abcam)	IF (1:100) PLA (1:100)
THOC5 [EP6904]	Rabbit	Synthetic peptide corresponding to amino acids in Human THOC5	ab137051 (Abcam)	IF (1:100) PLA (1:100)
Pol II (H-224)	Rabbit	Antibody raised against amino acids 1-224 of Pol II of Human origin	sc-9001 (Santa Cruz Biotechnology)	ChIP (2 μg)
FIP1L1	Rabbit	Synthetic peptide corresponding to a region between residue 544 and 594 of Human FIP1L1	ab80272 (Abcam)	ChIP (3 μg)
CPSF6	Rabbit	Synthetic peptide corresponding to a region between residue 1 and 50 of Human CPSF6	ab99347 (Abcam)	ChIP (2 μg)
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)
Dimethyl-Histone H3 (Lys9) (H3K9me2)	Rabbit	KLH-conjugated synthetic peptide corresponding to amino acids 6-13 of Histone H3 with C added to the C terminus for conjugation	07-212 (Millipore)	IF (1:200)

Antibody	Source	Epitope	Reference	Use
Histone H3 (di methyl K9) (H3K9me2)	Mouse	Synthetic peptide corresponding to Human Histone H3 amino acids 1-100 (di methyl K9) conjugated to KLH (Cysteine residue)	ab1220 (Abcam)	ChIP (5 μg)
Histone H3 (acetyl K9 + K14 + K18 + K23 + K27) (AcH3)	Rabbit	Synthetic peptide including acetyl-lysines contained in the N-terminal tail of Human Histone H3	ab47915 (Abcam)	WB (1:1000) TBS-T 5% milk or odyssey buffer ChIP (10 μl)
Histone H3	Rabbit	Synthetic peptide corresponding to Human Histone H3 amino acids 100 to the C-terminus conjugated to KLH	ab1791 (Abcam)	WB (1:1000) TBS-T 5% milk or odyssey buffer ChIP (5 μg)
Histone H3	Goat	Synthetic peptide corresponding to Human Histone H3 amino acids 100 to the C-terminus	ab12079 (Abcam)	WB (1:1000) odyssey buffer
Acetyl-Histone H4 (AcH4)	Rabbit	KLH-conjugated peptide corresponding to amino acids 2-19 of Tetrahymena histone H4, acetylated on lysines 5, 8, 12 and 16	06-598 (Millipore)	WB (1:1000) TBS-T 5% milk or odyssey buffer ChIP (5 μg)
Histone H4	Rabbit	Synthetic peptide corresponding to Human Histone H4 amino acids 1-100 conjugated to KLH	ab7311 (Abcam)	WB (1:1000) TBS-T 5% milk or odyssey buffer ChIP (5 μg)
Histone H4	Mouse	Synthetic peptide corresponding to Human Histone H4 aminoacids 50 to the C terminal conjugated to KLH	ab31830 (Abcam)	WB (1:1000) TBS-T 5% milk or odyssey buffer

KLH: Keyhole Limpet Haemocyanin; WB: Western Blotting; IF: Immunofluorescence; IP: Immunoprecipitation; ChIP: Chromatin immunoprecipitation; DRIP: DNA-RNA 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	Source	Use
Rabbit	Horseradish peroxidase	SIGMA (A6154)	WB (1:4000)
Mouse	Horseradish peroxidase	SIGMA (A4416)	WB (1:5000)
Goat	Horseradish peroxidase	Santa Cruz Biotechnology	WB (1:10000)

Specificity	Conjugation	Source	Use
Rabbit	IRDye 800CW	LI-COR (925-32211)	WB Odyssey (1:3000)
Mouse	IRDye 680RD	LI-COR (925-68074)	WB Odyssey (1:3000)
Goat	IRDye 680RD	LI-COR (925-68070)	WB Odyssey (1:3000)
Rabbit	PLA probe (PLUS oligonucleotide)	Olink Biosciences	PLA (1:5)
Mouse	PLA probe (MINUS oligonucleotide)	Olink Biosciences	PLA (1:5)
Rabbit	Alexa fluor 488	Molecular Probes	IF (1:500)
Rabbit	Alexa fluor 568	Molecular Probes	IF (1:500)
Rabbit	Alexa fluor 647	Molecular Probes	IF (1:500)
Mouse	Alexa fluor 488	Molecular Probes	IF (1:500)
Mouse	Alexa fluor 546	Molecular Probes	IF (1:500)
Mouse	Alexa fluor 594	Molecular Probes	IF (1:500)
Mouse	Alexa fluor 647	Molecular Probes	IF (1:500)

WB: Western Blotting; IF: Immunofluorescence; IP: Immunoprecipitation; PLA: Proximity Ligation Assay;

3. BACTERIA, YEAST STRAINS AND HUMAN CELL LINES

3.1. *Escherichia coli* strains

The DH5α strain (*F- endA1 gyr96 hsdR17 ΔlacU169(f80lacZΔM15) recA1 relA1 supE44 thi-1*) (Hanahan 1983) was used in all experiments with *E. coli*.

The following bacterial clones containing the cDNA of the listed genes were used.

Table M3. Bacterial clones used in this study.

Clone name	Gene	Vector	Resistance	Source
IRAUp969H0452D	Human *THOC1*	pDNR-LIB	Chloramphenicol	RZPD/ImaGenes
IRAUp969H063D	Human *UAP56*	pOTB7	Chloramphenicol	RZPD/ImaGenes
IRAUp969G051D	Human *URH49*	pOTB7	Chloramphenicol	RZPD/ImaGenes
IRAUp969B08104D	Human *THOC4* (*ALY*)	pOTB7	Chloramphenicol	RZPD/ImaGenes
IRCMp5012D0113D	Human *SAP130*	pCR4-TOPO	Kanamycin	RZPD/ImaGenes
IRATp970F0878D	Human *MFAP1*	pCMV-SPORT6	Ampicillin	RZPD/ImaGenes
IRATp970E01101D	Human *REV1*	pBluescriptR	Ampicillin	RZPD/ImaGenes
IRAUp969A1279D	Human *CCT5*	pOTB7	Chloramphenicol	RZPD/ImaGenes
IRATp970G0420D	Human *CDC23*	pCMV-SPORT6	Ampicillin	RZPD/ImaGenes

Clone name	Gene	Vector	Resistance	Source
IRCMp5012B0834D	Human *CUL1*	pCR-BluntII-TOPO	Kanamycin	RZPD/ImaGenes
LLAM13353-P01	Human *NUB1*	pCMV-SPORT6	Ampicillin	RZPD/ImaGenes
IRATp970H0132D	Human *SPATA22*	pBluescriptR	Ampicillin	RZPD/ImaGenes
IRAUp969G05108D	Human *CIP29*	pDNR-LIB	Chloramphenicol	RZPD/ImaGenes
IRATp970H0577D	Human *ZBTB44*	pCMV-SPORT6	Ampicillin	RZPD/ImaGenes

RZPD: RZPD German resource center for genome research.

3.2. *Saccharomyces cerevisiae* strains

Yeast strains used are shown in Table M4. Single and double mutants were obtained by backcrosses and tetrad dissection using a SINGER MSM 200 micromanipulator.

Table M4. S. cerevisiae strains used in this study.

Name	Genotype	Source
Y2HGold	*MAT**a**, trp1-901, leu2-3, 112,ura3-52, his3-200, gal4Δ, gal80Δ, LYS2:: GAL1$_{UAS}$-Gal1$_{TATA}$-His3, GAL2$_{UAS}$-Gal2$_{TATA}$-Ade2 URA3::MEL1$_{UAS}$-Mel1$_{TATA}$ AUR1-C MEL1*	Derived from P J69-2A strain (James et al. 1996)
Y187	*MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, gal80Δ, met-, URA3::GAL1$_{UAS}$-Gal1$_{TATA}$-LacZ, MEL1*	(Harper et al. 1993)
BY4741	*MAT**a** his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	Euroscarf
YOL004W	BY4741 *sin3Δ::KanMX4*	Euroscarf
YNL330C	BY4741 *rpd3Δ::KanMX4*	Euroscarf
YPL139C	BY4741 *ume1Δ::KanMX4*	Euroscarf
YMR075W	BY4741 *rco1Δ::KanMX4*	Euroscarf
YPR023C	BY4741 *eaf3Δ::KanMX4*	Euroscarf
YIL084C	BY4741 *sds3Δ::KanMX4*	Euroscarf
YMR263W	BY4741 *sap30Δ::KanMX4*	Euroscarf
YNL097C	BY4741 *pho23Δ::KanMX4*	Euroscarf
A3Y3T3	*MATα ade2 ura3 his3Δ200 hprlΔ3::HIS3 leu2-k::ADE2-URA3::leu2-k*	(Aguilera and Klein 1990)
LKWT-23B	*MATα ade2 ura3 his3 met15Δ0 leu2-k::ADE2-URA3::leu2-k*	JM. Santos-Pereira
LKWT-18D	*MAT**a** ade2 ura3 his3 leu2-k::ADE2-URA3::leu2-k*	JM. Santos-Pereira
LKH1-32A	*MAT**a** ade2 ura3 his3 hprlΔ3::HIS3 leu2-k::ADE2-URA3::leu2-k*	JM. Santos-Pereira
LKH1-10D	*MATα ade2 ura3 his3 met15Δ0 hprlΔ3::HIS3 leu2-k::ADE2-URA3::leu2-k*	JM. Santos-Pereira
LKH1-1B	*MATα ade2 ura3 his3 met15Δ0 hprlΔ3::HIS3 leu2-k::ADE2-URA3::leu2-k*	JM. Santos-Pereira

Name	Genotype	Source
LKSIN3-1B	*MATα ade2 ura3 his3 sin3Δ::Kan leu2-k::ADE2-URA3::leu2-k*	JM. Santos-Pereira
LKH1SIN3-4B	*MAT**a** ade2 ura3 his3 hprlΔ3::HIS3 sin3Δ::Kan leu2-k::ADE2-URA3::leu2-k*	JM. Santos-Pereira
LKH1SIN3-23C	*MAT**a** ade2 ura3 his3 met15Δ0 hprlΔ3::HIS3 sin3Δ::Kan leu2-k::ADE2-URA3::leu2-k*	JM. Santos-Pereira
LKRPD3-1C	*MAT**a** ade2 ura3 his3 met15Δ0 rpd3Δ::Kan leu2-k::ADE2-URA3::leu2-k*	JM. Santos-Pereira
LKH1RPD3-18A	*MATα ade2 ura3 his3 met15Δ0 hprlΔ3::HIS3 rpd3Δ::Kan leu2-k::ADE2-URA3::leu2-k*	JM. Santos-Pereira
CEN.PK2	*MAT**a** his3Δ1 leu2-3,112 ura3-52 trp1-289 MAL2-8c SUC2*	Euroscarf
CEN.RO20-3A (YBR152W)	CEN.PK *MATα his3Δ1 leu2-3,112 ura3-52 trp1-289 spp381Δ::LEU2*	Euroscarf
YPH2622	*MAT**a** can1Δ::MFA1pr-HIS3::LEU2 his3Δ1 met15Δ0 spp381-ts::URA3*	(Ben-Aroya et al. 2008)
YPH2622-2D	*MATα his3Δ1 leu2Δ0 met15Δ0 spp381-ts::URA3*	This study
YDR007W	BY4741 *trp1Δ::KanMX4*	Euroscarf
YPH2622-2D-4B	*MATα his3Δ1 leu2Δ0 trp1Δ::KanMX4 spp381-ts::URA3*	This study

UAS: Upstream Activation Sequence; TATA: TATA box.

3.3. Human cells

Human cells used in this study are listed in the Table M5.

Table M5. Human cell lines used in this study.

Cell line	Description	Medium	Source
HeLa	Human cervical adenocarcinoma epitelial cells	DMEM	ECACC
HEK293T	Human Embryonic Kidney 293 cells. It contains the SV40 T-antigen.	DMEM	Dr. Amelia Nieto CNB-CSIC, Madrid (España)

ECACC: European Collection of Authenticated Cell Cultures.

All experiments were performed using HeLa cells except co-immunoprecipitation experiments, which were performed using HEK293T cells.

4. PLASMIDS

Plasmids used are shown in Table M6.

Table M6. Plasmids used in this study.

Plasmids used in *S. cerevisiae*			
Plasmid	**Description**	**Resistance**	**Reference/Source**
pGAL4BD	Yeast two-hybrid pGBKT7 bait vector for expressing proteins fused to the GAL4 DNA-binding domain	Kanamycin	Clontech
pGAL4BD-53	Yeast two-hybrid pGBKT7-53 positive control plasmid in combination with pGADT7-T plasmid. 53=murine p53.	Kanamycin	Clontech
pGAL4BD-Lam	Yeast two-hybrid pGBKT7-Lam negative control plasmid in combination with pGADT7-T plasmid. Lam=Laminin C.	Kanamycin	Clontech
pGAL4BD-THOC1	pGBKT7 DNA-BD vector containing the Human THOC1 cDNA (cloned into the *NcoI-BamHI* sites of the MCS) fused to the GAL4 DNA-binding domain	Kanamycin	This study
pGAL4BD-UAP56	pGBKT7 DNA-BD vector containing the Human UAP56 cDNA (cloned into the *NcoI-BamHI* sites of the MCS) fused to the GAL4 DNA-binding domain	Kanamycin	This study
pGALBD-URH49	pGBKT7 DNA-BD vector containing the Human URH49 cDNA (cloned into the *EcoRI-SalI* sites of the MCS) fused to the GAL4 DNA-binding domain	Kanamycin	This study
pGAL4BD-ALY	pGBKT7 DNA-BD vector containing the Human ALY cDNA (cloned into the *NcoI-BamHI* sites of the MCS) fused to the GAL4 DNA-binding domain	Kanamycin	This study
pGAL4AD	Yeast two-hybrid pGADT7-RecAB prey vector for expressing proteins fused to the GAL4-AD (activation domain)	Ampicillin	Clontech
pGAL4AD-T	Yeast two-hybrid pGADT7-T control plasmid in combination with pGBKT7-53 or pGBKT7-Lam plasmids. T=SV40 T-antigen.	Ampicillin	Clontech

Plasmids used in *S. cerevisiae*			
Plasmid	**Description**	**Resistance**	**Reference/Source**
pWJ1213	YCp *HIS* containing the *RAD52::YFP* fusion	Ampicillin	(Lisby et al. 2001)
pWJ1344	YCp *LEU* containing the *RAD52::YFP* fusion	Ampicillin	(Lisby et al. 2001)
pRS316	YCp vector based on the *URA3* marker	Ampicillin	(Sikorski and Hieter 1989)
pRS316L	pRS316 containing two direct repeats of *LEU2* gene sharing 600 bp of homology	Ampicillin	(Prado and Aguilera 1995)
pRS316LYΔNS	pRS316L containing the complete YIp5 sequence, except a 1.92-kb *Sph*I-*Nsi*I deletion, inserted at the *Bgl*II site located between the repeats	Ampicillin	(Prado et al. 1997)
pRS316GL-*lacZ*	p316GLB with the 3-kb *Bam*HI fragment of *LacZ* from pPZ inserted at *Bgl*II between the *leu2* repeats	Ampicillin	J. Lafuente
pRS316L-*lacZ*	pRS316LB with the 3-kb *Bam*HI fragment of *LacZ* from pPZ inserted in the *Bgl*II site located between the repeats	Ampicillin	J. Lafuente
pRS314	YCp vector based on the *TRP* marker	Ampicillin	(Sikorski and Hieter 1989)
pRS314L	pRS314 containing two direct repeats of *LEU2* gene sharing 600 bp of homology	Ampicillin	(Prado and Aguilera 1995)
pRS314LYΔNS	pRS314L containing the complete YIp5 sequence, except a 1.92-kb *Sph*I-*Nsi*I deletion, inserted at the *Bgl*II site located between the repeats	Ampicillin	(Prado et al. 1997)
pRS314GL-*lacZ*	p314GLB with the 3-kb *Bam*HI fragment of *LacZ* from pPZ inserted at *Bgl*II between the *leu2* repeats	Ampicillin	(Piruat and Aguilera 1998)
pSCh204(pL-*lacZ*)	pRS314LB with the 3-kb *Bam*HI fragment of *LacZ* from pPZ inserted in the *Bgl*II site located between the repeats	Ampicillin	(Chavez and Aguilera 1997)
pCM189	Centromeric plasmid containing *Tet* promoter and *URA* marker	Ampicillin	(Gari et al. 1997)
pCM184	Centromeric plasmid containing *Tet* promoter and *TRP* marker	Ampicillin	(Gari et al. 1997)
pCM189-RNH1	*RNH1* ORF cloned into pCM189 in *Not*I site	Ampicillin	(Castellano-Pozo et al. 2013)
pCM184-RNH1	*RNH1* ORF cloned into pCM184 in *Not*I site	Ampicillin	(Santos-Pereira et al. 2013)

Plasmids used in mammalian cell cultures			
Plasmid	**Description**	**Resistance**	**Reference/Source**
pEGFP-C2	Mammalian expression vector that allows transient expression of N-terminal EGFP fusion proteins under the control of the P_{CMV}	Kanamycin	Clontech
pEGFP-SAP130	pEGFP-C2 containing the Human SAP130 cDNA (cloned into the *Bgl*II-*Kpn*I sites of the MCS) fused to EGFP	Kanamycin	This study
pEGFP-MFAP1	pEGFP-C2 containing the Human MFAP1 cDNA (cloned into the *Bgl*II-*Kpn*I sites of the MCS) fused to EGFP	Kanamycin	This study
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
pCDNA3	Vector containing a P_{CMV} for expression in mammalian cells	Ampicillin	(ten Asbroek et al. 2002)
pCDNA3-RNaseH1	pcDNA3 containing the human RNase H1 gene under the P_{CMV}	Ampicillin	(ten Asbroek et al. 2002)

MCS: Multiple cloning site; P_{CMV}: Cytomegalovirus promoter; EGFP: Enhanced Green Fluorescent Protein; ORF: Open reading frame.

The different Human cDNA sequences cloned into pEGFP-C2 plasmid were obtained by PCR from the corresponding clone (see Table M3).

5. BACTERIAL AND YEAST 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. Yeast transformation

Yeast transformation was performed using the lithium acetate method as previously described (Gietz et al. 1995).

5.3. Human cells transfection

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

5.3.1. siRNA transfection

siRNA used are shown in Table M7.

Table M7. siRNAs used in this study.

	siRNA	Time	Source
siC	ON-TARGETplus Non-targeting Pool (D-001810)	48, 72 h	Dharmacon
siMFAP1	ON-TARGETplus SMARTpool human MFAP1 (L-020071-02)	48, 72 h	Dharmacon
siSAP130	ON-TARGETplus SMARTpool human SAP130 (L-017129-02)	48, 72 h	Dharmacon
siSIN3	ON-TARGETplus SMARTpool human SIN3A (L-012990-00)	48, 72 h	Dharmacon
siSAP30	ON-TARGETplus SMARTpool human SAP30 (L-011538-00)	48, 72 h	Dharmacon
siSUDS3	ON-TARGETplus SMARTpool human SUDS3 (L-027133-02)	48, 72 h	Dharmacon
siTHOC1	ON-TARGETplus SMARTpool human THOC1 (L-019911-00)	48, 72 h	Dharmacon

Cells were transfected with siRNA using DharmaFECT 1 (Dharmacon) at 30-50% confluence. 24 hours before transfection cells were cultured in antibiotic-free medium. Transfection in a well of a 6-well plate was performed using the following protocol:

- *Mixture A* (final volume 100 µl): 50 µl culture serum-free medium (medium without antibiotics or FBS), 36 µl H2O, 9 µl 5X siRNA buffer (300 mM KCl, 30 mM HEPES-pH 7.5, 1.0 mM MgCl2) (Dharmacon) and 5 µl siRNA 20 µM (100 nM).

- *Mixture B* (final volume 100 µl): 95 µl serum-free medium and 5 µl DharmaFECT 1.

Each mixture was incubated at room temperature (RT) for 5 min, mixed and incubated for 20 min. Meanwhile, medium was replaced by 800 µl serum-free medium. For transfection of primary cells, this medium was supplemented with the normal concentration of FBS. Transfection solution was added carefully drop by drop to the cell culture and was incubated for 4-5 hours. Afterwards, 2 ml of a high-concentrate complete medium was added to have a final normal concentration of all the medium components.

Co-transfection with two siRNA for double depletions was performed using 3.5 µl of each siRNA in the mixture A.

5.3.2. Plasmid transfection using Lipofectamine 2000 or FuGENE 6

- For plasmid transfection using Lipofectamine 2000 (Invitrogen), cells were transfected at 80% confluence. 24 hours before transfection cells 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): 2-3 µg DNA in Opti-MEM (Gibco).
- *Mixture B* (final volume 100 µl): 4-6 µl Lipofectamine 2000 in Opti-MEM.
  Each mixture was incubated at RT for 5 min, mixed and incubated for 15 min at RT. Transfection solution was added carefully drop by drop to the cell culture.

- For plasmid transfection using FuGENE 6 (Roche), cells were transfected at 80% confluence. 24 h before transfection cells were cultured in antibiotic-free medium (3 ml for 6-well plates). Transfection in a well of a 6-well plate was performed using the following protocol:
- *Mixture* (final volume 150 µl): 9 µl FuGENE in Opti-MEM was incubated for 5 min at RT. Then, 3 µg of DNA was added and the mixture was incubated for 20 at RT. Transfection solution was finally added drop by drop to the cell culture.

6. PROTEIN-PROTEIN INTERACTION METHODS

6.1. Yeast two-hybrid system

The screening to identify new protein partners of THO/TREX was performed using the GAL4-based Matchmaker Gold Yeast Two-Hybrid System (Clontech). In this system, a bait protein is expressed as a fusion to the Gal4 DNA-binding domain (DNA-BD), while libraries of prey proteins are expressed as fusions to the Gal4 activation domain (AD) (Fields and Song 1989; Chien et al. 1991). When bait and library (prey) fusion proteins interact, the DNA-BD and AD are brought into proximity to activate transcription of four independent reporter genes (*AUR1-C, ADE2, HIS3* and *MEL1*). *AUR1-C* is a dominant version of the *AUR1* gene that encodes the enzyme inositol phosphoryl ceramide synthase. Its expression confers strong resistance to the otherwise highly toxic drug Aureobasidin A (AbA). The expression of *HIS3* permits the cell biosynthesizes histidine and therefore grow on SC–His minimal medium. The expression of *ADE2* allows cells to grow on SC–Ade minimal medium. *MEL1* encodes an α-galactosidase that when expressed is secreted by the yeast cells and allows yeast colonies turn blue in the presence of the chromogenic substrate X-α-Gal. These four reporter genes are under the control of three distinct Gal4-responsive promoters (G1, G2 and M1) that are different because they have different Gal4-binding sites thus reducing the number of false positives (Giniger et al. 1985; Giniger and Ptashne 1988).

As bait, we choose THOC1, UAP56, URH49 and ALY as representative subunits of THO/TREX. The full length cDNA of each one was obtained by PCR using primers detailed in Table M8 from IMAGE clones (see Table M3) and cloned into pGBKT7 vector fused to the Gal4 DNA-BD (pGAL4BD-*BAIT*). This plasmid contains the *TRP1* gene as selection marker in yeast. As prey, Mate & Plate library-Universal Human (normalized) (Clontech) was used. This library was constructed from human cDNA, obtained from a mixture of poly A+ RNAs from a collection of adult human (male and female) tissues chosen to represent a broad range of expressed genes. The library had been previously normalized to preferentially remove abundant cDNAs derived from high-copy-number mRNAs. cDNA library is cloned into pGADT7-RecAB vector, which contains the *LEU2* gene as selection marker in yeast and transformed into yeast strain Y187.

186

The screening was performed according to the Matchmaker Gold Yeast Two-Hybrid System manual. The yeast strain Y2HGold was transformed with vector pGAL4BD-*BAIT* and grown in 50 ml SD-Trp at 30 °C at 250 rpm until the OD_{600} reaches 0.8 (16-20 h). The cells were harvested by centrifugation at 1000 g for 5 min and cell density was adjusted to >1 x 10^8 cells per ml in SD-Trp. 5 ml of the bait strain cell culture were mated with one vial (1 ml) of the library strain Y187 cell culture transformed with the cDNA library in 45 ml of 2X YPAD containing 50 µg/ml kanamycin and cultivated at 30 °C for 20-24 h at 50 rpm. When zygotes were formed, cells were harvested by centrifugation at 1000 g for 10 min, the cell pellet was washed with 0.5X YPAD/kan and finally resuspended in 10 min of 0.5X YPAD/kan. The culture was plated on to SD-Trp-Leu+(X-α-Gal)+Aba agar plates and incubated at 30 °C for 3-5 days. The blue colonies appeared were streaked on to high stringency plates SD-Ade-His-Trp-Leu+(X-α-Gal)+Aba and incubated at 30 °C for 3-5 days. Prey plasmids from the blue colonies appeared on high stringency plates were rescued using the Easy Yeast Plasmid Isolation Kit (Clontech) and inserts were sequenced. Afterward, positive interactions between bait and candidates selected were verified by re-transformation of Y2HGold strain and subsequent growth on selective media.

In order to calculate the number of clones screened, the final volume of mated culture was scored and 100 µl of the mated culture diluted 1/100 were plated on SD-Trp-Leu (diploids selection). The number of cfu (colony forming units) was counted and the number of clones screened was calculated:

Nº clones screened= (nº cfu/Plating volume) * Final vol. of mated culture * Dilution factor.

6.2. 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, 4 h at 4 °C. 20 and 180 µL of supernatant were used for input and immunoprecipitation. After two washed in PBS-0.5% BSA, 180 µL of 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 2 h at 4 °C. 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.

6.3. Proximity Ligation Assay (PLA)

For detection of protein-protein interactions *in situ*, PLA was performed with reagents from Duolink In Situ Red Starter Kit (Olink Biosciences) in accordance with manufacturer´s instruction. Schematic drawing of the different PLA steps is shown in Figure M1. Cells were cultured on glass coverslips and fixed in optimal conditions compatible with the two primary antibodies to be used (see Materials and methods 11). Coverslips were blocked with PBS-3% BSA for 1 h at RT, incubated with primary antibodies diluted in PBS-3%BSA for 2 h 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.

<u>For each coverslip:</u>

PLA probes (40 µl): 8 µl PLA probe anti-Mouse MINUS, 8 µl PLA probe anti-Rabbit PLUS, 24 µl PBS-3% BSA.

Ligation reaction (40 µl): 8 µl of 5X Buffer ligase, 1 µl Ligase, 31 µl MQ H_2O.

Amplification reaction (40 µl): 8 µl of 5X Buffer amplification, 0.5 µl Polymerase, 31.5 µl MQ H_2O.

PLA probes, buffers, enzymes and mounting media are provided in the Duolink In Situ Red Starter Kit.
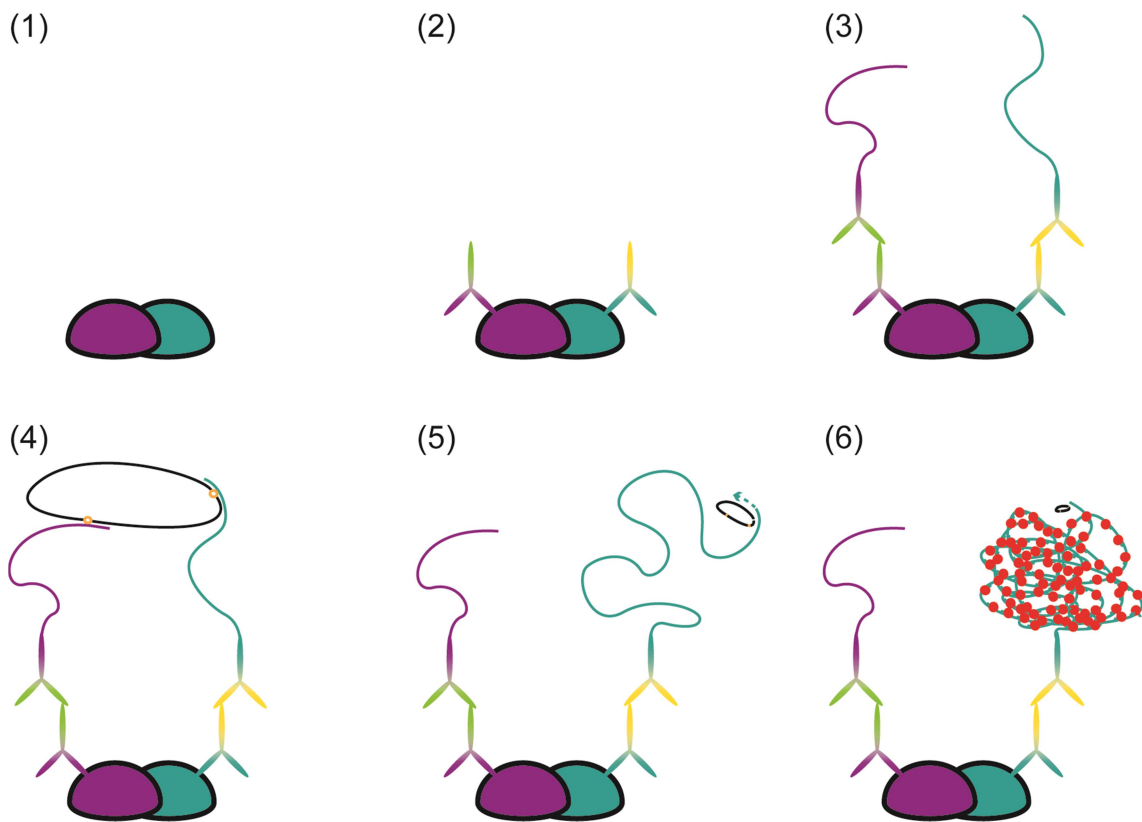
Figure M1. *In situ* proximity ligation assay (PLA).
Schematic drawing of the different PLA steps. **(1)** Two target proteins that interact. **(2)** Incubation with target primary antibodies. **(3)** Incubation with secondary antibodies conjugates with oligonucleotides (PLUS and MINUS PLA probes). **(4)** Ligation reaction to form a complete DNA circle (from the PLA probes oligonucleotides, connector oligos and the ligase). **(5)** Rolling circle amplification reaction (using the ligated DNA circle as template and the polymerase). **(6)** Addition of fluorescent probes that hybridize to the amplification product to reveal interaction.

7. PROLIFERATION ASSAYS IN HUMAN CELLS

Hela cells were plated in 1 well of a 6-well plate and were transfected with siRNA at 30-50% confluence the following day. 48 h after transfection, cells were detached using accutase, diluted in complete medium, counted and plated in wells of 96-well plates (2000 cells/well) in a final volume of 100 µl. The first quantification, which was considered the day 0, was performed after 7 h. Cells were transfected again 3 days after first transfection with 2/5 of siRNA standard amount as described in Materials and methods 5.3.1.

Cell number quantification was performed using the Cell Proliferation reagent WST-1 (4-[3-84-Iodophenyl)-2-(4-nitrophenyl)2H-5-tetrazolio]-1,3-benzene disulfonate) (Roche) following the manufacture´s protocol. 10 µl of WST-1 reagent were added to the culture and incubated for 2 hours at 37 °C and 5% CO_2 in darkness. WST-1 is a tetrazolium salt that can be cleaved to formazan by the mitochondrial dehydrogenases in the sample. Thus, the amount of formazan dye. For that, samples were shaken thoroughly (720 rpm) for 1 minute and the absorbance was measured using microplate (ELISA) reader VARIOSKAN FLASH (Thermo) at 450 nm with a reference wavelength of 690 nm. Absorbance values were normalized to the value at the day 0 and represented as arbitrary units (A.U.). Measurements were performed each 24 hours.

8. ANALYSIS OF SUB-G1 DNA CONTENT (APOPTOSIS)

HeLa cells were transfected with siRNA. 72 h after transfection, floating and adherent siRNA-depleted cells from two well of a 6-well plate ($\approx 5 \times 10^5$ cells) were harvested, washed with 5 ml of cold PBS, centrifuged at 3500 rpm for 2 min and resuspended in 100 µl PBS. For fixation, 900 µl of cold 70% ethanol was added drop by drop during vortexing and cells were incubated at least 5 minutes at 4 °C. Afterwards, cells were centrifuged at 3500 rpm for 2 min, washed with 2 ml PBS, resuspended in 250 µl PBS and 250 µl of DNA extraction solution was added. After incubation for 10 min at 37 °C, cells were centrifuged and pellet was resuspended in 300 µl of PI/RNase A solution and incubated for 30 min in darkness. 5000 cells were scored by FACS and the percentage of cells with sub-G1 content was calculated.

10X PBS: 81.9 g/L NaCl, 2.01 g/L KCl, 14.2 g/L Na_2HPO_4, 2.45 g/L KH_2PO_4.

DNA extraction solution: 190 ml Na_2HPO_4 0.2 M, 8 ml Citric Acid 0.1 M pH 7.8.

PI/RNase A solution: 100 µg/ml RNase A, 5 µg/ml Propidium Iodide in PBS.

9.   CELL CYCLE ANALYSIS IN HUMAN CELLS

9.1.   FACS analysis

For cell cycle analysis, HeLa cells were incubated with EdU (20 µM) for 20 min 72 h after siRNA transfection to allow EdU incorporation in replicating cells. Cells were harvested with accutase, washed with PBS and centrifuged (10000 rpm 10 s). Pellet was resuspended in 300 µl cold PBS and cells were fixed by adding 700 µl cold 96% ethanol drop by drop and incubated at least 1 h on ice. After centrifugation (10000 rpm) ethanol was removed and 800 µl of 0.5% Triton X-100 in PBS was added to the pellet, mixed by vortexing and incubated on ice for 15 min. Then, cells were centrifuged (10000 rpm), washed in 800 µl of 1% BSA, 0.5% Tween-20 in PBS and centrifuged again. Afterwards, pellet was resuspended in 100 µl Click-iT reaction (85.8 µl of 1X Click-iT reaction buffer, 4 µl of 100 mM $CuSO_4$, 0.24 µl of Alexa Fluor azide, 10 µl of 1X reaction buffer additive), incubated for 30 min at RT in darkness and washed three times in 800 µl of 1% BSA, 0.5% Tween-20 in PBS. For RNA degradation and DNA stain, cells were incubated with 700 µl of PBS with 7 µl RNase A 10 mg/ml and 5 µl of 7-Amino-Actinomycin D (7-AAD) at least for 30 min before being scored by FACS. Three population of cells were analyzed based on their DNA content (1n=haploid or 2n=diploid DNA content of the genome) and EdU signal: G1 (with 1n DNA content and no EdU signal), S (with DNA content between 1n and 2n and EdU signal) and G2 (with 2n DNA content and no EdU incorporation) -phase cells.

9.2.   Cell cycle progression (Synchronization by double thymidine block)

Synchronization of cells population in G_1/S to monitored S phase progression was performed using a double thymidine block. Thymidine blocks DNA synthesis by inhibiting the synthesis of specific nucleotides, resulting in a reversible arrest in S phase with partly synthetized (Crosby 2007). Hela cells were transfected with siRNA at 30-40% confluence. 24 hours after transfection, growth medium was replaced with fresh

medium containing 2 mM thymidine and cells were incubated for 19 h under normal conditions. After three washed in PBS, cells were released in fresh complete medium and incubated under normal conditions for 7 h. Following incubation, medium was replaced with fresh medium containing 2 mM thymidine and cells were incubated for 17-19 h under normal conditions. Finally, cells were washed three times in PBS and released in fresh complete medium. 20 min before collect the samples, cells were incubated with with EdU (20 µM) and then processed as described in Materials and Methods 9.1. Time points were collected 0, 1, 2, 4, 6 and 9 h after release.

10. SISTER CHROMATIN EXCHANGE (SCE) ASSAY

SCE assay was performed as previously described (Bayani and Squire 2005). Briefly, 24 hours after siRNA transfection HeLa cells were incubated with 10 µM BrdU for 42 hours followed by 3 hours treatment with 0.1 µg/ml of KaryoMAX colcemid solution (Invitrogen). Cells were resuspended in 0.075 M KCl and incubated at 37 °C for 10 min followed by 3 changes of Carnoy fixative (3:1 methanol:acetic acid). Cells were dropped onto slides and baked overnight at 65 °C. To differentially stain the two chromatids, the slides were incubated with 20 µg/ml Hoechst solution for 20 min, then exposed for 1 hour to UVA irradiation in 2X SSC and incubated in 2X SSC at 60°C for 20 min before standard Giemsa staining was performed.

20X SSC (sodium chloride/sdium citrate): 3 M NaCl, 0.3 M Na$_3$citrate 2H$_2$O. pH 7 adjusted with 1 M HCl.
Giemsa: 1:20 in buffer Weise pH 6.8 (Millipore).

11. IMMUNOFLUORESCENCE

Cells were cultured on glass coverslips, fixed in formaldehyde or methanol and specific target molecules were visualized in a fluorescence microscope after incubation with specific primary antibodies and with fluorophore-conjugated secondary antibodies.

Types of cell fixation used:
- Formaldehyde fixation: Cells were fixed in 2% formaldehyde in PBS for 20 min at RT and permeabilized with 70% ethanol for 5 min at -20 °C, 5 min at 4 °C, and

washed twice in PBS. This fixation was used for immunofluorescence with THOC1, γH2AX, 53BP1 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 MFAP1-SC35 antibodies (Figure R25) and PLA experiments using MFAP1, SAP130, SIN3, HDAC2, THOC1, THOC2, THOC5 and UAP56 and antibodies (see Figure M2).

- 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.

- Methanol-Formaldehyde fixation: Cells were pre-permeabilized with cold 0.25% triton in PBS for 1 min on ice, fixed in cold absolute methanol for 15 min at -20 °C, washed in 0.5% triton in PBS for 5 min at 4 °C, fixed in 2% formaldehyde in PBS for 20 min at RT and washed twice in PBS. This fixation was used for immunofluorescence with H3K9me2 antibody.
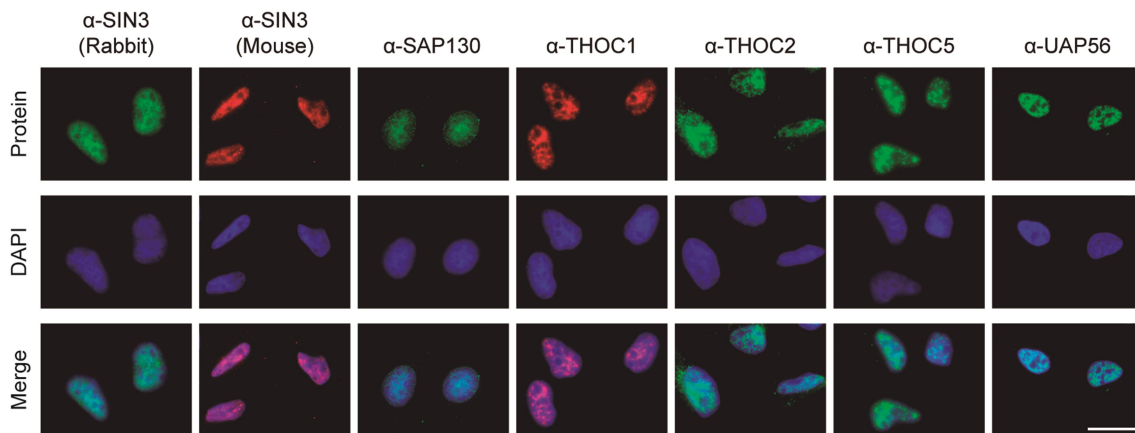
Figure M2. Immunofluorescence with antibodies used in PLA experiments.
Immunofluorescence of endogenous SIN3, SAP130, THOC1, THOC2, THOC5 and UAP56 proteins using compatible fixation conditions for each pair of antibodies used in PLA experiments.

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

After blocking with 3% bovine serum albumin (BSA) in PBS, the coverslips were incubated with primary antibodies diluted in 3% BSA in PBS for 2 h at RT. For antibody dilutions see Table M1. Then 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 incubated for 1 h at RT. 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. Finally, coverslips were washed in water and a drop of Immu-mount mounting medium (Thermo) was used for mounting. For S9.6 and nucleolin immunofluorescence see Materials and Methods 12.3.2.

Detection of cells in S phase by immunofluorescence:

When required, 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.

In this assay the modified thymidine analogue EdU (5-ethynyl-2'-deoxyuridine) is efficiently incorporated into DNA during active synthesis and fluorescently labeled with a bright, photostable Alexa Fluor dye in a reaction that allows the visualization of cells in S phase.

Cells were cultured on glass coverslips, incubated with EdU (10 µM) for 20 min and fixed in formaldehyde or methanol. Then 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 $CuSO_4$, 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.

12. GENOME INSTABILITY ANALYSIS

12.1. Analysis of 53BP1 and γH2AX foci

Cells were cultured on glass coverslips and transfected with siRNA as indicated in Material and Methods 5.3.1, fixed in 2% formaldehyde and immunofluorescence was performed with rabbit polyclonal anti-53BP1 (NB100-304 Abyntec Biopharma) or mouse monoclonal anti-γH2AX (JBW301, 05-636 Millipore) (see Materials and Methods 11 and Table M1). More than 100 cells from each experiment were analyzed (see Materials and Methods 13).

12.2. 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. In the case of comet assay experiment overexpressing RNase H1, cells were transfected with siRNA for 72 h and with pcDNA3 (- RNase H1) or pcDNA3-RNase H1 (+ RNase H1) 48 h before the experiment.

12.2.1. 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 dH_2O 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.

12.2.2. Neutral comet assay

Cells were collected and immobilized on CometSlides as in alkaline come assay. Cells were lysed for 1 h at 4 °C and immersed in prechilled 1X neutral electrophoresis buffer for 30 min at 4 °C. Electrophoresis was performed in prechilled 1X neutral electrophoresis buffer at 35 V for 15 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 SYBR Green at 4 °C for 30 min.

10X Neutral electrophoresis buffer (500 ml): 60.57 g Tris Base and 204.12 g of sodium acetate dissolved in H_2O. Adjust to pH=9.0 with glacial acetic acid. 1X stock was obtained by diluting the 10X stock in dH_2O.

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

For 53BP1 and 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 13).

12.3. DNA-RNA hybrids detection

12.3.1. DNA-RNA immunoprecipitation (DRIP)

DRIP was performed mainly as described (Ginno et al. 2012) with few differences. After 72 h of siRNA transfection, pellet from one confluent 6-cm plate of HeLa cells was collected using accutase, washed in PBS and resuspended in 800 µl of TE. Then, 20.75 µl SDS 20% and 2.5 µl proteinase K (20 mg/ml) were added and pellet was incubated at 37 °C overnight. DNA was extracted gently with phenol-chloroform. 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 *Hind*III, *EcoR*I, *BsrG*I, *Xba*I and *Ssp*I, 2 mM spermidine and BSA. For the negative control, half of the DNA was treated with 3 µl RNase H (New England BioLabs) overnight. 5 µg of the digested DNA were bound to 10 µl of S9.6 antibody (1mg/ml) in 500 µl 1X binding buffer in TE, overnight at 4 °C. DNA-antibody complexes were immunoprecipitated using Dynabeads Protein A (Invitrogen) during 2 h at 4 °C and washed 3 times with 1X binding buffer in TE. DNA was eluted in 180 µL elution buffer, treated 45 min with 7 µl proteinase K (20 mg/ml) at 55ºC and cleaned using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). qPCR was performed with the primers listed in Table M8.

10X Binding buffer: 100 mM NaPO$_4$ pH 7.0, 1.4 M NaCl, 0.5% triton X-100.
Elution buffer: 50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS.

DRIP quantification and normalization: Input and immunoprecipitate (IP) were eluted in 150 µl of double-distilled H$_2$O. 4 µl of 1:25 dilutions of the Input and 4 µl of IP were used for qPCR. The relative abundance of DNA-RNA hybrids immunoprecipitated in each region was normalized to the Input signal obtained.

12.3.2. S9.6 immunofluorescence

Cells were cultured on glass coverslips and fixed in methanol (see Materials and Methods 11). For S9.6 and nucleolin immunofluorescence, coverslips were blocked in 2% BSA in PBS overnight at 4 °C, incubated with anti-S9.6 (hybridoma cell line HB-8730) (1:200) and anti-nucleolin (ab50279, Abcam) (1:1000) primary antibodies diluted in 2% BSA in PBS o/n at 4 °C, washed in PBS and incubated with secondary antibodies conjugated with Alexa Fluor 594 and 488 diluted in 2% BSA in PBS (1:1000) for 1 h at RT. Washed, DAPI staining and mounting as described above. More than 100 cells from each experiment were scored (see Materials and Methods 13).

12.4. Anaphase bridges analysis

HeLa cells were cultured on glass coverslips and transfected with siRNA for 72 h. In order to increase the proportion of cells in mitosis, cells were treated with 50 ng/ml nocodazol in complete medium for 4 h at 37 °C, washed twice carefully in PBS and incubated for 1 h in 2 ml complete medium before to be harvested. Then, cells were fixed and permeabilized directly in the well adding 2 ml of 4% formaldehyde – 0.5% Triton X-100 in PBS during 20 min at RT. After immersing the coverslips carefully in PBS vectashield mounting medium with DAPI was used for the mounting. The percentage of cells in anaphase with anaphase bridges was calculated. More than 100 cells from each experiment were analyzed (see Materials and Methods 13).

13. MICROSCOPY IMAGES ACQUISITION, DATA ANALYSIS AND STATISTICAL ANALYSIS

13.1. Fluorescence and confocal microscopy

Fluorescence microscopy images were captured with a Leica DM6000 microscope equipped with a DFC390 camera (Leica) and data acquisition was performed with LAS AF (Leica). A Leica TCS SP5 confocal microscope was used to capture images of H3K9me2 immunofluorescence. A 100x objective was used for SCE, a 63x objective was used for immunofluorescence (53BP1 and γH2AX and H3S10-P foci, H3K9me2 IF, PLA, S9.6 and nucleolin IF) and HeLa DAPI staining (Anaphase bridges), a 40x

objective was used for DNA combing experiments and a 10x objective was used for comet assays.

13.2. Data analysis

- 53BP1, γH2AX, H3S10-P foci measurements were analyzed and processed with the MetaMorph v7.5.1.0. software using the *granularity* application.
- S9.6 and H3K9me2 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).
- Anaphase cells with bridges and SCE events were counted manually.

For all experiments, at least three biological repeats (n) were performed. More than 100 cells were scored in each repeat.

13.3. Statistical analysis

- 53BP1, γH2AX, H3S10-P foci and anaphase bridges: Graphs shows the mean of the percentage of cells with foci or anaphases with bridge 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 and H3K9me2 signal intensity per nucleus: Graphs shows the median of the measurements from at least three biological repeats. Data were analyzed with GraphPad Prism software. For statistical analysis, Mann-Whitney *U*-test, two tailed was performed and P value < 0.05 was considered as statistically significant. (****, $P < 0.0001$; ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$).
- 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.

198

- 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.

14. CHROMATIN IMMUNOPRECIPITATION (ChIP) ASSAY

After 72 h of siRNA transfection, HeLa cells were crosslinked for 10 min with crosslinking solution - formaldehyde, 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 30 s on and 30 s off in Bioruptor (Diagenode), to obtain approx. 400 bp fragments. For each immunoprecipitation, 25 µg of chromatin were diluted up to 1300 µl with IP buffer. 100 µl and 1200 µl of diluted chromatin were used for input and immunoprecipitation, respectively. Chromatin was incubated overnight at 4 ºC with 5 µg of antibody. A negative control without antibody was used to calculate the background signal. Chromatin-antibody complexed were immunoprecipitated for 2 h with 30 µl of Dynabeads Protein A/G (Invitrogen) at 4 ºC and washed once with wash buffer 1, once with wash buffer 2, once with wash buffer 3 and twice with 1X TE. Input and immunoprecipitate 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 qPCR was performed with the primers listed above. Signal values in the different regions were calculated as the ratio between the DNA amount immunoprecipitated subtracting the background signal (IP) and the total amount of DNA (input) of each region.

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

ChIP quantification and normalization: Input and immunoprecipitate (IP) were eluted in 150 µl of double-destilled H_2O. 4 µl of 1:10 dilutions of the Input and 4 µl of IP were used for qPCR. IP/Input ratios in the different regions were calculated and multiplied by 100. The IP signal of the negative control without antibody was considered as background and was subtracted to IP before dividing it by Input.

15. MICROARRAY ANALYSES OF GENE EXPRESSION AND ALTERNATIVE SPLICING

Gene expression and alternative splicing microarray was performed using the Affymetrix platform GeneChip Human Transcriptome array 2.0 (HTA 2.0). This array contains a median of 10 probes per exon or ncRNA and 4 probes per junction (>6.0 million distinct probes in total).

Total RNA was isolated from HeLa cells transfected with siRNA for 72 h using the RNeasy Mini Kit (Qiagen). The quality of isolated RNA was confirmed with a 2100 Bioanalyzer (Agilent Technologies). A total of 250 ng of RNA per microarray were used. cDNA synthesis and labelling and microarray hybridization to HTA 2.0, scanning, and data extraction were performed by the CABIMER´s Genomics Unit according to the recommended Affymetrix protocols.

15.1. Microarray data analyses

For each condition, microarray analysis was conducted in triplicate from different biological replicates. Affymetrix Command Console 2.0 Software was used to transform scanned image signals (saved as DAT image files) into digital signals (saved as CEL files), which recorded the fluorescence density of probes. Raw data from CEL files were normalized using the SST (Signal Space transformation)-RMA (Robust Microarray Average) method (Irizarry et al. 2003) and the average of the replicates was calculated using the Tukey´s Bi-weight algorithm, thus obtaining chp files. Data were processed using Transcriptome Analysis Console (TAC) Software from Affymetrix, to detect differentially expressed genes (DEGs) and alternatively spliced genes (ASGs).

For gene expression analysis, genes showing expression changes with P-value<0.05 (ANOVA) and a |linear fold change| >1.5 were considered significant.

For alternative splicing analysis, the filter criteria to consider that the alternative splicing of a gene is significantly altered were as follows: A gene is expressed in both conditions, A PRS/Junction must be expressed in at least one condition, a gene must contain at least one PRS, |Splicing Index (SI)| was >2 and P-value<0.01 (ANOVA). Splicing Index (SI) was calculated as:

$$SI = Log_2 \left[\frac{\left(\dfrac{\text{Exon 1 Condition 1 Intensity}}{\text{Gene 1 Condition 1 Intensity}} \right)}{\left(\dfrac{\text{Exon 1 Condition 2 Intensity}}{\text{Gene 1 Condition 2 Intensity}} \right)} \right]$$

Moreover, the analysis using TAC Software provides the Splicing Event Estimate method as tool to classify splicing events in the five basic modes of alternative splicing (Figure M3), not being a tool intended for a *de novo* event discovery.

Gene Ontology (GO) analyses, enriched biological processes (BPs) and functional annotation clustering analyses were performed using DAVID (Database for Annotation, Visualization and Integrated Discovery) bioinformatics resources 6.8 tool (http://david-d.ncifcrf.gov/summary.jsp) (Huang da et al. 2009). A P-value<0.05 or P-value<0.01 was established to consider GO terms as significantly enriched in gene expression and alternative splicing analyses, respectively. Adjusted P-value (Q-value) by the Benjamini-Hochberg procedure was also provided.
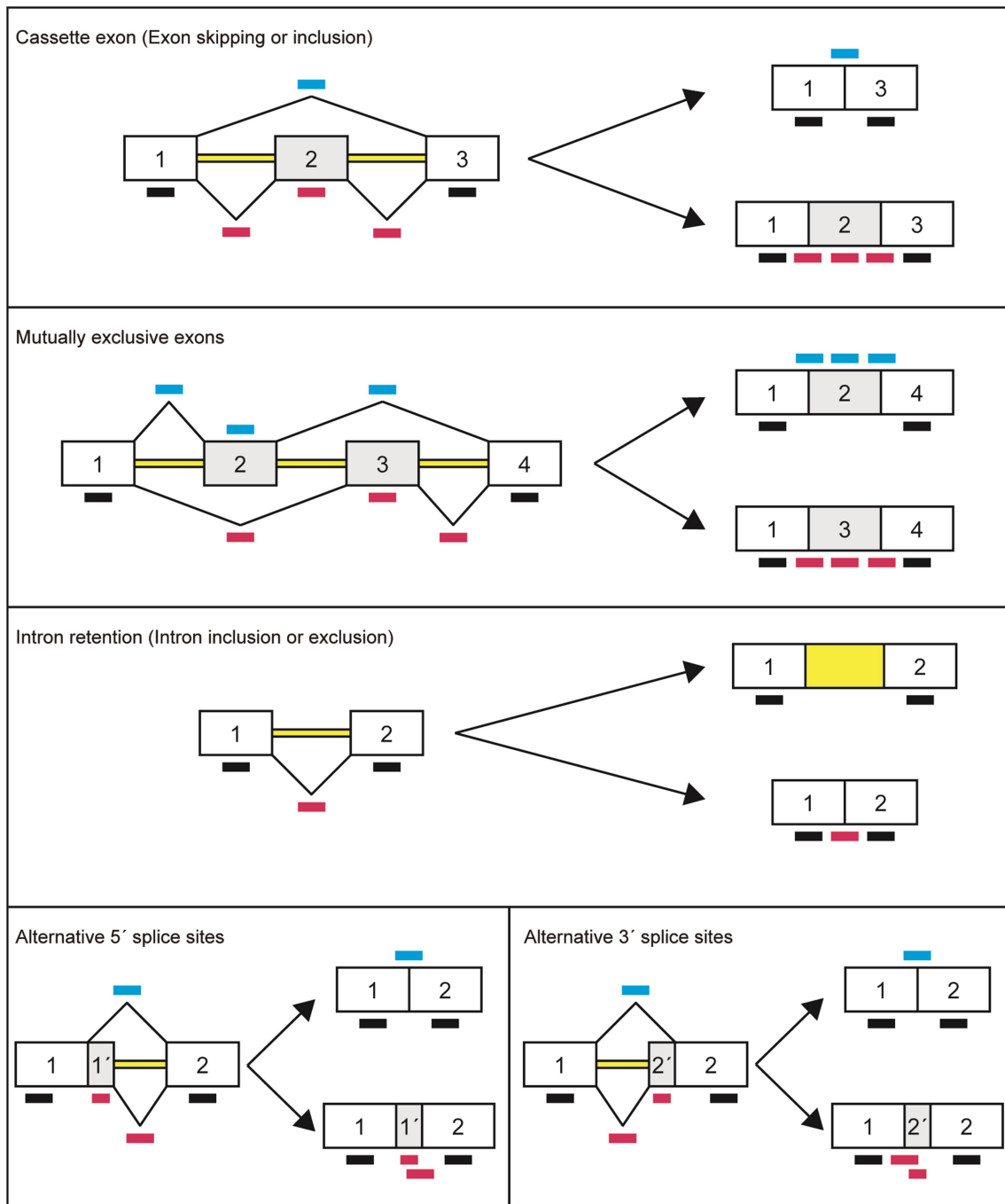
Figure M3. Basic modes of alternative splicing.
Schematic diagrams of the five main types of alternative splicing: Cassette exon (Exon skipping or inclusion), mutually exclusive exons, intron retention (Intron inclusion or exclusion), alternative 5´splice sites and alternative 3´splice sites. Exons are depicted as white boxes and alternative exons are indicated as grey boxes. Introns are represented as yellow lines and splicing events are indicated with diagonal lines. Short black, blue and pink lines represent different assay probes, covering exon-exon and splice junctions.

16. YEAST TECHNIQUES

16.1. Genotoxic damage sensitivity assay

Mid-log cultures were grown in YPAD or SC medium. Drops of 3 or 7 µl of 10-fold dilutions in sterile water were plated on solid YPAD or SC medium, respectively, containing the drugs at the concentrations indicated in the figures. For UV irradiation, drops were dried before irradiation. Plates were incubated during 2-6 days (in the dark for UV-irradiated plates) at the indicated temperature.

16.2. Recombination assays

Recombination frequencies were calculated as the median value of six independent colonies. The average of three independent transformants is plotted. Recombinants were obtained by platting appropriate dilutions in applicable selective medium. To calculate total number of cells, they were plated in the same media as the original transformation used. All plates were grown for 3-5 days at 30ºC.

Plasmid L and LYΔNS systems

Both systems are based on *leu2Δ3'* and *leu2Δ5'* truncations of the *LEU2* gene that share 600 bp of homology. The difference between them resides in the length of the sequence placed between the repeats. The L system contains a 31-bp fragment and the LYΔNS system contains a 3.7-kb fragment as intervening sequence (Prado and Aguilera 1995; Prado et al. 1997). In this system, recombination between the repeats results in a wild-type *LEU2* gene. Recombinants are selected in plates without leucine (Figure M4A).

Plasmid L-lacZ systems

This system contains the sequence of the 3 kb long *lacZ* gene from *E. coli* cloned between the direct repeats (*leu2Δ3'* and *leu2Δ5'* truncations of the *LEU2*) (Chavez and Aguilera 1997). In this system, recombination between the repeats results in a wild-type *LEU2* gene. Recombinants are selected in plates without leucine (Figure M4B).

Chromosomal *leu2-k::ADE2-URA3::leu2-k* (Lk-AU) system

This system is located in chromosome III. It consists in two direct-repeats of the *leu2-K* allele, which lacks a 7-bp region that contains a *Kpn*I target. It contains an 11-kb

fragment containing *ADE2* and *URA3* markers between the two *leu2-k* repeats (Aguilera and Klein 1989). In this system, recombination results in deletion of *ADE2* and *URA3* genes. Recombinants are selected in plates with FOA, since loss of *URA3* confers cells resistance to this toxic. In addition, recombinants generate red-white colonies due to a red pigment that accumulates by the loss *ADE2* gene (Figure M4C).
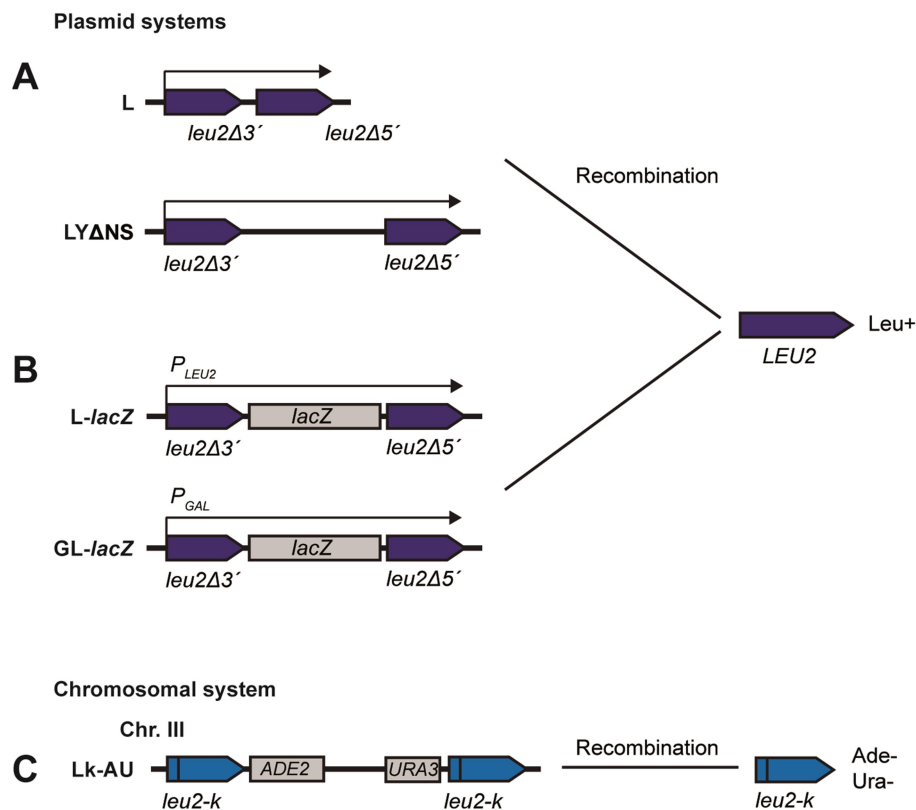
Figure M4. Recombination systems used in *S. cerevisiae*.
Schematic representation of the different recombination systems used in yeast. **(A)** L and LYΔNS systems, **(B)** L-lacZ and GL-lacZ systems and **(C)** Lk-AU system. The outcome of the recombination event and the marker used for recombinant selection are shown.

16.3. Detection of Rad52-YFP foci

Spontaneous Rad52-YFP foci from mid-log growing cells carrying plasmid pWJ1213 or pWJ1344 were visualized and counted by fluorescence microscopy in a Leica DC 350F microscope, as previously described (Lisby et al. 2001). More than 100 S/G2 cells where inspected for each experimental replica.

17. POLYMERASE CHAIN REACTION (PCR) ANALYSIS

17.1. Non-quantitative PCR

PCRs performed to check constructions and to amplify DNA fragment to be cloned or to be used as probed were performed following standard and manufacture's protocols with the polymerases described in Materials and Methods 2.3. DNA primers used are listed in Table M8.

17.2. Quantitative PCR analysis

Real-time quantitative PCRs (qPCRs) were performed on a 7500 Fast Real-Time PCR system (Applied Biosystems, Carlsbad, CA). For PCRs, 6 µl H_2O, 2 µl primer mixture (each 10 µM), 2 µl DNA and 10 µl SYBR® green qPCR Mix (Bio-rad) were used. The following PCR reaction was used: 1 cycle (10 minutes 95 °C), 40 cycles (15 s 95 °C and 1 minute 65 °C) and 1 dissociation cycle (15 s 95 °C, 1 minute 65 °C, 15 s 95 °C and 15 s 60 °C). DNA primers were designed using Primer express 3.0 Software (Applied Biosystems) and are listed in Table M8. 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).

17.2.1. Reverse Transcription quantitative PCR (RT-qPCR) analysis

Relative mRNA levels in human cells were determined using Relative qPCRs. To obtain the samples, cDNA was synthesized 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 of the indicated genes were normalized to mRNA expression of the Hypoxanthine PhosphoRibosylTransferase (HPRT) housekeeping gene.

17.2.2. qPCR analysis for ChIPs and DRIP quantification

Absolute qPCRs were used for ChIP and DRIP quantification.

Table M8. DNA primers used in this study.

Non-quantitative PCR primers		
Primer	**Sequence 5´to 3´**	**Use**
pGBKT7-THOC1 up	CATATGGCCATGGAGGCCGAATTCATGTCTCCGACGCCGCCGCTCTTC	Construction of bait plasmid: THOC1 PCR
pGBKT7-THOC1 low	GCAGGTCGACGGATCCCTAACTATTTGTCTCATTGTCATT	
pGBKT7-UAP56 up	TATGGCCATGGAGGCCGAATTCATGGCAGAGAACGATGTGGACAAT	Construction of bait plasmid: UAP56 PCR
pGBKT7-UAP56 low	GCAGGTCGACGGATCCCTACCGTGTCTGTTCAATGTAGGA	
pGBKT7-URH49 up	CATGGAGGCCGAATTCATGGCAGAACAGGATGTGGAAAAC	Construction of bait plasmid: URH49 PCR
pGBKT7-URH49 low	CTGCAGGTCGACGGATCCTTACCGGCTCTGCTCGATGTATGT	
pGBKT7-ALY up	CATATGGCCATGGAGGCCGAATTCATGGCCGACAAAATGGACATGTCT	Construction of bait plasmid: ALY PCR
pGBKT7-ALY low	GCAGGTCGACGGATCCTTAACTGGTGTCCATTCTCGCATT	
T7 Sequencing up	TAATACGACTCACTATAGGGC	Sequencing of prey inserts
3'DNA-BD Seq. low **3'AD Seq. low**	TTTTCGTTTTAAAACCTAAGAGTC	
pFLAG-THOC1 up	CGCGAATTCAGATCTCATGTCTCCGACGCCGCCGCTCTTC	Plasmid construction: THOC1 PCR
pFLAG-THOC1 low	GCAGGTCGACGGATCCCTAACTATTTGTCTCATTGTCATT	
pFLAG-UAP56 up	GACAAGCTCAAGCTTATGGCAGAGAACGATGTGGACAAT	Plasmid construction: UAP56 PCR
pFLAG-UAP56 low	GCAGGTCGACGGATCCCTACCGTGTCTGTTCAATGTAGGA	
pFLAG-URH49 up	GACAAGCTCAAGCTTATGGCAGAACAGGATGTGGAAAAC	Plasmid construction: URH49 PCR
pFLAG-URH49 low	CTGCAGGTCGACGGATCCTTACCGGCTCTGCTCGATGTATGT	
pFLAG-ALY up	GACAAGCTCAAGCTTATGGCCGACAAAATGGACATGTCT	Plasmid construction: ALY PCR
pFLAG-ALY low	GCAGGTCGACGGATCCTTAACTGGTGTCCATTCTCGCATT	
pEGFP-URH49 up	CATGGAGGCCGAATTCATGGCAGAACAGGATGTGGAAAAC	Plasmid construction: URH49 PCR
pEGFP-URH49 low	CTGCAGGTCGACGGATCCTTACCGGCTCTGCTCGATGTATGT	
pEGFP-SAP130 up	GGCCGGACTCAGATCTCGATGGGCCCTCCGCGGCACCCCCAG	Plasmid construction: SAP130 PCR
pEGFP-SAP130 low	CCCGGGCCCGCGGTACCGCTAGACTTTTTCCTTTCGCTTCAA	

Non-quantitative PCR primers		
Primer	**Sequence 5´ to 3´**	**Use**
pEGFP-MFAP1 up	GGCCGGACTCAGATCTCGATGTCGGTCCCAAGCGCTCTC	Plasmid construction: MFAP1 PCR
pEGFP-MFAP1 low	CCCGCGGTACCGTCGACTCTAGGTAGTTTTCCGCTTCTT	
pEGFP-SMC2 up	GGCCGGACTCAGATCTCGATGCATATTAAGTCAATTATTCTA	Plasmid construction: SMC2 PCR
pEGFP-SMC2 low	CCCGCGGTACCGTCGACTTTAAACTTCCACATGTGCTCCTTT	
pEGFP-REV1 up	CGGACTCAGATCTCGAGCATGTTGCATGGAGGTCAATACCAT	Plasmid construction: REV1 PCR
pEGFP-REV1 low	CCCGCGGTACCGTCGACTCTAACACAGGCCAAGCCAACCTGA	
pEGFP-CDC23 up	GGCCGGACTCAGATCTCGATGGTCCCGGTGGCTGTGACG	Plasmid construction: CDC23 PCR
pEGFP-CDC23 low	CCCGCGGTACCGTCGACTCTATTTAAAGATTTAACTTT	
pEGFP-CCT5 up	GCTCAAGCTTCGAATTCATGGCGTCCATGGGGACCCTCGCC	Plasmid construction: CCT5 PCR
pEGFP-CCT5 low	CTAGATCCGGTGGATCCCTCATTCTTCAGATTCTCCAGGCTT	
pEGFP-NUB1 up	GGCCGGACTCAGATCTCGATGAAAACAAACGGCGGCCGC	Plasmid construction: NUB1 PCR
pEGFP-NUB1 low	CCCGCGGTACCGTCGACTTTAGTTTTTCTTTGTTGCTGA	
pEGFP-CUL1 up	AGCTTCGAATTCTGCAGTATGTCGTCAACCCGGAGCCAG	Plasmid construction: CUL1 PCR
pEGFP-CUL1 low	GATCCGGTGGATCCCTTAAGCCAAGTAACTGTAGGT	
pEGFP-SPATA22 up	GGCCGGACTCAGATCTCGATGAAGCGAAGCCTAAATGAAAAT	Plasmid construction: SPATA22 PCR
pEGFP-SPATA22 low	CCCGCGGTACCGTCGACTTCACGATCCTAATACTTCAAAAAG	
pEGFP-ZBTB44 up	GGCCGGACTCAGATCTCGATGGGTGTGAAAACATTTACTCAT	Plasmid construction: ZBTB44 PCR
pEGFP-ZBTB44 low	CCCGCGGTACCGTCGACTTCATTCTCCGTTCCTGGTCATTTC	

Quantitative PCR primers		
Primer	**Sequence 5´ to 3´**	**Use**
HPRT up **HPRT low**	GGACTAATTATGGACAGGACTG TCCAGCAGGTCAGCAAAGAA	Relative mRNA expression
MFAP1 up **MFAP1 low**	CGCTCAGGTCGCACCAA AGCTGAGTCAAAGGAGGTGGTATCT	Relative mRNA expression
SAP130 up **SAP130 low**	GCCACGGCCAAGCATACT GGAATGAGGGTTTTCCGAACT	Relative mRNA expression
SIN3 up **SIN3 low**	GCCTGAGCTTCGTGAACATCT CTC TCGACCACGTTGACACTTC	Relative mRNA expression

Quantitative PCR primers		
Primer	**Sequence 5´to 3´**	**Use**
THOC1 up **THOC1 low**	GGAACCCTGTGCAATGC TATG ACATGTTCTCCTCCTGTTTTCAATT	Relative mRNA expression
ACTB_1 up **ACTB_1 low**	CGGCCAACGCCAAAACT TCCCCTCCTTTTGCGAAAA	ChIP. Promoter region of the *β-Actin* gene
ACTB_2 up **ACTB_2 low**	AGGCATCCTCACCCTGAAGTAC TCCATGTCGTCCCAGTTGGT	ChIP. Exon 3 region of the *β-Actin* gene
ACTB_3 up **ACTB_3 low**	GGCGGCACCACCATGTAC CAGGGCAGTGATCTCCTTCTG	ChIP. Exon 5 region of the *β-Actin* gene
ACTB_4 up **ACTB_4 low**	CAGAGTGCAGGTGTGTGGAGAT CCTCCCCCACCCTCTAAGG	ChIP. Pause region of the *β-Actin* gene
GAPDH_1 up **GAPDH_1 low**	CCCCCTTCCTTACAAGTGTTCA GGGCAGAGGGCCAGGTT	ChIP. Promoter region of the *GAPDH* gene
GAPDH_2 up **GAPDH_2 low**	CTCTCTGCTCCTCCTGTTCGA CACCTGGCGACGCAAAA	ChIP. TSS region of the *GAPDH* gene
GAPDH_3 up **GAPDH_3 low**	CCGGGAAGGAAATGAATGG CAGGAGCGCAGGGTTAGTCA	ChIP. Intron 2 region of the *GAPDH* gene
GAPDH_4 up **GAPDH_4 low**	ATGCTGGCGCTGAGTACGT GCCTTCTCCATGGTGGTGAA	ChIP. Exon 5 region of the *GAPDH* gene
GAPDH_5 up **GAPDH_5 low**	GCGTGTAGCTGGGACCTAGGT CAAAGCAACTGAGCCGTTCA	ChIP. 5' region of the *GAPDH* gene
APOE up **APOE low**	GGGAGCCCTATAATTGGACAAGT CCCGACTGCGCTTCTCA	DRIP ChIP
APOE_2 up **APOE_2 low**	GGCACACAAGGACACTCAATACA CCCAAAGTGCTGGGATTACAG	DRIP ChIP
APOE_3 up **APOE_3 low**	AAGCTGGAGGAGCAGGCC ACTGGCGCTGCATGTCTTC	DRIP ChIP
BTBD19 up **BTBD19 low**	CCCCAAAGGGTGGTGACTT TTCACATTACCCAGACCAGACTGT	DRIP ChIP
RPL13A up **RPL13A low**	GCTTCCAGCACAGGACAGGTAT CAC CCACTACCCGAGTTCAAG	DRIP ChIP
RPL13A_2 up **RPL13A_2 low**	ACTGGGCAGGCCTCACACT CGCTTGCGGAGGAAAGC	DRIP ChIP
RPL13A_3 up **RPL13A_3 low**	GGGAGCAAGGAAAGGGTCTTA ACAATTCTCCGAGTGCTTTCAAG	DRIP ChIP
EGR1 up **EGR1 low**	GCCAAGTCCTCCCTCTCTACTG GGAAGTGGGCAGAAA GGATTG	DRIP ChIP
SNRPN up **SNRPN low**	TGCCAGGAAGCCAAATGAGT TCCCTCTTGGCAACATCCA	DRIP ChIP
PMA1 3´ up **PMA1 3´ low**	ATCGCTATTTTCGCTGATGTTG CGGGCTTTGGAGAGTAAGGA	Relative mRNA expression
GCN4 3´ up **GCN4 3´ low**	TTGTGCCCGAATCCAGTGA TGGCGGCTTCAGTGTTTCTA	Relative mRNA expression
ADH1 3´ up **ADH1 3´ low**	TGGTCAAGTCTCCAATCAAGGTT CCAACGATTTGACCCTTTTCC	Relative mRNA expression

18. PROTEIN ANALYSIS

18.1. Yeast protein extraction by TCA (trichloracetic acid)

For protein extraction, 10 ml of mid-log yeast culture were centrifuged and kept in ice. Pellets were resuspended in 200 µl of cold 10% TCA and 200 µl of glass beads, and cells were broken by vortexing 7 times 20 seg each time at 4ºC. Supernatant was recovered and beads were washed twice with 200 µl of cold 10% TCA. Samples were centrifuged 10 min at 3000 rpm and supernatant discarded. The remaining pellet was resuspended using 100 µl of 2x Loading Buffer supplemented with protease inhibitors, 50 µl of water and 50 µl of 1M Tris (not-adjusted pH). Prior to gel loading samples were boiled for 5 min and centrifuged 10 min at 3000 rpm at RT.

2X Loading buffer: 62.5 mM Tris-HCl pH 6.8, 25% glycerol, 2% SDS, 0.01% water-diluted Bromophenol Blue, 5% β-mercaptoethanol.
Protease inhibitor: 1 mM PMSF, 66 µg/ml chymostatin.

18.2. Human cells protein extraction

Pellet of HEK293T or HeLa cells was collected using accutase, washed in cold PBS and resuspended in ice-cold lysis buffer (100 µl/$1x10^6$ cells) during 30 min on ice with occasionally gently pipetting up and down. The lysate was centrifuged 10 min at 16000 g and the supernatant was transfer to a new tube. Prior to gel loading, 4X laemmi buffer was added to 1X final concentration and samples were boiled for 5 min.

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

18.3. SDS-PAGE

Proteins were separated in 29:1 acrylamide:bis-acrylamide gels with concentrations appropriate to the molecular size of the proteins of interest or in 4-20% gradient SDS-PAGE CriterionTM TGX TM Precast Gels (BioRad) and SDS-PAGE was performed according to previously described method (Laemmli 1970). Electrophoreses were

performed in a Mini-PROTEAN 3 Cell with Running Buffer at 100 V. Page RulerTM (Fermentas) was used as a protein marker.

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

18.4. 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.

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.

18.4.1. Non-fluorescent 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 during 2 h at RT or o/n at 4 °C in 1X TBS - 0.1% Tween 20 - 5% milk or Blocking Buffer solution. After 3 washes of 10 min each one, membranes were incubated with the corresponding secondary antibodies conjugated with the horseradish peroxidase for 1h hour and washed again. Finally, SuperSignalR West Pico (Pierce) or Immobilon Western Chemiluminescent HRP Substrate (Millipore) was used for chemiluminescence detection depending on the expected strength of the signal.

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.

18.4.2. 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 the transference. Commercial Odyssey Blocking Buffer (LI-COR Biosciences) was used to block the membrane for 1 h at RT or o/n at 4 °C. Primary antibody was prepared to the 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).

212

REFERENCES

Agalioti T, Chen G, Thanos D. 2002. Deciphering the transcriptional histone acetylation code for a human gene. *Cell* **111**: 381-392.

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

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

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

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

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

Aguilera A, Klein HL. 1989. Genetic and molecular analysis of recombination events in Saccharomyces cerevisiae occurring in the presence of the hyper-recombination mutation hpr1. *Genetics* **122**: 503-517.

Aguilera A, 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 RD, Innocente SA, Barrass JD, Beggs JD. 2010. Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell* **40**: 582-593.

Andersen DS, Tapon N. 2008. Drosophila MFAP1 is required for pre-mRNA processing and G2/M progression. *J Biol Chem* **283**: 31256-31267.

Aparicio JG, Viggiani CJ, Gibson DG, Aparicio OM. 2004. The Rpd3-Sin3 histone deacetylase regulates replication timing and enables intra-S origin control in Saccharomyces cerevisiae. *Mol Cell Biol* **24**: 4769-4780.

Ayer DE, Lawrence QA, Eisenman RN. 1995. Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell* **80**: 767-776.

Ayrapetov MK, Gursoy-Yuzugullu O, Xu C, Xu Y, Price BD. 2014. DNA double-strand breaks promote methylation of histone H3 on lysine 9 and transient formation of repressive chromatin. *Proc Natl Acad Sci U S A* **111**: 9169-9174.

Azvolinsky A, Giresi PG, Lieb JD, Zakian VA. 2009. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Mol Cell* **34**: 722-734.

Bachl J, Carlson C, Gray-Schopfer V, Dessing M, Olsson C. 2001. Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J Immunol* **166**: 5051-5057.

Balasubramanyam K, Swaminathan V, Ranganathan A, Kundu TK. 2003. Small molecule modulators of histone acetyltransferase p300. *J Biol Chem* **278**: 19134-19140.

Bayani J, Squire JA. 2005. Sister chromatid exchange. *Curr Protoc Cell Biol* **Chapter 22**: Unit 22 27.

Beaulieu AM, Sant'Angelo DB. 2011. The BTB-ZF family of transcription factors: key regulators of lineage commitment and effector function development in the immune system. *J Immunol* **187**: 2841-2847.

Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Studitsky VM, Reinberg D. 2003. FACT facilitates transcription-dependent nucleosome alteration. *Science* **301**: 1090-1093.

Ben-Aroya S, Coombes C, Kwok T, O'Donnell KA, Boeke JD, Hieter P. 2008. Toward a comprehensive temperature-sensitive mutant repository of the essential genes of Saccharomyces cerevisiae. *Mol Cell* **30**: 248-258.

Bentley DL. 2014. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**: 163-175.

Bhatia V, Barroso SI, Garcia-Rubio ML, Tumini E, Herrera-Moyano E, Aguilera A. 2014. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**: 362-365.

Bjork P, Wieslander L. 2017. Integration of mRNP formation and export. *Cell Mol Life Sci*.

Boeke JD, LaCroute F, Fink GR. 1984. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol Gen Genet* **197**: 345-346.

Braunschweig U, Gueroussov S, Plocik AM, Graveley BR, Blencowe BJ. 2013. Dynamic integration of splicing within gene regulatory pathways. *Cell* **152**: 1252-1269.

Brewer BJ. 1988. When polymerases collide: replication and the transcriptional organization of the E. coli chromosome. *Cell* **53**: 679-686.

Brewer BJ, Fangman WL. 1988. A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell* **55**: 637-643.

Brock RD. 1971. Differential mutation of the beta-galactosidase gene of Escherichia coli. *Mutat Res* **11**: 181-186.

Buratowski S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36**: 541-546.

Burns CG, Ohi R, Mehta S, O'Toole ET, Winey M, Clark TA, Sugnet CW, Ares M, Jr., Gould KL. 2002. Removal of a single alpha-tubulin gene intron suppresses cell cycle arrest phenotypes of splicing factor mutations in Saccharomyces cerevisiae. *Mol Cell Biol* **22**: 801-815.

Carrillo Oesterreich F, Preibisch S, Neugebauer KM. 2010. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell* **40**: 571-581.

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP et al. 2005. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**: 581-592.

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

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

Cerritelli SM, Crouch RJ. 2009. Ribonuclease H: the enzymes in eukaryotes. *FEBS J* **276**: 1494-1505.

Ciccia A, Elledge SJ. 2010. The DNA damage response: making it safe to play with knives. *Mol Cell* **40**: 179-204.

Cowley SM, Iritani BM, Mendrysa SM, Xu T, Cheng PF, Yada J, Liggitt HD, Eisenman RN. 2005. The mSin3A chromatin-modifying complex is essential for embryogenesis and T-cell development. *Mol Cell Biol* **25**: 6990-7004.

Crosby ME. 2007. Cell Cycle: Principles of Control. *The Yale Journal of Biology and Medicine* **80**: 141-142.

Custodio N, Carvalho C, Condado I, Antoniou M, Blencowe BJ, Carmo-Fonseca M. 2004. In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. *RNA* **10**: 622-633.

Chan YA, Aristizabal MJ, Lu PY, Luo Z, Hamza A, Kobor MS, Stirling PC, Hieter P. 2014. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS genetics* **10**: e1004288.

Chanarat S, Seizl M, Strasser K. 2011. The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes Dev* **25**: 1147-1158.

Chanarat S, Strasser K. 2013. Splicing and beyond: the many faces of the Prp19 complex. *Biochim Biophys Acta* **1833**: 2126-2134.

Chapman JR, Taylor MR, Boulton SJ. 2012. Playing the end game: DNA double-strand break repair pathway choice. *Mol Cell* **47**: 497-510.

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

Chavez S, 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 AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T, 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 I. 2015. Mammalian mitoribosomes revealed. *Nat Struct Mol Biol* **22**: 361-361.

Chen PB, Chen HV, Acharya D, Rando OJ, Fazzio TG. 2015. R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat Struct Mol Biol* **22**: 999-1007.

Chen YI, Moore RE, Ge HY, Young MK, Lee TD, Stevens SW. 2007. Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res* **35**: 3928-3944.

Cheng H, Dufu K, Lee CS, Hsu JL, Dias A, Reed R. 2006. Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **127**: 1389-1400.

Chi B, Wang Q, Wu G, Tan M, Wang L, Shi M, Chang X, Cheng H. 2013. Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Res* **41**: 1294-1306.

Chien CT, Bartel PL, Sternglanz R, Fields S. 1991. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A* **88**: 9578-9582.

Chon H, Sparks JL, Rychlik M, Nowotny M, Burgers PM, Crouch RJ, Cerritelli SM. 2013. RNase H2 roles in genome integrity revealed by unlinking its activities. *Nucleic Acids Res* **41**: 3130-3143.

Dardenne E, Pierredon S, Driouch K, Gratadou L, Lacroix-Triki M, Espinoza MP, Zonta E, Germann S, Mortada H, Villemin JP et al. 2012. Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness. *Nat Struct Mol Biol* **19**: 1139-1146.

Datta A, Jinks-Robertson S. 1995. Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science* **268**: 1616-1619.

David G, Grandinetti KB, Finnerty PM, Simpson N, Chu GC, Depinho RA. 2008. Specific requirement of the chromatin modifier mSin3B in cell cycle exit and cellular differentiation. *Proc Natl Acad Sci U S A* **105**: 4168-4172.

de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR. 2003. A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* **12**: 525-532.

Dias AP, Dufu K, Lei H, Reed R. 2010. A role for TREX components in the release of spliced mRNA from nuclear speckle domains. *Nat Commun* **1**: 97.

Dominguez-Sanchez MS, Barroso S, Gomez-Gonzalez B, Luna R, Aguilera A. 2011a. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS genetics* **7**: e1002386.

Dominguez-Sanchez MS, Saez C, Japon MA, Aguilera A, Luna R. 2011b. Differential expression of THOC1 and ALY mRNP biogenesis/export factors in human cancers. *BMC Cancer* **11**: 77.

Doostzadeh-Cizeron J, Evans R, Yin S, Goodrich DW. 1999. Apoptosis induced by the nuclear death domain protein p84N5 is inhibited by association with Rb protein. *Mol Biol Cell* **10**: 3251-3261.

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

Duch A, de Nadal E, Posas F. 2013. Dealing with transcriptional outbursts during S phase to protect genomic integrity. *J Mol Biol* **425**: 4745-4755.

Dufu K, Livingstone MJ, Seebacher J, Gygi SP, Wilson SA, 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.

Durfee T, Mancini MA, Jones D, Elledge SJ, Lee WH. 1994. The amino-terminal region of the retinoblastoma gene product binds a novel nuclear matrix protein that co-localizes to centers for RNA processing. *J Cell Biol* **127**: 609-622.

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

Fan HY, Merker RJ, Klein HL. 2001. High-copy-number expression of Sub2p, a member of the RNA helicase superfamily, suppresses hpr1-mediated genomic instability. *Mol Cell Biol* **21**: 5459-5470.

Fields S, Song O. 1989. A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245-246.

Fischer T, Strasser K, Racz A, Rodriguez-Navarro S, Oppizzi M, Ihrig P, Lechner J, Hurt E. 2002. The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. *EMBO J* **21**: 5843-5852.

Fleischer TC, Yun UJ, Ayer DE. 2003. Identification and characterization of three new components of the mSin3A corepressor complex. *Mol Cell Biol* **23**: 3456-3467.

Francisco-Mangilet AG, Karlsson P, Kim MH, Eo HJ, Oh SA, Kim JH, Kulcheski FR, Park SK, Manavella PA. 2015. THO2, a core member of the THO/TREX complex, is required for microRNA production in Arabidopsis. *Plant J* **82**: 1018-1029.

Furumizu C, Tsukaya H, Komeda Y. 2010. Characterization of EMU, the Arabidopsis homolog of the yeast THO complex member HPR1. *RNA* **16**: 1809-1817.

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

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

Gallardo M, Luna R, Erdjument-Bromage H, Tempst P, Aguilera A. 2003. Nab2p and the Thp1p-Sac3p complex functionally interact at the interface between transcription and mRNA metabolism. *J Biol Chem* **278**: 24225-24232.

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

Garcia-Rubio M, Chavez S, Huertas P, Tous C, Jimeno S, Luna R, Aguilera A. 2008. Different physiological relevance of yeast THO/TREX subunits in gene expression and genome integrity. *Mol Genet Genomics* **279**: 123-132.

Garcia-Rubio ML, Perez-Calero C, Barroso SI, Tumini E, Herrera-Moyano E, Rosado IV, Aguilera A. 2015. The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS genetics* **11**: e1005674.

Gari E, Piedrafita L, Aldea M, Herrero E. 1997. A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. *Yeast* **13**: 837-848.

Gatfield D, Le Hir H, Schmitt C, Braun IC, Kocher T, Wilm M, 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, 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.

Gietz RD, Schiestl RH, Willems AR, Woods RA. 1995. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* **11**: 355-360.

Giglia-Mari G, Zotter A, Vermeulen W. 2011. DNA damage response. *Cold Spring Harb Perspect Biol* **3**: a000745.

Giniger E, Ptashne M. 1988. Cooperative DNA binding of the yeast transcriptional activator GAL4. *Proc Natl Acad Sci U S A* **85**: 382-386.

Giniger E, Varnum SM, Ptashne M. 1985. Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* **40**: 767-774.

Ginno PA, Lim YW, Lott PL, Korf I, Chedin F. 2013. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res* **23**: 1590-1600.

Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell* **45**: 814-825.

Gomez-Gonzalez B, 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, Garcia-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marin A, Foiani M, 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, Cowling VH. 2014. Cap-binding complex (CBC). *Biochem J* **457**: 231-242.

Gonzalez-Aguilera C, Tous C, Gomez-Gonzalez B, Huertas P, Luna R, Aguilera A. 2008. The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol Biol Cell* **19**: 4310-4318.

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

Groh M, Lufino MM, Wade-Martins R, Gromak N. 2014. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS genetics* **10**: e1004318.

Grozinger CM, Schreiber SL. 2002. Deacetylase enzymes: biological functions and the use of small-molecule inhibitors. *Chem Biol* **9**: 3-16.

Grzenda A, Lomberk G, Zhang JS, Urrutia R. 2009. Sin3: master scaffold and transcriptional corepressor. *Biochim Biophys Acta* **1789**: 443-450.

Guo S, Hakimi MA, Baillat D, Chen X, Farber MJ, Klein-Szanto AJ, Cooch NS, Godwin AK, Shiekhattar R. 2005. Linking transcriptional elongation and messenger RNA export to metastatic breast cancers. *Cancer Res* **65**: 3011-3016.

Guo S, Liu M, Godwin AK. 2012. Transcriptional regulation of hTREX84 in human cancer cells. *PLoS One* **7**: e43610.

Gwizdek C, Hobeika M, Kus B, Ossareh-Nazari B, Dargemont C, Rodriguez MS. 2005. The mRNA nuclear export factor Hpr1 is regulated by Rsp5-mediated ubiquitylation. *J Biol Chem* **280**: 13401-13405.

Hanahan D. 1983. Studies on transformation of Escherichia coli with plasmids. *J Mol Biol* **166**: 557-580.

Hanawalt PC, Spivak G. 2008. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* **9**: 958-970.

Hansen T, Andersen CB, Echwald SM, Urhammer SA, Clausen JO, Vestergaard H, Owens D, Hansen L, Pedersen O. 1997. Identification of a common amino acid polymorphism in the p85alpha regulatory subunit of phosphatidylinositol 3-kinase: effects on glucose disappearance constant, glucose effectiveness, and the insulin sensitivity index. *Diabetes* **46**: 494-501.

Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ. 1993. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**: 805-816.

Harris ME, Bohni R, Schneiderman MH, Ramamurthy L, Schumperli D, Marzluff WF. 1991. Regulation of histone mRNA in the unperturbed cell cycle: evidence suggesting control at two posttranscriptional steps. *Mol Cell Biol* **11**: 2416-2424.

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

Hautbergue GM, Hung ML, Golovanov AP, Lian LY, Wilson SA. 2008. Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. *Proc Natl Acad Sci U S A* **105**: 5154-5159.

Heath CG, Viphakone N, Wilson SA. 2016. The role of TREX in gene expression and disease. *Biochem J* **473**: 2911-2935.

Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, Will CL, Pena V, Luhrmann R, Stelzl U. 2012. Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell* **45**: 567-580.

Heidemann M, Hintermair C, Voss K, Eick D. 2013. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* **1829**: 55-62.

Heinrich B, Zhang Z, Raitskin O, Hiller M, Benderska N, Hartmann AM, Bracco L, Elliott D, Ben-Ari S, Soreq H et al. 2009. Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in pre-mRNA. *J Biol Chem* **284**: 14303-14315.

Helleday T, Eshtad S, Nik-Zainal S. 2014. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* **15**: 585-598.

Herman RK, Dworkin NB. 1971. Effect of gene induction on the rate of mutagenesis by ICR-191 in Escherichia coli. *J Bacteriol* **106**: 543-550.

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

Heyer WD, Ehmsen KT, Liu J. 2010. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet* **44**: 113-139.

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

Hoeijmakers JH. 2009. DNA damage, aging, and cancer. *N Engl J Med* **361**: 1475-1485.

Horrigan SK, Rich CB, Streeten BW, Li ZY, Foster JA. 1992. Characterization of an associated microfibril protein through recombinant DNA techniques. *J Biol Chem* **267**: 10087-10095.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44-57.

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

Huertas P, Garcia-Rubio ML, Wellinger RE, Luna R, Aguilera A. 2006. An hpr1 point mutation that impairs transcription and mRNP biogenesis without increasing recombination. *Mol Cell Biol* **26**: 7451-7465.

Hung ML, Hautbergue GM, Snijders AP, Dickman MJ, Wilson SA. 2010. Arginine methylation of REF/ALY promotes efficient handover of mRNA to TAP/NXF1. *Nucleic Acids Res* **38**: 3351-3361.

Huvet M, Nicolay S, Touchon M, Audit B, d'Aubenton-Carafa Y, Arneodo A, Thermes C. 2007. Human gene organization driven by the coordination of replication and transcription. *Genome Res* **17**: 1278-1285.

Icardi L, Mori R, Gesellchen V, Eyckerman S, De Cauwer L, Verhelst J, Vercauteren K, Saelens X, Meuleman P, Leroux-Roels G et al. 2012. The Sin3a repressor complex is a master regulator of STAT transcriptional activity. *Proc Natl Acad Sci U S A* **109**: 12058-12063.

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. 2003. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249-264.

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

Jackson MR, Melideo SL, Jorns MS. 2012. Human sulfide:quinone oxidoreductase catalyzes the first step in hydrogen sulfide metabolism and produces a sulfane sulfur metabolite. *Biochemistry* **51**: 6804-6815.

Jackson SP, Bartek J. 2009. The DNA-damage response in human biology and disease. *Nature* **461**: 1071-1078.

James P, Halladay J, Craig EA. 1996. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144**: 1425-1436.

Jang TH, Kim SH, Jeong JH, Kim S, Kim YG, Park HH. 2015. Crystal structure of caspase recruiting domain (CARD) of apoptosis repressor with CARD (ARC) and its implication in inhibition of apoptosis. *Sci Rep* **5**: 9847.

Jazayeri A, McAinsh AD, Jackson SP. 2004. Saccharomyces cerevisiae Sin3p facilitates DNA double-strand break repair. *Proc Natl Acad Sci U S A* **101**: 1644-1649.

Jeggo PA, Pearl LH, Carr AM. 2016. DNA repair, genome stability and cancer: a historical perspective. *Nat Rev Cancer* **16**: 35-42.

Jelinic P, Pellegrino J, David G. 2011. A novel mammalian complex containing Sin3B mitigates histone acetylation and RNA polymerase II progression within transcribed loci. *Mol Cell Biol* **31**: 54-62.

Jimeno S, Luna R, Garcia-Rubio M, Aguilera A. 2006. Tho1, a novel hnRNP, and Sub2 provide alternative pathways for mRNP biogenesis in yeast THO mutants. *Mol Cell Biol* **26**: 4387-4398.

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

Jinks-Robertson S, Bhagwat AS. 2014. Transcription-associated mutagenesis. *Annu Rev Genet* **48**: 341-359.

Johnson SA, Cubberley G, Bentley DL. 2009. Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11. *Mol Cell* **33**: 215-226.

Johnson SA, Kim H, Erickson B, Bentley DL. 2011. The export factor Yra1 modulates mRNA 3' end processing. *Nat Struct Mol Biol* **18**: 1164-1171.

Kadamb R, Mittal S, Bansal N, Batra H, Saluja D. 2013. Sin3: insight into its transcription regulatory functions. *Eur J Cell Biol* **92**: 237-246.

Kadosh D, Struhl K. 1997. Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd3 histone deacetylase to target promoters. *Cell* **89**: 365-371.

Katahira J. 2012. mRNA export and the TREX complex. *Biochim Biophys Acta* **1819**: 507-513.

Katahira J, Okuzaki D, Inoue H, Yoneda Y, Maehara K, Ohkawa Y. 2013. Human TREX component Thoc5 affects alternative polyadenylation site choice by recruiting mammalian cleavage factor I. *Nucleic Acids Res* **41**: 7060-7072.

Kaufmann I, Martin G, Friedlein A, Langen H, Keller W. 2004. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J* **23**: 616-626.

Khanna A, Stamm S. 2010. Regulation of alternative splicing by short non-coding nuclear RNAs. *RNA Biol* **7**: 480-485.

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

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

Kishore S, Khanna A, Zhang Z, Hui J, Balwierz PJ, Stefan M, Beach C, Nicholls RD, Zavolan M, Stamm S. 2010. The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Hum Mol Genet* **19**: 1153-1164.

Kito K, Yeh ET, Kamitani T. 2001. NUB1, a NEDD8-interacting protein, is induced by interferon and down-regulates the NEDD8 expression. *J Biol Chem* **276**: 20603-20609.

Kittler R, Putz G, Pelletier L, Poser I, Heninger AK, Drechsel D, Fischer S, Konstantinova I, Habermann B, Grabner H et al. 2004. An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature* **432**: 1036-1040.

Koffa MD, Clements JB, Izaurralde E, Wadd S, Wilson SA, Mattaj IW, Kuersten S. 2001. Herpes simplex virus ICP27 protein provides viral mRNAs with access to the cellular mRNA export pathway. *EMBO J* **20**: 5769-5778.

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

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

Kornblihtt AR, de la Mata M, Fededa JP, Munoz MJ, Nogues G. 2004. Multiple links between transcription and splicing. *RNA* **10**: 1489-1498.

Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ. 2013. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Biol* **14**: 153-165.

Kota KP, Wagner SR, Huerta E, Underwood JM, Nickerson JA. 2008. Binding of ATP to UAP56 is necessary for mRNA export. *J Cell Sci* **121**: 1526-1537.

Kumar R, Corbett MA, van Bon BW, Woenig JA, Weir L, Douglas E, Friend KL, Gardner A, Shaw M, Jolly LA et al. 2015. THOC2 Mutations Implicate mRNA-Export Pathway in X-Linked Intellectual Disability. *Am J Hum Genet* **97**: 302-310.

La Salle S, Palmer K, O'Brien M, Schimenti JC, Eppig J, Handel MA. 2012. Spata22, a novel vertebrate-specific gene, is required for meiotic progress in mouse germ cells. *Biol Reprod* **86**: 45.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685.

Lahue RS, Frizzell A. 2012. Histone deacetylase complexes as caretakers of genome stability. *Epigenetics* **7**: 806-810.

Lamond AI, Spector DL. 2003. Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol* **4**: 605-612.

Larochelle M, Lemay JF, Bachand F. 2012. The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. *Nucleic Acids Res* **40**: 10240-10253.

Leaw CL, Ren EC, Choong ML. 2004. Hcc-1 is a novel component of the nuclear matrix with growth inhibitory function. *Cell Mol Life Sci* **61**: 2264-2273.

Lee S, Chang J, Renvoise B, Tipirneni A, Yang S, Blackstone C. 2012. MITD1 is recruited to midbodies by ESCRT-III and participates in cytokinesis. *Mol Biol Cell* **23**: 4347-4361.

Lee SU, Maeda T. 2012. POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. *Immunol Rev* **247**: 107-119.

Lehner B, Semple JI, Brown SE, Counsell D, Campbell RD, Sanderson CM. 2004. Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics* **83**: 153-167.

Lenzken SC, Loffreda A, Barabino SM. 2013. RNA splicing: a new player in the DNA damage response. *Int J Cell Biol* **2013**: 153634.

Lewis MJ, Liu J, Libby EF, Lee M, Crawford NP, Hurst DR. 2016. SIN3A and SIN3B differentially regulate breast cancer metastasis. *Oncotarget* **7**: 78713-78725.

Li G, Reinberg D. 2011. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* **21**: 175-186.

Li J, Mahajan A, Tsai MD. 2006. Ankyrin repeat: a unique motif mediating protein-protein interactions. *Biochemistry* **45**: 15168-15178.

Li M, Xu X, Liu Y. 2011. The SET2-RPB1 interaction domain of human RECQ5 is important for transcription-associated genome stability. *Mol Cell Biol* **31**: 2090-2099.

Li X, Manley JL. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* **122**: 365-378.

Li Y, Lin AW, Zhang X, Wang Y, Wang X, Goodrich DW. 2007. Cancer cells and normal cells differ in their requirements for Thoc1. *Cancer Res* **67**: 6657-6664.

Li Y, Wang X, Zhang X, Goodrich DW. 2005. Human hHpr1/p84/Thoc1 regulates transcriptional elongation and physically links RNA polymerase II and RNA processing factors. *Mol Cell Biol* **25**: 4023-4033.

Lippert MJ, Kim N, Cho JE, Larson RP, Schoenly NE, O'Shea SH, Jinks-Robertson S. 2011. Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. *Proc Natl Acad Sci U S A* **108**: 698-703.

Lisby M, Rothstein R, Mortensen UH. 2001. Rad52 forms DNA repair and recombination centers during S phase. *Proc Natl Acad Sci U S A* **98**: 8276-8282.

Lischka P, Toth Z, Thomas M, Mueller R, Stamminger T. 2006. The UL69 transactivator protein of human cytomegalovirus interacts with DEXD/H-Box RNA helicase UAP56 to promote cytoplasmic accumulation of unspliced RNA. *Mol Cell Biol* **26**: 1631-1643.

Losada A, Hirano T. 2005. Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes Dev* **19**: 1269-1287.

Luna R, Gaillard H, Gonzalez-Aguilera C, Aguilera A. 2008. Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. *Chromosoma* **117**: 319-331.

Luna R, Jimeno S, Marin M, Huertas P, Garcia-Rubio M, 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 AG, 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 ML, Zhou Z, Magni K, Christoforides C, Rappsilber J, Mann M, Reed R. 2001. Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413**: 644-647.

Lybarger S, Beickman K, Brown V, Dembla-Rajpal N, Morey K, Seipelt R, Rymond BC. 1999. Elevated levels of a U4/U6.U5 snRNP-associated protein, Spp381p, rescue a mutant defective in spliceosome maturation. *Mol Cell Biol* **19**: 577-584.

Ma L, Gao X, Luo J, Huang L, Teng Y, Horvitz HR. 2012. The Caenorhabditis elegans gene mfap-1 encodes a nuclear protein that affects alternative splicing. *PLoS genetics* **8**: e1002827.

Madireddy A, Kosiyatrakul ST, Boisvert RA, Herrera-Moyano E, Garcia-Rubio ML, Gerhardt J, Vuono EA, Owen N, Yan Z, Olson S et al. 2016. FANCD2 Facilitates Replication through Common Fragile Sites. *Mol Cell* **64**: 388-404.

Maizels N. 2005. Immunoglobulin gene diversification. *Annu Rev Genet* **39**: 23-46.

Makarov EM, Makarova OV, Urlaub H, Gentzel M, Will CL, Wilm M, Luhrmann R. 2002. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**: 2205-2208.

Mancini A, Niemann-Seyde SC, Pankow R, El Bounkari O, Klebba-Farber S, Koch A, Jaworska E, Spooncer E, Gruber AD, Whetton AD et al. 2010. THOC5/FMIP, an mRNA export TREX complex protein, is essential for hematopoietic primitive cell survival in vivo. *BMC Biol* **8**: 1.

Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ. 2002. The human and mouse replication-dependent histone genes. *Genomics* **80**: 487-498.

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

Meinel DM, Burkert-Kautzsch C, Kieser A, O'Duibhir E, Siebert M, Mayer A, Cramer P, Soding J, Holstege FC, Strasser K. 2013. Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II. *PLoS genetics* **9**: e1003914.

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

Mischo HE, Gomez-Gonzalez B, Grzechnik P, Rondon AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell* **41**: 21-32.

Moldovan GL, D'Andrea AD. 2009. How the fanconi anemia pathway guards the genome. *Annu Rev Genet* **43**: 223-249.

Monji M, Nakatsura T, Senju S, Yoshitake Y, Sawatsubashi M, Shinohara M, Kageshita T, Ono T, Inokuchi A, Nishimura Y. 2004. Identification of a novel human cancer/testis antigen, KM-HN-1, recognized by cellular and humoral immune responses. *Clin Cancer Res* **10**: 6047-6057.

Moore MJ, Wang Q, Kennedy CJ, Silver PA. 2010. An alternative splicing network links cell-cycle control to apoptosis. *Cell* **142**: 625-636.

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

Nakayama KI, Nakayama K. 2006. Ubiquitin ligases: cell-cycle control and cancer. *Nat Rev Cancer* **6**: 369-381.

Nelson DM, Ye X, Hall C, Santos H, Ma T, Kao GD, Yen TJ, Harper JW, Adams PD. 2002. Coupling of DNA synthesis and histone synthesis in S phase independent of cyclin/cdk2 activity. *Mol Cell Biol* **22**: 7459-7472.

Nelson JR, Lawrence CW, Hinkle DC. 1996. Deoxycytidyl transferase activity of yeast REV1 protein. *Nature* **382**: 729-731.

Nishito Y, Hasegawa M, Inohara N, Nunez G. 2006. MEX is a testis-specific E3 ubiquitin ligase that promotes death receptor-induced apoptosis. *Biochem J* **396**: 411-417.

Ohno M, Shimura Y. 1996. A human RNA helicase-like protein, HRH1, facilitates nuclear export of spliced mRNA by releasing the RNA from the spliceosome. *Genes Dev* **10**: 997-1007.

Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**: 635-648.

Osinalde N, Olea M, Mitxelena J, Aloria K, Rodriguez JA, Fullaondo A, Arizmendi JM, Zubiaga AM. 2013. The nuclear protein ALY binds to and modulates the activity of transcription factor E2F2. *Mol Cell Proteomics* **12**: 1087-1098.

Papasaikas P, Tejedor JR, Vigevani L, Valcarcel J. 2015. Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol Cell* **57**: 7-22.

Papasaikas P, Valcarcel J. 2016. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci* **41**: 33-45.

Paulsen RD, Soni DV, Wollman R, Hahn AT, Yee MC, Guan A, Hesley JA, Miller SC, Cromwell EF, Solow-Cordero DE 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 ZP et al. 2015. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* **161**: 774-789.

Pellegrino J, Castrillon DH, David G. 2012. Chromatin associated Sin3A is essential for male germ cell lineage in the mouse. *Dev Biol* **369**: 349-355.

Perrod S, Cockell MM, Laroche T, Renauld H, Ducrest AL, Bonnard C, Gasser SM. 2001. A cytosolic NAD-dependent deacetylase, Hst2p, can modulate nucleolar and telomeric silencing in yeast. *EMBO J* **20**: 197-209.

Pile LA, Schlag EM, Wassarman DA. 2002. The SIN3/RPD3 deacetylase complex is essential for G(2) phase cell cycle progression and regulation of SMRTER corepressor levels. *Mol Cell Biol* **22**: 4965-4976.

Pile LA, Wassarman DA. 2000. Chromosomal localization links the SIN3-RPD3 complex to the regulation of chromatin condensation, histone acetylation and gene expression. *EMBO J* **19**: 6131-6140.

Piruat JI, 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.

Pommier Y, Sun Y, Huang SN, Nitiss JL. 2016. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol* **17**: 703-721.

Prado F, Aguilera A. 1995. Role of reciprocal exchange, one-ended invasion crossover and single-strand annealing on inverted and direct repeat recombination in yeast: different requirements for the RAD1, RAD10, and RAD52 genes. *Genetics* **139**: 109-123.

-. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J* **24**: 1267-1276.

Prado F, Piruat JI, Aguilera A. 1997. Recombination between DNA repeats in yeast hpr1delta cells is linked to transcription elongation. *EMBO J* **16**: 2826-2835.

Preker PJ, Lingner J, Minvielle-Sebastia L, Keller W. 1995. The FIP1 gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase. *Cell* **81**: 379-389.

Qian J, Wang Q, Dose M, Pruett N, Kieffer-Kwon KR, Resch W, Liang G, Tang Z, Mathe E, Benner C et al. 2014. B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* **159**: 1524-1537.

Rappsilber J, Ryder U, Lamond AI, Mann M. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**: 1231-1245.

Rasheva VI, Knight D, Bozko P, Marsh K, Frolov MV. 2006. Specific role of the SR protein splicing factor B52 in cell cycle control in Drosophila. *Mol Cell Biol* **26**: 3468-3477.

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

Rehwinkel J, Herold A, Gari K, Kocher T, Rode M, Ciccarelli FL, Wilm M, Izaurralde E. 2004. Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. *Nat Struct Mol Biol* **11**: 558-566.

Reid JL, Moqtaderi Z, Struhl K. 2004. Eaf3 regulates the global pattern of histone acetylation in Saccharomyces cerevisiae. *Mol Cell Biol* **24**: 757-764.

Ren L, McLean JR, Hazbun TR, Fields S, Vander Kooi C, Ohi MD, Gould KL. 2011. Systematic two-hybrid and comparative proteomic analyses reveal novel yeast pre-mRNA splicing factors connected to Prp19. *PLoS One* **6**: e16719.

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

Rondon AG, Jimeno S, Garcia-Rubio M, Aguilera A. 2003. Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. *J Biol Chem* **278**: 39037-39043.

Rougemaille M, Dieppois G, Kisseleva-Romanova E, Gudipati RK, Lemoine S, Blugeon C, Boulay J, Jensen TH, Stutz F, Devaux F et al. 2008. THO/Sub2p

functions to coordinate 3'-end processing with gene-nuclear pore association. *Cell* **135**: 308-321.

Ruepp MD, Aringhieri C, Vivarelli S, Cardinale S, Paro S, Schumperli D, Barabino SM. 2009. Mammalian pre-mRNA 3' end processing factor CF I m 68 functions in mRNA export. *Mol Biol Cell* **20**: 5211-5223.

Saguez C, Schmid M, Olesen JR, Ghazy MA, Qu X, Poulsen MB, Nasser T, Moore C, Jensen TH. 2008. Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. *Mol Cell* **31**: 91-103.

Saldi T, Cortazar MA, Sheridan RM, Bentley DL. 2016. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J Mol Biol* **428**: 2623-2635.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor laboratory Press.

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

Santos-Pereira JM, Herrero AB, Garcia-Rubio ML, Marin A, Moreno S, Aguilera A. 2013. The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes Dev* **27**: 2445-2458.

Sanz LA, Hartono SR, Lim YW, Steyaert S, Rajpurkar A, Ginno PA, Xu X, 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 GP, Heron M, Williams H, Soding J, Stewart A, Svejstrup JQ. 2014. RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell* **157**: 1037-1049.

Schmid M, Jensen TH. 2008. The exosome: a multipurpose RNA-decay machine. *Trends Biochem Sci* **33**: 501-510.

Selth LA, Sigurdsson S, Svejstrup JQ. 2010. Transcript Elongation by RNA Polymerase II. *Annu Rev Biochem* **79**: 271-293.

Serpeloni M, Moraes CB, Muniz JR, Motta MC, Ramos AS, Kessler RL, Inoue AH, daRocha WD, Yamada-Ogatta SF, Fragoso SP et al. 2011. An essential nuclear protein in trypanosomes is a component of mRNA transcription/export pathway. *PLoS One* **6**: e20730.

Shahbazian MD, Grunstein M. 2007. Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem* **76**: 75-100.

Sherman F, Fink GR, Lawrence CW. 1986. *Methods in Yeast Genetics*. Cold Spring Harbor, Cold Spring Harbord, NY.

Shiloh Y. 2003. ATM and related protein kinases: safeguarding genome integrity. *Nat Rev Cancer* **3**: 155-168.

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

Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* **122**: 19-27.

Silverstein RA, Ekwall K. 2005. Sin3: a flexible regulator of global gene expression and genome stability. *Curr Genet* **47**: 1-17.

Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* **516**: 436-439.

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

Smith KT, Martin-Brown SA, Florens L, Washburn MP, Workman JL. 2010. Deacetylase inhibitors dissociate the histone-targeting ING2 subunit from the Sin3 complex. *Chem Biol* **17**: 65-74.

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

Sollier J, Cimprich KA. 2015. Breaking bad: R-loops and genome integrity. *Trends Cell Biol* **25**: 514-522.

Sollier J, Stork CT, Garcia-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* **56**: 777-785.

Song F, Fan C, Wang X, Goodrich DW. 2013. The Thoc1 encoded ribonucleoprotein is a substrate for the NEDD4-1 E3 ubiquitin protein ligase. *PLoS One* **8**: e57995.

Sordet O, Redon CE, Guirouilh-Barbat J, Smith S, Solier S, Douarre C, Conti C, Nakamura AJ, Das BB, Nicolas E et al. 2009. Ataxia telangiectasia mutated activation by transcription- and topoisomerase I-induced DNA double-strand breaks. *EMBO Rep* **10**: 887-893.

Sorensen BB, Ehrnsberger HF, Esposito S, Pfab A, Bruckmann A, Hauptmann J, Meister G, Merkl R, Schubert T, Langst G et al. 2017. The Arabidopsis THO/TREX component TEX1 functionally interacts with MOS11 and modulates mRNA export and alternative splicing events. *Plant Mol Biol* **93**: 283-298.

Spiess C, Meyer AS, Reissmann S, Frydman J. 2004. Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends Cell Biol* **14**: 598-604.

Sridhara SC, Carvalho S, Grosso AR, Gallego-Paez LM, Carmo-Fonseca M, de Almeida SF. 2017. Transcription Dynamics Prevent RNA-Mediated Genomic Instability through SRPK2-Dependent DDX23 Phosphorylation. *Cell Rep* **18**: 334-343.

Stirling PC, Bloom MS, Solanki-Patil T, Smith S, Sipahimalani P, Li Z, Kofoed M, Ben-Aroya S, Myung K, Hieter P. 2011. The complete spectrum of yeast chromosome instability genes identifies candidate CIN cancer genes and functional roles for ASTRA complex components. *PLoS genetics* **7**: e1002057.

Stirling PC, Chan YA, Minaker SW, Aristizabal MJ, Barrett I, Sipahimalani P, Kobor MS, 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 AG, Aguilera A, Struhl K, Reed R et al. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**: 304-308.

Stutz F, Bachi A, Doerks T, Braun IC, Seraphin B, Wilm M, Bork P, Izaurralde E. 2000. REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6**: 638-650.

Su C, Gao G, Schneider S, Helt C, Weiss C, O'Reilly MA, Bohmann D, Zhao J. 2004. DNA damage induces downregulation of histone gene expression through the G1 checkpoint pathway. *EMBO J* **23**: 1133-1143.

Sulli G, Di Micco R, 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.

Sundaramoorthy S, Vazquez-Novelle MD, Lekomtsev S, Howell M, Petronczki M. 2014. Functional genomics identifies a requirement of pre-mRNA splicing factors for sister chromatid cohesion. *EMBO J* **33**: 2623-2642.

Taipale M, Kaminen N, Nopola-Hemmi J, Haltia T, Myllyluoma B, Lyytinen H, Muller K, Kaaranen M, Lindsberg PJ, Hannula-Jouppi K et al. 2003. A candidate gene for developmental dyslexia encodes a nuclear tetratricopeptide repeat domain protein dynamically regulated in brain. *Proc Natl Acad Sci U S A* **100**: 11553-11558.

Takahashi T, Burguiere-Slezak G, Van der Kemp PA, Boiteux S. 2011. Topoisomerase 1 provokes the formation of short deletions in repeated sequences upon high transcription in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **108**: 692-697.

Tamura T, Mancini A, Joos H, Koch A, Hakim C, Dumanski J, Weidner KM, Niemann H. 1999. FMIP, a novel Fms-interacting protein, affects granulocyte/macrophage differentiation. *Oncogene* **18**: 6488-6495.

Taniguchi I, Ohno M. 2008. ATP-dependent recruitment of export factor Aly/REF onto intronless mRNAs by RNA helicase UAP56. *Mol Cell Biol* **28**: 601-608.

Tejedor JR, Papasaikas P, Valcarcel J. 2015. Genome-wide identification of Fas/CD95 alternative splicing regulators reveals links with iron homeostasis. *Mol Cell* **57**: 23-38.

ten Asbroek AL, van Groenigen M, Nooij M, Baas F. 2002. The involvement of human ribonucleases H1 and H2 in the variation of response of cells to antisense phosphorothioate oligonucleotides. *Eur J Biochem* **269**: 583-592.

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

Tran DD, Saran S, Williamson AJ, Pierce A, Dittrich-Breiholz O, Wiehlmann L, Koch A, Whetton AD, Tamura T. 2014. THOC5 controls 3'end-processing of immediate early genes via interaction with polyadenylation specific factor 100 (CPSF100). *Nucleic Acids Res* **42**: 12249-12260.

Tresini M, Warmerdam DO, Kolovos P, Snijder L, Vrouwe MG, Demmers JA, van IWF, Grosveld FG, Medema RH, Hoeijmakers JH et al. 2015. The core spliceosome as target and effector of non-canonical ATM signalling. *Nature* **523**: 53-58.

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.

van Attikum H, Gasser SM. 2009. Crosstalk between histone modifications during the DNA damage response. *Trends Cell Biol* **19**: 207-217.

van Oevelen C, Bowman C, Pellegrino J, Asp P, Cheng J, Parisi F, Micsinai M, Kluger Y, Chu A, Blais A et al. 2010. The mammalian Sin3 proteins are required for muscle development and sarcomere specification. *Mol Cell Biol* **30**: 5686-5697.

van Oevelen C, Wang J, Asp P, Yan Q, Kaelin WG, Jr., Kluger Y, Dynlacht BD. 2008. A role for mammalian Sin3 in permanent gene silencing. *Mol Cell* **32**: 359-370.

Venkatesh S, Workman JL. 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol* **16**: 178-189.

Vermeulen W, Fousteri M. 2013. Mammalian transcription-coupled excision repair. *Cold Spring Harb Perspect Biol* **5**: a012625.

Viphakone N, Hautbergue GM, Walsh M, Chang CT, Holland A, Folco EG, Reed R, Wilson SA. 2012. TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nat Commun* **3**: 1006.

Wahba L, Amon JD, Koshland D, 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.

Wahl MC, Will CL, Luhrmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**: 701-718.

Wan Y, Zheng X, Chen H, Guo Y, Jiang H, He X, Zhu X, Zheng Y. 2015. Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing. *J Cell Biol* **209**: 235-246.

Wang X, Chang Y, Li Y, Zhang X, Goodrich DW. 2006. Thoc1/Hpr1/p84 is essential for early embryonic development in the mouse. *Mol Cell Biol* **26**: 4362-4367.

Wang X, Chinnam M, Wang J, Wang Y, Zhang X, Marcon E, Moens P, Goodrich DW. 2009a. Thoc1 deficiency compromises gene expression necessary for normal testis development in the mouse. *Mol Cell Biol* **29**: 2794-2803.

Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K. 2009b. Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**: 1019-1031.

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

Wellinger RE, Prado F, 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.

Wilkins BJ, Rall NA, Ostwal Y, Kruitwagen T, Hiragami-Hamada K, Winkler M, Barral Y, Fischle W, Neumann H. 2014. A cascade of histone modifications induces chromatin condensation in mitosis. *Science* **343**: 77-80.

Will CL, Luhrmann R. 2011. Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**.

Williams GT, Farzaneh F. 2012. Are snoRNAs and snoRNA host genes new players in cancer? *Nat Rev Cancer* **12**: 84-88.

Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, Helin K. 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**: 343-348.

Wu HY, Shyy SH, Wang JC, Liu LF. 1988. Transcription generates positively and negatively supercoiled domains in the template. *Cell* **53**: 433-440.

Yamaguchi Y, Wada T, Handa H. 1998. Interplay between positive and negative elongation factors: drawing a new view of DRB. *Genes Cells* **3**: 9-15.

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.

Yelina NE, Smith LM, Jones AM, Patel K, Kelly KA, Baulcombe DC. 2010. Putative Arabidopsis THO/TREX mRNA export complex is involved in transgene and endogenous siRNA biosynthesis. *Proc Natl Acad Sci U S A* **107**: 13948-13953.

Yoshida M, Kijima M, Akita M, Beppu T. 1990. Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *J Biol Chem* **265**: 17174-17179.

Yu K, Roy D, Bayramyan M, Haworth IS, Lieber MR. 2005. Fine-structure analysis of activation-induced deaminase accessibility to class switch region R-loops. *Mol Cell Biol* **25**: 1730-1736.

Zeman MK, Cimprich KA. 2014. Causes and consequences of replication stress. *Nat Cell Biol* **16**: 2-9.

Zhou Z, Licklider LJ, Gygi SP, Reed R. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature* **419**: 182-185.

Zhou Z, Luo MJ, Straesser K, Katahira J, Hurt E, Reed R. 2000. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**: 401-405.

Zilfou JT, Hoffman WH, Sank M, George DL, Murphy M. 2001. The corepressor mSin3a interacts with the proline-rich domain of p53 and protects p53 from proteasome-mediated degradation. *Mol Cell Biol* **21**: 3974-3985.