NEW FACTORS INVOLVED IN TRANSCRIPTION-ASSOCIATED GENOME INSTABILITY

José Antonio Mérida Cerro

Tesis Doctoral

Universidad de Sevilla
2019

New factors involved in transcription-associated genome instability

Trabajo realizado en el Departamento de Genética, Facultad de Biología (Universidad de Sevilla) y en el departamento de Biología del Genoma, CABIMER (Universidad de Sevilla-CSIC-UPO), con el fin de optar al grado de de Doctor en Biología Molecular, Biomedicina e Investigación Clínica por el licenciado José Antonio Mérida Cerro.

Sevilla, 2019

El doctorando:

José Antonio Mérida Cerro

Los directores de tesis:

Andrés Aguilera López

Ana Beatriz García Rondón

«No os preocupéis demasiado esta noche pensando en el camino. Pues los caminos que seguiréis todos vosotros ya se extienden quizás a vuestros pies, aunque no los veáis aun».

J. R. R. Tolkien

INDEX OF CONTENTS

INDEX OF FIGURES AND TABLES

ABBREVIATIONS AND ACRONYMS

5-FOA	5-fluorotic acid.
AID	Activation-induced cytidine deaminase / Auxin-induced Degron.
ARS	Autonomously replicating sequence.
ATM	Ataxia telangiectasia mutated.
A.U.	Arbitrary units.
BER	Base excision repair.
BIR	Break-induced replication.
bp	Base pair.
cDNA	Complementary DNA.
ChIP	Chromatin immunoprecipitation.
CPF-CF	Cleavage and polyadenylation factor dependent termination.
CPT	Camptothecin.
CTD	Carboxy-terminal domain.

CSA	Cockayne syndrome group A.
CSB	Cockayne syndrome group B.
DDR	DNA damage response.
DEPC	Diethyl pyrocarbonate.
DNA	Deoxyribonucleic acid.
DNAPK	DNA-dependent protein kinase.
DOX	Doxycyclin.
DRIP	DNA:RNA hybrid immunoprecipitation.
DSIF	DRB sensitivity inducing factor.
DSB	Double-strand break.
DSBR	Double-strand break repair.
EMSA	Electrophoretic mobility shift assay.
FACS	Fluorescence-activated cell sorting.
FISH	Fluorescence in situ hybridization.
FRT	Flipase recognition target.
Gal	Galactose.
GFP	Green fluorescent protein.
GGR	Global genome repair.
Glu	Glucose.
GTF	General transcription factor.
HA	Hemagglutinin.
HR	Homologous recombination.
HU	Hydroxyurea.
Hyg	Hygromycin.
IF	Immunofluorescence.
Kan	Kanamycin.
kb	Kilobase.
MMR	Mismatch repair.
MMS	Monomethyl sulfate.
mRNA	Messenger RNA.
mRNP	Messenger ribonucleoparticle.
NAA	1-Naphthaleneacetic acid.
Nat	Nourseothricin.
ncRNA	Non-coding RNA.
NER	Nucleotide excision repair.
NHEJ	Non-homologous end joining.
NNS	Nrd1-Nab3-Sen1 dependent termination.
ORF	Open reading frame.

pA	Polyadenylation.
PCR	Polymerase chain reaction.
PIC	Pre-initiation complex.
PMSF	Phenylmethanesulfonyl fluoride.
qPCR	Quantitative polymerase chain reaction.
RBP	RNA binding protein.
rDNA	Ribosomal DNA.
RNA	Ribonucleic acid.
RNAPII	RNA polymerase II.
RNase	Ribonuclease.
RPA	Replication protein A.
rRNA	Ribosomal RNA.
RRM	RNA-recognition motif.
SC	Synthetic complete medium.
SD	Synthetic defined medium.
SDSA	Synthesis-dependent strand annealing.
SPO	Sporulation medium.
siRNA	Small interfering RNA.
snRNA	Small nucleolar RNA.
SSA	Single-strand annealing.
SSB	Single-strand break.
ssDNA	Single-stranded DNA.
TAR	Transcription-associated recombination.
TCR	Transcription-coupled repair.
TRC	Transcription-replication conflicts.
TSS	Transcription start site.
UTR	Untranslated region.
UV	Ultraviolet.
WT	Wild type.
YFP	Yellow fluorescent protein.

RESUMEN

El mantenimiento de la integridad del DNA y la transmisión fiel de su información a la descendencia es una prioridad para los seres vivos. Al mismo tiempo, la molécula de DNA es el sustrato de numerosas reacciones vitales para la célula, como la replicación o la transcripción, las cuales pueden suponer una fuente de daño cuando se desregulan o entran en conflicto con otros procesos. Esto puede producir alteraciones genéticas que incluyen la pérdida de información, mutaciones y reordenaciones cromosómicas en un proceso conocido como inestabilidad genómica. La transcripción es uno de los procesos centrales en el metabolismo del DNA que puede generar inestabilidad genómica en determinadas circunstancias. La transcripción está acoplada al procesamiento del transcrito y desemboca en la producción de una ribonucleopartícula mensajera (mRNP) apta para ser exportada al citoplasma. Se ha demostrado que la ausencia de determinados factores de ensamblaje de la mRNP produce fallos en la elongación de la transcripción, generalmente asociados a la acumulación de híbridos de DNA:RNA y que pueden dar lugar a colisiones entre las maquinarias de transcripción y replicación, mutaciones y daño en el DNA. Estos híbridos de DNA:RNA se forman cuando el transcrito naciente hibrida con la hebra molde del DNA, dando lugar a un dúplex de DNA:RNA y a una hebra de cadena sencilla de DNA en una estructura conocida como bucle R (*R-loop*). Si bien estas estructuras se forman naturalmente, la acumulación de *R-loops* es una marca de inestabilidad genómica.

El objetivo de esta tesis es estudiar nuevos factores que puedan generar inestabilidad genómica asociada a híbridos de DNA:RNA mediante defectos en el ensamblaje de la mRNP, así como entender el efecto que tienen la acumulación de *R-loops* y otras fuentes de daño poco estudiadas, como las roturas de cadena sencilla (SSBs), sobre la transcripción y la propia RNA polimerasa II (RNAPII).

Recientemente se ha demostrado que no solo la ausencia de factores de ensamblaje de la mRNP puede generar inestabilidad genómica. Un exceso de la proteína Yra1, que participa en la formación y exportación de la mRNP, es capaz de producir inestabilidad genómica dependiente de *R-loops*. En un primer capítulo de esta tesis, y empleando *Saccharomyces cerevisiae* como organismo modelo, buscamos otros genes cuya sobreexpresión pudiera causar inestabilidad genómica asociada a híbridos de DNA:RNA. Para ello, realizamos un escrutinio de sobreexpresión en mutantes *hpr1* que acumulan *R-loops*, identificando aquellos genes cuya sobreexpresión produjera una reducción en el crecimiento como indicativo de un aumento en la cantidad de híbridos de DNA:RNA. Mediante esta aproximación hemos identificado tres

proteínas de unión al RNA cuyo exceso produce un aumento de daño en el DNA que es dependiente de *R-loops*: Dis3, la unidad catalítica esencial del exosoma; She2, implicada en el transporte al citoplasma de ciertos mRNAs; y Rie1, una proteína de unión al RNA de función desconocida.

Hemos mostrado que la sobrexpresión de *DIS3* posiblemente produce su agregación y por tanto la reducción de los niveles de esta proteína disponibles para formar el exosoma, causando fenotipos similares a los de la pérdida de función de esta proteína: Aumento del daño en el DNA dependiente de híbrido y defectos en el procesamiento del RNA ribosómico. La sobreexpresión de *DIS3* da lugar a un ligero aumento de híbridos a nivel global en el núcleo, como sucede en el mutante condicional de *dis3*, probablemente debido a que el exceso de Dis3 produce una desregulación de la transcripción que conlleva la expresión de RNA aberrantes, antisentido y no codificantes que no son degradados. Por otra parte, la sobreexpresión de *SHE2* produce una acumulación de *R-loops* en genes que codifican para RNAs con estructuras en forma de tallo-lazo (stem-loop), posiblemente porque el exceso de She2 favorece la re-hibridación del transcrito con el DNA, o estabiliza los híbridos de DNA:RNA. Por último, hemos mostrado cómo la sobrexpresión de *RIE1* provoca la entrada de su proteína, que es citoplasmática a niveles endógenos, en el núcleo. Aquí, Rie1 produce una fuerte acumulación de *R-loops*, posiblemente debido a la generación de defectos en el procesamiento de los ARNs de manera directa, por unión a RNA, o bien por mediación de otros factores.

En el segundo capítulo de la tesis, hemos diseñado una serie de sistemas moleculares que permiten controlar la expresión del gen *LYS2* de la levadura al tiempo que inducimos la formación de *R-loops* o de cortes de cadena sencilla (SSBs) en dicho gen, permitiéndonos estudiar sus efectos sobre la transcripción. Hemos determinado que la formación de híbridos de DNA:RNA a niveles fisiológicos bloquean transitoriamente el avance de la RNAPII, produciendo una disminución en la tasa de elongación. Esta RNAPII bloqueada podría ser eliminada del DNA, ya que los niveles de transcrito disminuyen en presencia de *R-loops* y observamos una caída drástica de la cantidad de RNAPII durante cinéticas de elongación. Por otra parte, la inducción de SSBs en la cadena molde (pero no en la cadena no-molde) producen una acumulación de RNAPII corriente arriba del sitio dañado, generando un bloqueo que puede causar la eliminación de la RNAPII del DNA, como sugiere la disminución de transcrito. Finalmente, empleando estos sistemas para inducir la formación de *R-loops* y SSBs, diseñamos unos sistemas genéticos que nos permiten medir recombinación entre cromosomas homólogos. Gracias a estos sistemas hemos observado que los *R-loops* son una fuente débil de daño en el DNA en comparación a los SSBs y nos permitirán determinar qué factores son necesarios para

los procesos de reparación, transcripción o eliminación de la RNAPII bloqueada por a *R-loops* o SSBs.

1. INTRODUCTION

The DNA is the genetic material encoding the information required for life as it has the ability to faithfully duplicate and transmit this information to the daughter cells. In eukaryotes, the information encoded into the DNA is transcribed into RNA molecules that are processed and packaged into messenger ribonucleoparticles (mRNPs), that are exported to the cytoplasm and translated into proteins with a plethora of different functions. An important group of RNAs do not encode for proteins, yet they have important regulatory functions modulating gene expression, or they are part of macromolecular complexes like the ribosome or the spliceosome. DNA in the cell is coated with different proteins, mainly histones, in a structure called chromatin. The degree of chromatin compaction varies through the cell cycle and according to the processes that are taking place in the DNA, as transcription.

The DNA is susceptible of being damaged by certain metabolic cell products but also by the processes that take place on it, like replication or transcription. In addition to these endogenous sources of DNA damage, exogenous agents as chemical compounds, ionizing radiation or ultraviolet light (UV) also alter the DNA molecule, compromising the hereditary information and negatively affecting transcription and replication. Improper DNA damage signalling or repair could lead to loss of genetic data. We refer to all these types of different alterations in DNA, from single nucleotide changes to gross chromosomal rearrangements, as genome instability, that is a pathological mark and a hallmark of cancer.

1.1. Eukaryotic transcription is a highly coordinated process coupled to mRNA processing

Gene expression is a highly regulated process that comprises several steps, from changes in chromatin that allow the assembly of the transcription machinery, to the synthesis, processing and export of the mRNA to the cytoplasm. All these processes are tightly coupled and changes or failures in any of them could lead to an incorrect transcription, formation of non-B DNA structures, mutations and eventually to genome instability (Aguilera & García-Muse 2012). In eukaryotes, transcription is performed by three different RNA polymerases specialized in different products: RNAPI focuses on ribosomal RNA synthesis; while RNAPII transcribes the DNA encoding messenger RNAs (mRNAs), small nuclear/nucleolar RNAs (sn/snoRNAs) and non-coding RNAs

(ncRNAs); and RNAPIII synthetizes transfer RNA (tRNA) and the 5S rRNA subunit (Cramer *et al.*, 2008).

RNA polymerase II (RNAPII) is an enzymatic complex of 12 subunits in yeast and human cells (Rpb1 to Rpb12). Rpb1 is the largest subunit of RNAPII that, when it is ubiquitylated and targeted by the proteasome for degradation, removes the entire RNAPII holoenzyme from the chromatin. It contains a carboxy-terminal domain (CTD), composed by tandem repeats of a consensus sequence: $Y_1S_2P_3T_4S_5P_6S_7$. This aminoacidic heptad is repeated 26 times in yeast and 52 in mammal cells. The CTD is highly conserved in all the eukaryotes and it is essential for RNAPII function, acting as a scaffold that allow the polymerase to interact with different transcription, splicing, mRNA processing factors and chromatin remodelers. The serine (S), threonine (T) and Tyrosine (Y) residues can be phosphorylated, while the two prolines (P) can undergo isomerization. The different combinations of these modifications compose a code that coordinates the recruitment of the different factors at the precise time of the transcription cycle (Heidermann *et al.*, 2013). Although RNAPII is able to unwind the double-stranded DNA and polymerize RNA, it needs additional factors in order to form initiation and elongation complexes (Cramer *et al.*, 2000).

Transcription is conceptually divided in three main steps: initiation, elongation and termination; but there are several processes coupled to these steps, as RNA 5'-end capping, splicing, mRNA folding, mRNP assembly, RNA 3'-end cleave and polyadenylation, mRNA quality control and mature mRNP export (Vinciguerra & Stutz 2004). During initiation, the core RNAPII assembles with general transcription factors (GTFs) in a promoter consensus sequence, forming the close pre-initiation complex (PIC) upstream of the transcription start site (TSS). In the presence of ribonucleotide triphosphates, TFIIH melts DNA, providing a ssDNA template that is positioned in the RNAPII catalytic centre, permitting the RNA chain synthesis and constituting the open PIC (Sainsbury *et al.*, 2015; Shandilya & Roberts 2012). After that, Kin28/Cdk7, a kinase of the TFIIH complex, phosphorylates Ser5 and Ser7 residues of RNAPII CTD, triggering promoter escape and the recruitment of different initiation and early transcription factors (Akhtar *et al.*, 2009), starting RNAPII elongation. When the nascent transcript reaches a certain length (about 25 nt), a stable elongation complex is formed and the initiation factors are released (Shandilya & Roberts 2012). As soon as the nascent RNA exits

RNAPII, a 7-methylguanosine cap is added to the 5'-end of the RNA, in a CTD Ser5 phosphorylation-dependent manner (Zorio & Bentley 2004).

Downstream TSS, Ser2 CTD residue is phosphorylated by Cdk1 and Bur1, (Cdk9 in humans) releasing RNAPII from proximal pausing and allowing the recruitment of additional elongation and splicing factors (Heidemann *et al.*, 2013), while Tyr1 phosphorylation impedes the premature interaction with 3'-end processing and export factors (Mayer *et al.*, 2012). During elongation, the RNAPII repeats a nucleotide addition cycle in which it adds a ribonucleotide complementary to the template DNA to the nascent mRNA by the formation of a phosphodiester bond. Elongation by yeast RNAPII has an average rate of 1.2 Kb/min, but the speed is not uniform, greatly depending on the sequence transcribed or the cell cycle stage (Palangat & Larson 2012). Moreover, the presence of damage, obstacles (for example, torsional stress) or misincorporation of ribonucleotides could pause, stall or backtrack the RNAPII. When this occurs, elongation factors like the Spt4/Spt5 (DSIF) complex or Dst1 (TFIIS) are required to resume transcription (Cheung & Cramer 2012). mRNA splicing takes place co-transcriptionally during elongation. Thus, splicing may affect the elongation rate and the other way around, the elongation rate could affect splicing, confirming the close relation between both processes (Zorio & Bentley 2004). The assembly of RNA-binding proteins to the nascent mRNA is also cotranscriptional and may influence transcription efficiency, as reported for some mRNP assembly mutants (Rondón *et al.*, 2003; Chávez *et al.*, 2001).

Termination implies the release of the RNAPII and the mRNA from the DNA template. It may occur in different manners, comprising two main routes: the polyadenylation (pA) dependent or the Nrd1-Nab3-Sen1 (NNS) dependent pathway. The pA-mediated termination depends on the recognition of polyadenylation signals in the 3' untranslated region (UTR) of the mRNA by the cleavage and polyadenylation factors (CPF) loaded in RNAPII CTD. Dephosphorylation of CTD Tyr1 residue allows its interaction with the Ser2 phosphorylated residues. The target RNA sequence is then cleaved at the poly(A) site and adenosine nucleotides are added to the 3'-end of the transcript by Pap1 polymerase. Different proteins bind to this poly(A) tail, protecting it and preparing the mRNA for its export to the cytoplasm. Simultaneously, Rat1, a 5'-3' exonuclease, targets the unprotected 5'-end of the RNA still attached to RNAPII and degrades it. This, together with the conformational changes occurring at RNAPII mediated by the CPFs interaction with the CTD and the nascent RNA, dissociates the

elongation complex. On the other hand, the NNS-dependent pathway takes place mainly in non-coding RNAs (ncRNAs), cryptic unstable transcripts and small-nucleolar or nuclear RNAs (snoRNAs or snRNAs) (Arigo *et al.*, 2006). NNS-dependent termination implicates recognition of UGUAG and UCUUGU sequence motifs in the RNA by Nrd1 and Nab3 (Creamer *et al.*, 2011). These proteins target the RNA to the exosome immediately after its release. The exosome is a complex with 3'-5' exonuclease activity that, aided by TRAMP, degrades ncRNAs or trims sn/snoRNAs and rRNA to its final length (LaCava *et al.*, 2005). NNS complex interaction with RNAPII is partially dependent on CTD Ser5 phosphorylation, indicating that this pathway takes place preferentially in short genes and predominantly inside the gene body, as the levels of Ser5-P are quickly reduced during elongation (Vasiljeva *et al.*, 2008; Porrua & Libri 2015). However, NNS-dependent termination could also take place as a fail-safe mechanism in genes with weak polyadenylation signals or during the lack of some termination factors like Rat1, avoiding RNAPII readthrough (Rondón *et al.*, 2009). Finally, the stable binding of a protein into the DNA is also a mechanism used by the cell to terminate transcription. For example, Reb1, Rap1 or Abf1 proteins can terminate transcription by roadblock-induced stalling of the elongation complex and ubiquitylation-mediated removal of RNAPII, while transcripts are subsequently degraded by the TRAMP-exosome complex using the NNS termination pathway (Colin *et al.*, 2014). All these different but redundant proteins act limiting pervasive readthrough and favouring transcription insulation (Candelli *et al.*, 2018).

1.2. mRNP biogenesis

Naked mRNAs are quickly degraded and can not be exported from the nucleus. In order to increase their stability and to be exported through the nuclear pore, the mRNAs are coated with different RNA-binding proteins (RBPs) in complexes termed messenger ribonucleoparticles (mRNPs). The assembly of the mRNP is co-transcriptional and starts as soon as the nascent RNA exits the RNAPII, with the attachment of different RNA-processing and RNA-binding proteins to the transcript, impeding its re-hybridization with the template DNA. The composition of the mRNP is dynamic, changing in time depending on the processes that take place, as the factors bound are key components of mRNA maturation, folding, quality control and export (Björk & Wieslander 2017).

During mRNP biosynthesis, RBPs interact with the mRNA in different manners: recognizing specific short degenerated sequences through their RNA recognition motifs (RRMs); binding to the poly(A) site or to consensus sequences in the transcripts; binding to sequence-independent secondary or tertiary RNA structures, or interacting with other proteins bounded to the RNA (Björk & Wieslander 2017). This is the case of the THO complex, which interacts with the RNAPII CTD (Meinel *et al.*, 2013) and with the nascent RNA to establish a scaffold for multiple RBPs to bind to the mRNA (Luna *et al.*, 2012). In yeast, the THO complex includes five proteins: Tho2, Hpr1, Mft1, Thp2 and Tex1; while the human contains six subunits: THOC1, THOC2, THOC5, THOC6, THOC7 and TEX1. This core complex could associate with Sub2/UAP56 and Yra1/ALY and other proteins to form the TREX complex, that interacts with Mex67 and other mRNA export factors, hence intimately relating mRNP assembly with export. Mutations in the core components of the THO complex cause pleiotropic phenotypes: transcription impairment especially in long, GC-rich genes, defects in 3'-end mRNA formation and export and genomic instability (Luna *et al.*, 2012). Failure in other steps of mRNP biogenesis, as mRNA splicing or mRNA cleavage and polyadenylation, could also be a source of genomic instability as it was reported for the depletion of SRSF1 splicing factor in human, or the yeast cleavage and polyadenylation mutant *fip1* (Li & Manley 2005; Stirling *et al.*, 2012). In all the cases mentioned earlier, genome instability is caused by the accumulation of R-loops.

1.3. R-loops as byproducts of transcription

R-loops are non-B DNA structures formed by an DNA:RNA hybrid in a Watson-Crick double helix and an a displaced single-stranded DNA (Crossley *et al.*, 2019). They may occur physiologically in the genome due to accumulation of negative supercoiling in transcribed genes, a high GC content in the DNA, the presence of single-strand breaks (SSBs) in the template or the formation of G-quadruplexes in the non-template strand; situations that favour the hybridization of the nascent mRNA with the template DNA (Belotserkovskii *et al.*, 2018). R-loops have a role in transcription regulation. DRIP-seq analysis (in which the RNA of the DNA:RNA hybrids are isolated and sequenced) have shown that R-loops localize at promoter regions of genes, upstream TSS, protecting the promoters from DNA methylation and silencing (Ginno *et al.*, 2012). An alternative mechanism to induce transcription was described for R-loops formed by antisense

transcription of ncRNAs from the promoters. They enhance chromatin opening and the binding of transcription factors to the promoter to induce sense transcription (Boque-Sastre *et al.*, 2015). Other evidences have related R-loop formation in termination regions with an active role in this process. More in detail, some termination events rely in the presence of G-rich sequences located after the poly(A) sequence, that pause the RNAPII probably by the formation of an R-loop (Skourti-Stathaki *et al.*, 2011). This would help Rat1 to reach the elongating polymerase, promoting termination. Sen1 cooperates with Rat1 in this termination pathway, probably because the helicase activity of Sen1 is necessary to unwind the hybrid so that Rat1 could reach RNAPII and release the elongation complex (Skourti-Stathaki *et al.*, 2011; Rosonina *et al.*, 2006). In human cells, the presence of R-loops in the termination regions establishes a H3K9me2 heterochromatin mark that facilitates RNAPII pausing prior to termination (Skourti-Stathaki *et al.*, 2014). R-loops also have a great physiological relevance for the cell during mitochondrial DNA replication, where short DNA:RNA hybrids form at the replication origin to prime DNA synthesis. In mammalian cells, during immunoglobulin class-switch recombination, formation of co-transcriptional R-loops favours mutation of the displaced ssDNA strand by the activated-induced cytidine deaminase (AID), leading to directed chromosomal rearrangements (Pavri 2017; Aguilera & García-Muse 2012). Finally, R-loops also have a role in telomere maintenance, as their accumulation in cells that lack some of the components of the THO complex or RNase H, that degrades the RNA moiety of the DNA:RNA hybrids, leads to telomere lengthening (Luke *et al.*, 2008).

Taking apart the physiological role of DNA:RNA hybrids, their accumulation in the cell is a source of genomic instability. This was first demonstrated in the THO complex *hpr1* mutant, whose transcription-associated hyperrecombination was suppressed by RNase H overexpression (Huertas & Aguilera 2003), greatly suggesting that defects in mRNP assembly, as those produced in the THO complex mutants, facilitate the re-hybridation of the nascent RNA with the template DNA, increasing R-loops beyond physiological levels. The displacement of one strand of DNA in the R-loop provides a substrate for DNA-modifying enzymes like AID that acts on ssDNA but also for spontaneous mutations and endonucleases, but we still unknown mechanisms for R-loops to generate DSBs. Thus R-loops induce hyperrecombination even in wild type conditions, as the translocations produced between immunoglobulin and c-myc regions that are responsible for Burkytt's lymphoma (Ramiro *et al.*, 2004; Aguilera & García-

Muse 2012). Recent studies have shown that probably R-loops are not deleterious by themselves. The persistence of DNA:RNA hybrids, like the produced by RNase H or different RNA biogenesis mutants, trigger chromatin compaction mediated by histone H3S10 phosphorylation (Castellano-Pozo et al 2013; García-Pichardo *et al.*, 2017). The prevalent idea is that a more compacted chromatin would establish barriers for a proper replication progression, generating DNA damage (Castellano-Pozo *et al.*, 2013; García-Pichardo *et al.*, 2017; García-Rondón & Aguilera 2019; Gómez-González & Aguilera 2019).

To avoid a pathological accumulation of R-loops, cells present different mechanisms. The previously mentioned RNase H enzymes or the helicases can remove directly the DNA:RNA hybrids. There are two classes of RNase H: RNase H1 and RNase H2. Both degrade DNA:RNA hybrids in addition to other specialized substrates in the cell, removing RNA primers of the mitochondrial DNA replication in the case of RNase H1 or cleaving ribonucleotides misincorporated to the DNA duplex (Cerritelli & Crouch 2009; Wahba *et al.*, 2011). Several DNA:RNA helicases can unwind hybrids and limit their formation. Is the case of Sen1/SETX, that restricts the natural occurrence of co-transcriptional R-loops, but also has a role in transcription termination and in resolution of colliding replication forks (Mischo *et al.*, 2011; Costantino & Koshland 2018);

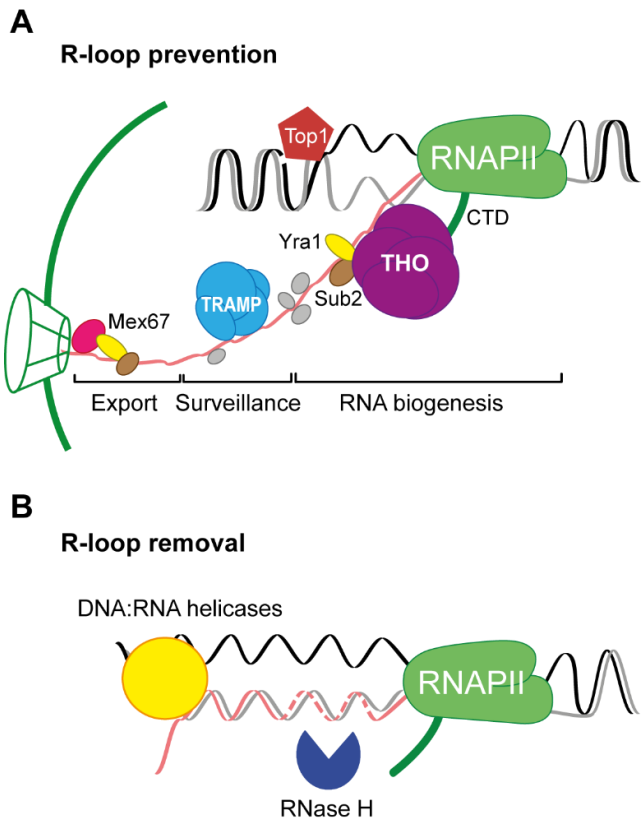


Figure I1. Mechanisms to prevent and remove R-loops.
(A) R-loop accumulation is prevented by specific mRNP assembly and export proteins that impede the re-hybridation of the nascent RNA with the DNA template. Top1 releases the negative supercoiling behind RNAPII that could facilitate R-loop formation. **(B)** the RNA moiety of DNA:RNA hybrids can be degraded by RNase H enzymes. Different DNA:RNA helicases, as Sen1 or Pif1, can resolve R-loops unwinding the hybrid. Adapted from Santos-Pereira & Aguilera 2015.

Pif1/PIF1, a conserved helicase that is involved in mitochondrial DNA maintenance, rDNA replication and telomeric DNA synthesis (Tran *et al.*, 2017); or the human DDX19, that resolves DNA:RNA hybrids during transcription-replication collisions (Hodroj *et al.*, 2017), among others. Another mechanism used by the cell to prevent R-loop formation is avoiding the accumulation of DNA negative supercoiling during transcription by topoisomerase enzymes (TOP). Indeed, in TOP1-deficient human cells RNase H sensitive DNA breaks increase in transcribed genes, demonstrating that negative supercoiling of the DNA favours R-loop formation and the role of Top1 at preventing it (Tuduri *et al.*, 2009). In addition, as we early mentioned, the proper mRNP assembly prevents the re-hybridation of the nascent mRNA with the template strand, avoiding R-loop formation. Finally, chromatin also has an important role in preventing DNA:RNA hybrid accumulation as shown by the increase in R-loops detected in mutants of the histone chaperone FACT (Herrera-Moyano *et al*., 2014), in human cells depleted for Sin3A (Salas-Armenteros *et al.*, 2017) or in some histone H3 and H4 mutants (García-Pichardo *et al.*, 2017). All these processes take place concurrently in the cell to avoid R-loop accumulation and formation in regions where they may produce a deleterious effect for the DNA metabolism.

1.4. *YRA1* overexpression increases R-loop-mediated genome instability

Not only the absence of mRNP proteins affects mRNA metabolism and therefore increases R-loop formation but also their excess. Yra1 is an essential RNA-binding protein that interacts with RNAPII CTD and with other mRNP factors, like the THO complex or Sub2, acting as an adaptor for mRNA-export proteins, like Mex67, contributing to the formation of an export-competent mRNP (Strasser & Hurt 2000). Yra1 contains an RNA-binding domain (RBD/RRM) but also two conserved domains, REF-N and REF-C, that are required for its interaction with the RNA (MacKellar & Greenleaf 2011). The levels of Yra1 protein in the cell are tightly regulated through a negative feedback mechanism involving the splicing of its intron. If Yra1 intron is artificially eliminated from the gene, this regulation is bypassed and *YRA1* is overexpressed (Rodríguez-Navarro *et al.*, 2002), leading to a negative effect on mRNA export, hyperrecombination and a strong growth inhibition. Recently, we observed that overabundance of Yra1 in yeast increases its recruitment to R-loop prone sequences, and causes accumulation of DNA:RNA hybrids in those regions, inducing

hyperrecombination that is reduced with RNase H. Yra1 binds to R-loops *in vitro* and presumably the excess of the protein stabilizes R-loops and promotes transcription-replication collisions, inducing DNA damage together with all the phenotypes described (García-Rubio *et al.*, 2018; Gavaldá *et al.*, 2016).

1.5. The impact of DNA damage on transcription

Not only does transcription damages the DNA, but DNA damage also affects the transcription process. Pre-existing damage in the DNA may impair transcription when an elongating RNA polymerase encounters a lesion in the template. Different outcomes could result depending on the type of damage. DNA lesions that do not significantly distort the DNA backbone may not block transcription elongation but promote RNA polymerase pausing and error-prone elongation, bypassing the damage. That is the case of the oxidative damage 8-oxo-2′-deoxyguanosine (8-oxo-dG), where RNAPII can either insert a matched cytosine or a mismatched adenine, introducing a mutation in the transcript but bypassing the lesion (Kitsera *et al.*, 2011). On the contrary, bulky lesions that significantly change the DNA helix structure, as double-strand breaks (DSBs) or pyrimidine dimers, are able to stall RNAPII (Shanbhag *et al.*, 2010; Woudstra *et al.*, 2002; Pankotai *et al.*, 2012). Single-stranded DNA breaks (SSBs), one of the most common forms of DNA damage, reduce the amount of transcript *in vitro* when they localize in the template strand of the DNA (Zhou & Doetsch 1993; Kathe *et al.*, 2004; Neil *et al.*, 2012). SSBs are naturally produced in the cell during elongation by topoisomerase I (Top1) to release topological stress. An erroneous or abortive activity of Top1, in which the cleavage complex intermediate that forms stays longer in the DNA or collides with the transcription machinery, could lead to the formation of TOP1-linked SSBs. In this situation Top1 remains covalently bound to the nick, arresting transcription (Desai *et al.*, 2003; Pommier *et al.*, 2003). Finally, little is known on how R-loops affect transcription. *In vitro* experiments suggest that R-loops impair transcription elongation. First, a blocked T7 RNA polymerase at the proximity of a G-rich promoter could be rescued by RNase H overexpression (Belotserkovskii *et al.*, 2017). While the presence of preformed RNA:DNA hybrids in the template or the transcription of an R-loop prone sequence reduce the RNA outcome (Tous & Aguilera, 2007; Tornaletti *et al.*, 2008).

When RNAPII is paused or arrested by a DNA damage, it needs to backtrack to resume elongation (Awrey *et al.*, 1997), displacing the nascent transcript from the polymerase catalytic site. TFIIS, Dst1 in yeast, triggers the hydrolysis of the 3'-end of the nascent RNA to position it within the active site of the RNAPII. RNAPII backtracking permits the access of the repair machineries to the damage site without removing the elongation complex. The addition of TFIIS inhibitors negatively affects nucleotide excision repair (NER), the yeast *dst1* mutant is lethal and its transient depletion in human cells greatly increases the amount of polyubiquitylated RNAPII, consistent with RNAPII backtracking being a general mechanism to rescue compromised elongation complexes (Sigurdsson *et al.*, 2010). Finally, if transcription can not be resumed, RNAPII is poly-ubiquitinated and degraded by the proteasome, to allow the repair of the lesion (Ratner *et al.*, 1998). In this process, Rpb1 is initially monoubiquitylated by the E3 ubiquitin ligase Rsp5/NEDD4 in the Lys-63 residue (Huibregtse *et al.*, 1997). This mark is recognized by

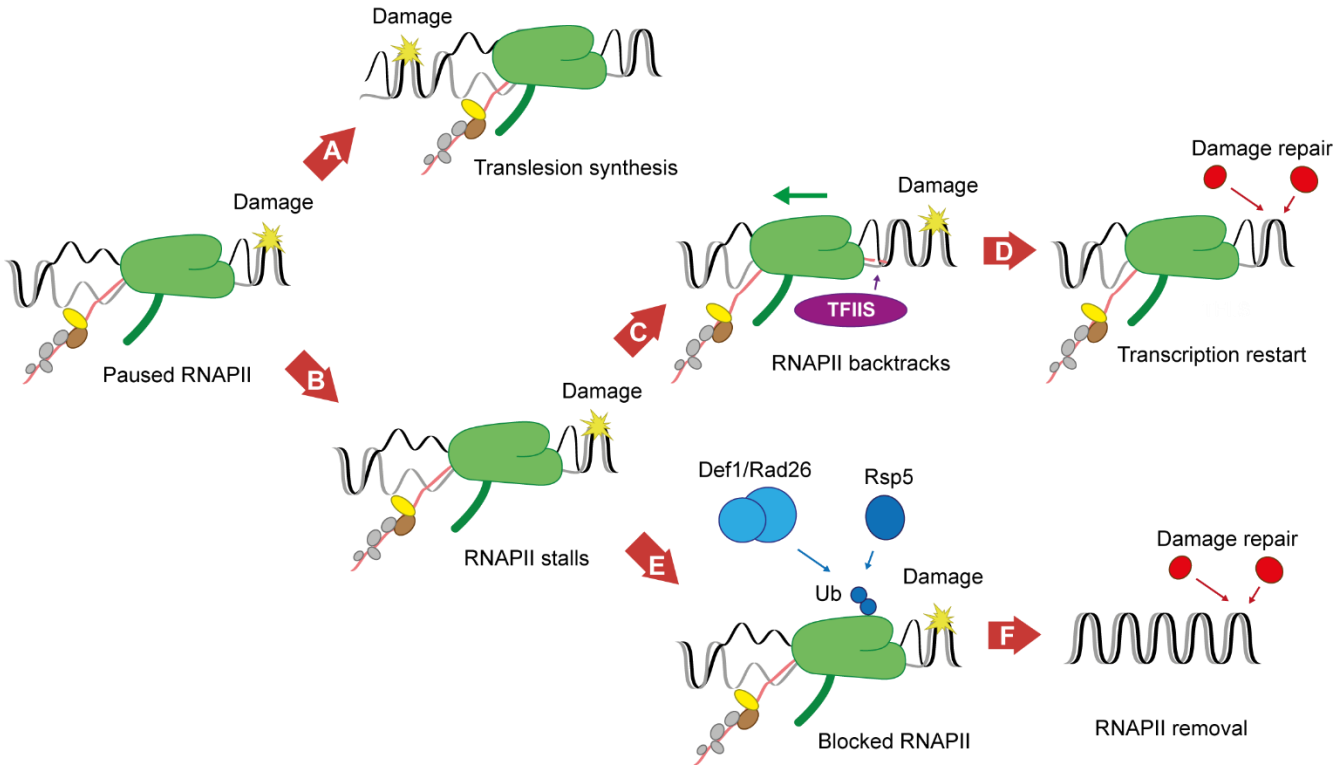


Figure I2. Possible outcomes of an RNAPII encountering a DNA damage.
Model to explain the different outcomes of RNAPII encounterin a DNA damage in the template. **(A)** Some kind of damages can be bypassed by RNAPII in a translesion synthesis process. **(B)** If the RNAPII can not bypass the damage, it stalls at the damage site. Two alternative possibilities can take place: **(C)** if the RNAPII can be backtracked by TFIIS and the damage can be repaired, **(D)** transcription would be restarted; **(E)** if the damage persists or RNAPII can not be backtracked, the blocked RNAPII is ubiquitinated **(F)** and removed from the chromatin, allowing the DNA repair.

the Elc1–Cul3 elongin E3 ligase complex (ElonginA/B/C-Cullin 5 complex in humans) that poly-ubiquitinates the Lys-48 residue, triggering Rpb1 proteolysis (Harreman *et al.*, 2009).

1.5.1. Effect of UV light induced DNA damage on transcription

DNA bulky lesions, mostly those produced by ultraviolet (UV) light, block transcription by direct obstruction of the RNAPII advance, as it was discovered in cells from Cockayne's syndrome and xeroderma pigmentosum patients, that are deficient in the repair of this kind of damage (Mayne & Lehmann 1982; Protić-Sabljić & Kraemer 1985; Mellon *et al.*, 1987). Pyrimidine dimers produced by UV light are primarily repaired by nucleotide excision repair (NER), that recognizes bulky adducts and structures that disrupt the DNA double helix. NER machinery induces incisions at both sides of the damaged DNA template, removing the oligonucleotides that contains the adduct, latterly refilling and ligating a new synthetized DNA fragment complementary to the undamaged strand (Spivak 2015). In NER, two similar pathways could be differentiated: global genomic repair (GGR), in which the lesion is recognized by yeast Rad7/Rad16 and depends on the ubiquitin ligase Elc1; or transcription coupled repair (TCR), that relies on RNAPII stalling to recognize the damage and it is mediated by Rad26/Rad9. In both cases, these two pathways converge by recruiting TFIIH to the lesion, which unwinds the DNA and catalyses the incision of the damaged template (Spivak 2015; Lejeune *et al.*, 2009; Li & Smerdon 2004).

In response to UV damage, the level of transcribing Ser5-P RNAPII decreases, and this reduction persists longer at promoters, indicating a lack of initiation and a defect in elongation. If the damage is not repaired, RNAPII is ubiquitinated by Rsp5/NEDD4 in order to resolve the stall (Harreman *et al.*, 2009; Jansen *et al.*, 2002). The process is not completely understood in eukaryotic cells, but it seems to be mediated by Rad26/Def1 in yeast, as *DEF1* deletion increases sensitivity to UV light, decreases RNAPII degradation and fails to produce polyubiquitylated Rpb1 (Woudstra *et al.*, 2002). In contrast, *rad26Δ* mutants present an increase in Ser5-P RNAPII at short times after damage induction, followed by a complete RNAPII delocalization from the DNA if the damage persists, suggesting a stronger stalling of RNAPII (Jansen *et al.*, 2002). In human cells, it has been proposed that ERCC6 (Cockayne syndrome group B or CSB, homolog of Rad26), binds

to RNAPII when it is blocked by UV damage (Iyama & Wilson 2016) and recruits other TCR factors, including ERCC8 (CSA), that is a recognition factor for the DCX E3 ubiquitin ligase complex, but also TFIIS, that restores backtracked RNAPIIs, and UVSSA, that facilitates CSA and CSB-dependent ubiquitylation of the stalled RNAPII (Schwertman *et al.*, 2013; Fei & Chen 2012). In yeast, different mutants of RNA-processing or export factors, as *mex67-5*, *sub2-206* or *hpr1Δ* are more sensitive to UV light in the absence of GGR factors, suggesting a role of mRNP assembly in TCR (Gaillard *et al.*, 2007). Moreover, UV light irradiation of the *hpr1Δ* mutant reduces RNAPII occupancy and processivity and is lethal for the double mutant *hpr1*Δ *def1*Δ. All together this data suggests that impaired RNAPIIs that encounter a damaged DNA tend to block and they are eliminated by Def1 (Gaillard *et al.*, 2007).

To summarize, in the presence of bulky DNA lesions produced by UV light, RNAPII stalls. If the damage persists in time due to lack of some mRNP or repair factors, the elongation complex is removed from the chromatin by polyubiquitination and degradation of Rpb1 by the proteasome. This process is modulated by Rad26/Def1 in yeast and is accompanied of transcription repression at promoter level.

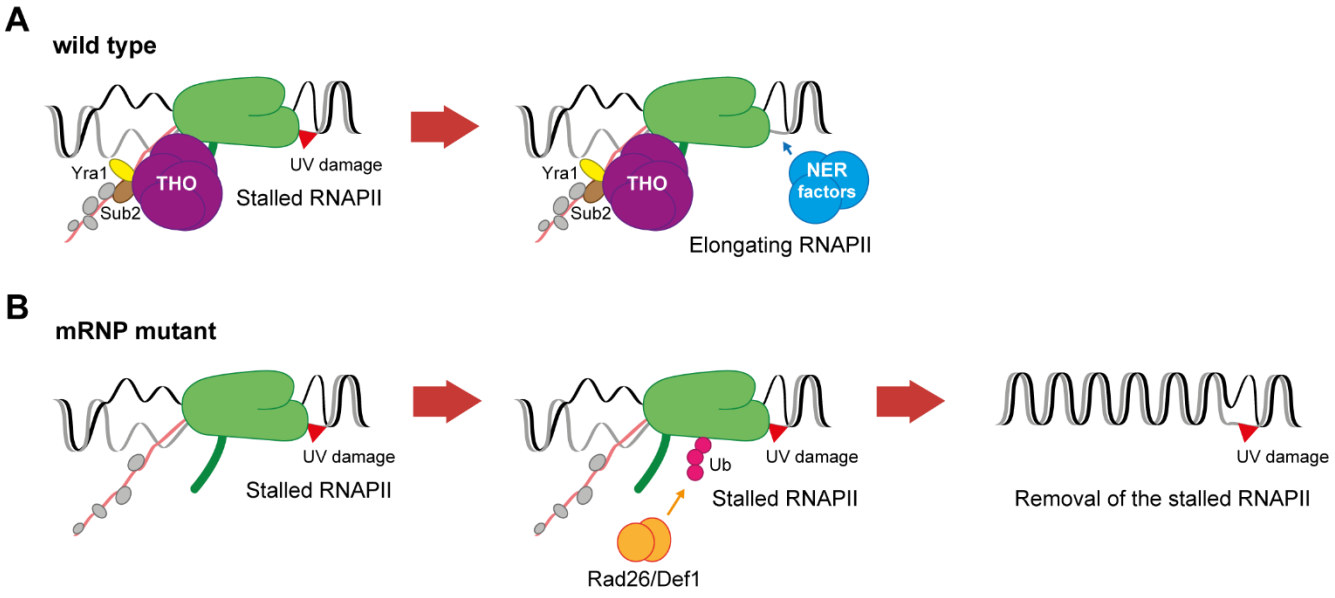


Figure I3. UV light induced damage stalls and removes RNAPII.
Model to explain the effect on transcription of UV damage. **(A)** In a wild-type strain the lesion is detected by RNAPII that recruits the NER machinery to repair the damage and restart transcription. **(B)** In mRNP deficient cells, the RNAPII stalls at the UV damage and a deficient recruitment of the NER machinery leads to ubiquitylation of RNAPII by the Rad26/Def1 complex and the removal of RNAPII. At the same time transcription initiation is repressed at promoter level.

1.5.2. Double-strand DNA breaks block transcription

Double-strand breaks (DSBs) are the most harmful damage in the DNA, as they result in cell death if they are not repaired. RNAPII is greatly affected by the presence of DSBs in the DNA, repressing initiation, inhibiting transcription elongation and finally releasing it from the DNA if the damage persists (Shanbhag *et al.*, 2010). To avoid these deleterious effects of DSBs, cells repair these damages by two different and competing pathways: nonhomologous end joining (NHEJ), that resolves DSBs by ligating the two ends, potentially leading to information loss during the processing of the break and to chromosomal rearrangements; or homologous recombination (HR), that requires DNA resection to generate ssDNA that searches for homology in the genome to repair the DSB in an error-free manner (Pardo *et al.*, 2009). The choice of one or other pathway depends on the cell cycle stage and on 5'-end DSB resection, that is irreversible and impedes NHEJ. While NHEJ machinery is efficient in all the cycle stages, DSB-end resection is more efficient during S and G2 phases, when the presence of a sister chromatid to copy information facilitates homologous recombination (Pardo *et al.*, 2009). In both cases, the presence of a DSB is detected and bound by the Ku complex (formed by yKu70/KU70 and yKu80/KU80, that are part of the human DNA-PK complex) and the MR(X)N complex (composed by Mre11, Rad50 and Xrs2/NBS1). Human DNA-PK, together with ATM/Tel1 and ATR/Mec1, mayor transductor kinases that would trigger the DNA damage response (DDR), a system that includes a set of DNA repair and damage tolerance processes, but also cell cycle checkpoints (Pardo *et al.*, 2009; Matsuzaki *et al.*, 2008; Heyer *et al.*, 2010).

Induction of a single DSB in the proximity of a promoter region of a reporter human gene leads to a strong transcription repression in an ATM-dependent manner. This decrease in transcription is a direct effect of the damage as it is restored following the dynamics of the break repair and it is associated with a reduction in Ser2-P RNAPII, while global RNAPII level remains unaltered. It has been proposed that transcription inhibition occurs as a consequence of a local condensation of the chromatin that extends several kilobases from the damage site (Shanbhag *et al.*, 2010). An alternative model poses that DNAPKcs, the catalytic subunit of the DNA-PK, but not ATM signalling, is the responsible for the transcription inhibition when DSBs are produced in the gene body (Pankotai *et al.*, 2012). In both cases, polyubiquitylation of RNAPII would remove the stalled RNAPII, while ubiquitylation of histone H2A, but also other heterochromatic

marks, as HP1, would condensate chromatin in order to repress transcription. Moreover, in yeast it has been confirmed the spread of transcription inhibition in DSB proximal genes. Surprisingly, this inhibition was not dependent on Tel1 or Mec1 but on the resection of the DNA (Manfrini *et al.*, 2015).

All this data suggests a model in which changes in chromatin induced by DDR in response to DSBs could prevent transcription elongation; while local DSB signalling, mediated by DNA-PK, is able to inhibit both, elongation and initiation of transcription, by ubiquitylation of RNAPII and PARylation of the mRNA, a modification driven by the Poly ADP-ribose polymerase 1 (PARP1) that would remove the nascent RNA from the chromatin (Pankotai & Soutoglou 2013; Chou *et al.*, 2010).

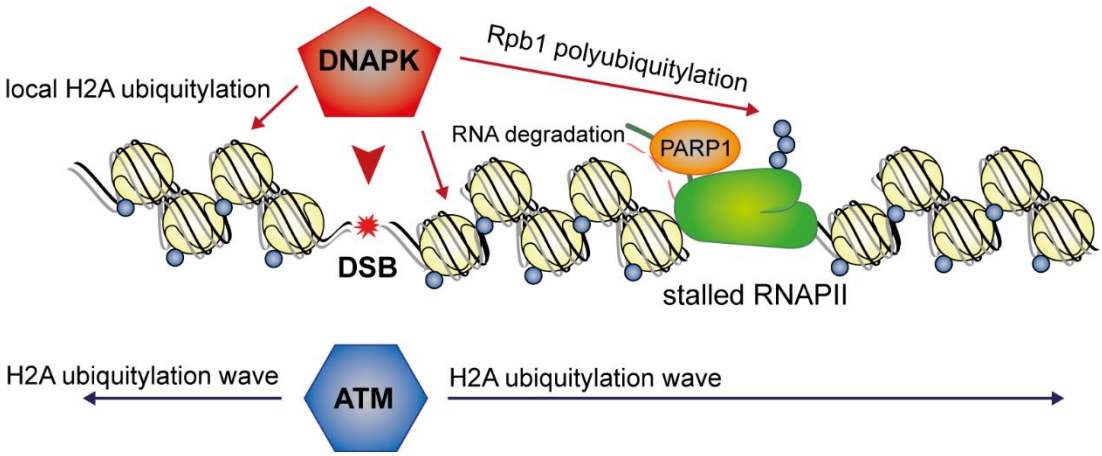


Figure I4. DSB blocks transcription and removes RNAPII.
Model summarizing the effect of DSB in transcription. The presence of DSBs lead to a histone H2A ubiquitylation wave that expanded from the damage site in both directions. This process is triggered by the ataxia telangiectasia mutated (ATM) kinase. H2A ubiquitylation impedes the decondensation of the chromatin that is needed for transcription elongation, thus blocking RNAPII. DNA-dependent protein kinase (DNAPK) recognises and binds to the DSB, inducing local histone ubiquitylation and RNAPII degradation by the proteasome. In the presence of DSBs, PARP1 locally degrades nascent mRNA and contributes to establishing a repressive chromatin structure.

2. OBJETIVES

The aim of this thesis is to improve our knowledge on how overexpression of genes involved in mRNP biosynthesis could induce R-loop formation, leading to genomic instability and how different types of DNA damage could affect RNA polymerase II transcription. For that, we pursued the following specific objectives:

1 – Identify new mRNP proteins that, as Yra1, when they are in excess, affect R-loop metabolism.

2 – Determine the fate of an elongating RNAPII that encounters an R-loop *in vivo*.

3 – Analyse the effects of single-strand break over RNAPII mediated transcription.

4 – Develop a new genetic system to study recombination with the homologous chromosome induced by R-loops and single-strand breaks.

3. RESULTS

3.1. Identification of new RNA-binding factors that induce genome instability when overexpressed

During elongation, R-loops are formed naturally. These non-canonical structures have several roles in transcription regulation, replication or during switch-class recombination. However, R-loop formation is highly regulated as indicated by the presence of several pathways that impedes their overaccumulation. Probably, most important path to prevent R-loop accumulation is a proper mRNP assembly and the correct coupling of the different steps in transcription, from chromatin regulation to the export of a competent mRNP. In that sense, different studies showed that the absence or downregulation of different components of the THO/TREX complexes, export factors or chromatin remodelers leaded to DNA:RNA hybrid increase and R-loop dependent genomic instability (García-Benitez et al., 2017; García-Pichardo et al., 2017; Gómez-González et al., 2011). But, if previous studies have shown how the lack of different RNA-binding proteins are able to induce R-loop dependent genome instability, none of them focused in the possible outcomes of their overabundance. Yra1 is an RNA-binding protein that is tightly regulated in the cell. The overexpression of *YRA1* gene increases DNA damage, transcription-dependent hyper-recombination and R-loop accumulation, that are hallmark of genomic instability. This genomic instability is probably a consequence to the ability of an excess of Yra1 to stabilize R-loops (Gavaldá *et al.*, 2016; García-Rubio *et al.*, 2018). In the light of this new role of Yra1, we wondered whether an excess of other RNA-binding proteins (RBPs) whose overexpression is detrimental for the cell could also induce genome instability in an R-loop-dependent manner.

3.1.1. Screening of the MW90 overexpression library reports candidates that induce genome instability

Considering that the overexpression of *YRA1* severely reduces growth on yeast lacking Hpr1 (García-Rubio *et al.*, 2018) probably due to a synergistic effect through the stabilization of R-loops in a strain that overproduce them, we wondered if overexpression of other proteins could also reproduce this phenotype. This would allow us to identify new mRNA processing proteins that affect R-loop metabolism. To explore that possibility, we performed a screening by overexpressing a library, MW90 (Waldherr *et al.*, 1993), containing DNA fragments from the genome of the budding yeast *Saccharomyces cerevisiae* cloned into YEp351, a multicopy expression plasmid. The genomic library was overexpressed in HPR1DGK, a diploid strain with only one copy of *HPR1* gene that is tagged with the Auxin-induced Degron system (AID) (Nishimura *et*

al., 2009) (*HPR1-DG*). The AID system causes a rapid depletion of the Hpr1 protein by the addition of auxins (NAA) to the media. Considering that expression of the MW90 genomic library is constitutive, the capability to deplete Hpr1 at will let us to differentiate between the growth inhibition that could be produced by the overexpression of the library itself and the synergistic growth defects that we want to select as a screening criterion. We also choose to use a diploid strain as we thought that it could improve the fitness of the cells.

We transformed this diploid HPR1DGK strain with the MW90 genomic library using a high efficiency yeast transformation protocol. About 36.800 transformants were obtained. 95% of the colonies presented a normal growth phenotype, but 1750 transformants showed a slower growth rate in medium without auxin. We wanted to focus

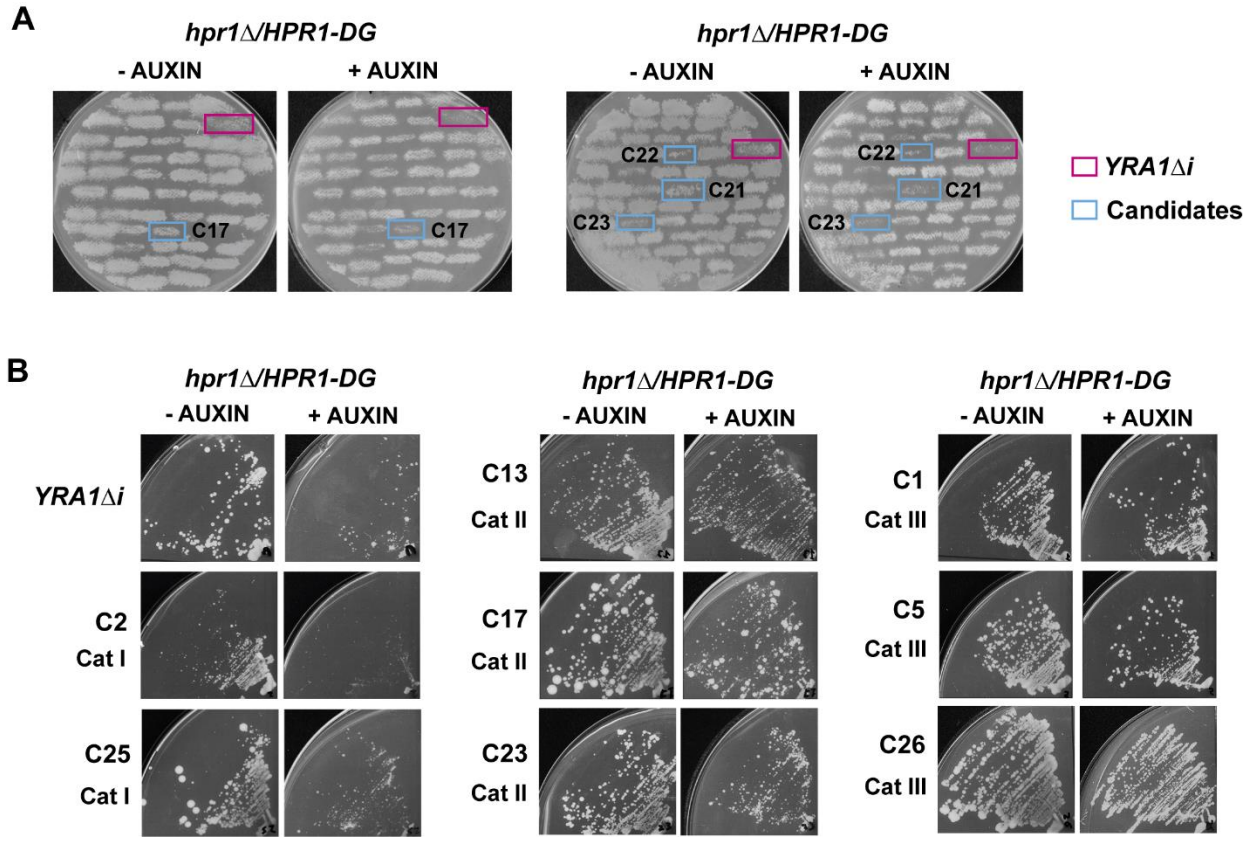


Figure R1. Screening for genes whose overexpression increased genetic instability.
(A) Example of isolated transformants of HPR1DGK, an *hpr1Δ/HPR1-DG* diploid strain, containing different plasmids from the MW90 overexpression library or YEp351-YRA1Δi, a multicopy vector containing *YRA1* gene without its intron (*YRA1*Δi). The clones were replicated in medium without (-Auxin) or with NAA (+Auxin) in order to deplete Hpr1. Candidate clones that decreased growth in Hpr1 depletion conditions are marked in blue. **(B)** Streaks of HPR1DGK in plates with or without auxin, containing different candidates grouped in three different categories depending of their reduction in growth when Hpr1 is depleted.

in the clones with reduced growth because, as reported for *YRA1* overexpression, we expected that their excess would have harmful effects in the cells even without depleting Hpr1. For this reason, all the colonies with a low growth phenotype were isolated in SC-leu plates to maintain the selection of the MW90 plasmid and replicated into SC-leu plates with 0.5 mM of NAA auxin in order to deplete Hpr1. We also included 500 of the normal growth transformants to check whether, by the criterion of selection, we were overseeing possible candidates. As positive control, a strain carrying YEp351-YRA1Δi was included. This plasmid has the same backbone that the MW90 library and contains *YRA1* without its intron (*YRA1Δi*) that ensures the overexpression of the gene (Rodríguez-Navarro *et al.*, 2002). Of the 2250 transformants checked, only 32 decreased their growth when Hpr1 was depleted in the auxin containing medium (Figure R1-A), fulfilling the selection criterion. None of the 500 clones with normal growth in plates without auxin reported a growth reduction in Hpr1 depleted cells. Hence, although it was possible that by restricting the analysis to the slowly growing transformants we may lose candidates, the possibility was low. The 32 candidates and *YRA1* control were streaked in SC-leu and SC-leu with NAA to confirm the result (Figure R1-B). Among them, only 14 candidates corroborated the reduction in growth upon depletion of Hpr1; 4 of them with a phenotype similar to the *YRA1Δi* overexpressed clone, and 10 with a milder reduction in growth. The plasmid for the MW90 library of each clone was isolated but only 2 of them (of the second category) had a genomic fragment inserted, with the other 12 being only empty plasmid. These two clones, hereafter referred to as C17 and C23, were sequenced. The C17 plasmid contained an 8.3 kb fragment from Chromosome XV, including (Figure R2-A):

- A carboxyterminal-truncated copy of *TSR4* gene, a pre-rRNA processing protein.
- *DIS3*, one of the catalytic subunit of the exosome core complex (Dziembowski *et al.*, 2007).
- *TAT2*, a tryptophan and tyrosine permease (Schmidt *et al.*, 1994)
- *YOL019W-A*, a putative unknown function ORF (Kumar *et al.*, 2002).
- Two tRNA genes, *SUF17* and *SUP3*.

The C23 plasmid contained a fragment from chromosome VII of 7.2 kb, including only two genes (Figure R2-A):

- *MGA1*, encoding for a protein similar to heat shock transcription factors (Lorenz & Heitman 1998).
- *RIE1* (*YGR250C*), an RNA-binding protein (Feroli *et al.*, 1997).

Next, we tested whether the reduction in growth rate produced by the overexpression of the library plasmids C17 and C23 in Hpr1-depletion conditions was related to an increase in genomic instability. To do so we measure DNA damage by quantifying Rad52-YFP foci in wild type cells containing either C17 or C23 plasmid. We included YEp351-YRA1Δi as a positive control. Rad52 is a protein involved in DNA repair that accumulates, forming repair centres that can be visualized like sub-nuclear foci

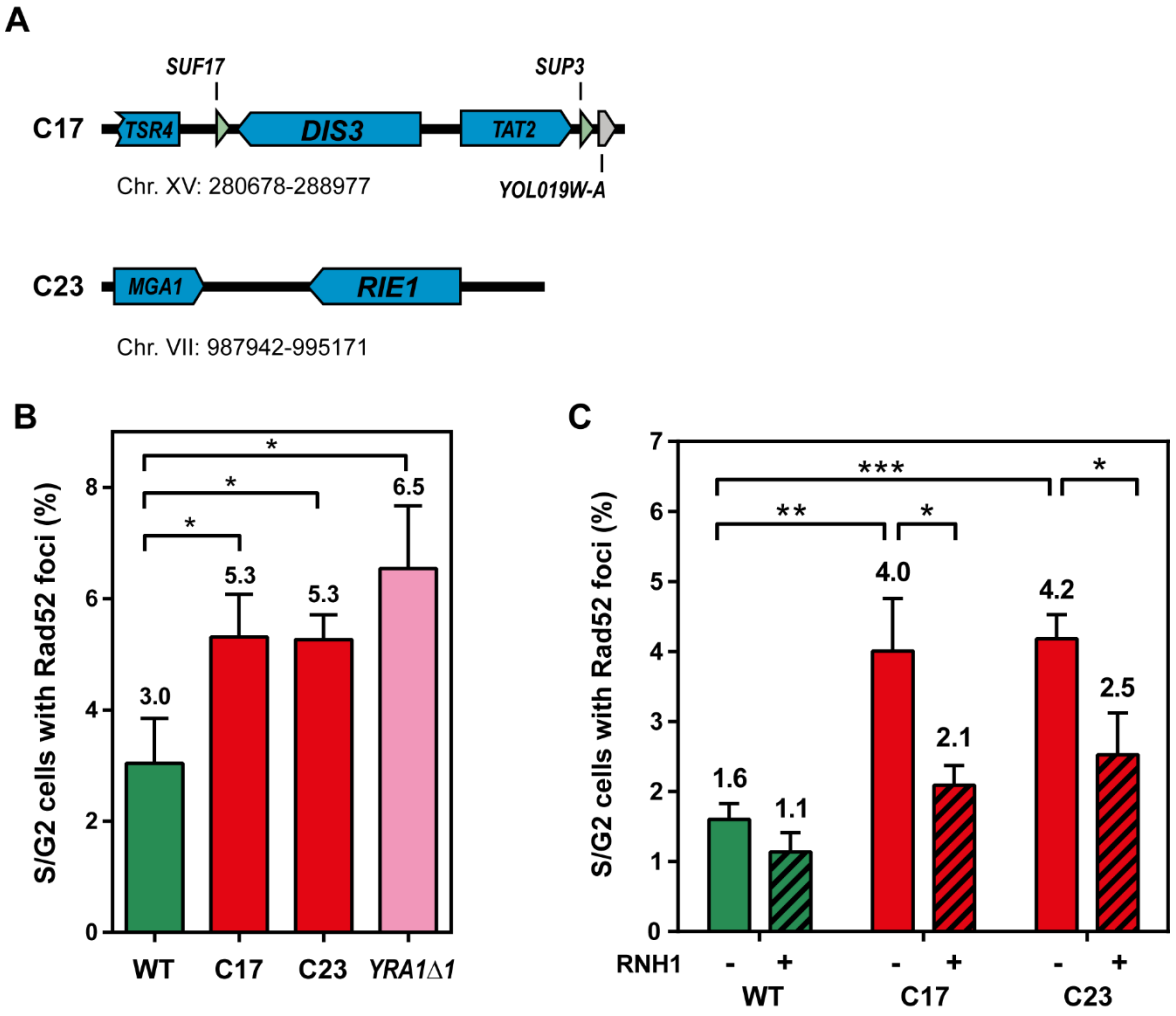


Figure R2. C17 an C23 plasmids increased R-loop dependent DNA damage.
(A) Schematic map of the genomic regions inserted into YEp351 multicopy vector in the C17 and C23 plasmids from the MW90 library. Rad52-YFP foci formation in W303-1A *RAD5* strain (Ybp249) transformed with pWJ1213 and either C17, C23 or YEp351 empty vector (WT) **(B)** and with pGALRH1 **(C)**. Cell were growth in 2% glucose medium (RNH1-) or 2% galactose medium (RNH1+) in order to overexpress RNase H1. Average and SD from at least three independent experiments are shown. *, p≤0,05; **, p≤0,01; ***, p≤0,005 (Student's t-test).

if it is fused to a fluorescent protein like YFP (Lisby *et al.*, 2001). We observed a significative increase in Rad52-YFP foci in both cases, C17 and C23 overexpression (Figure R2-B). Therefore, overexpression of the genes contained in either C17 or C23 causes DNA damage.

In order to determine if the increase in damage is R-loop dependent, we overexpressed RNase H1 (*RNH1*, an endonuclease that removes R-loops by specifically cleaving the RNA moiety in DNA:RNA hybrids) together with C17 or C23. We observed a significative reduction in Rad52-YFP accumulation when *RNH1* was overexpressed in both candidates (Figure R2-C). Therefore, the increase in DNA damage cause by overexpression of either C17 or C23 was at least partially mediated by DNA:RNA hybrids.

3.1.2. *DIS3* and *RIE1* overexpression produces DNA damage from C17 and C23

Both, C17 and C23 plasmids contained more than one complete gene inserted. To establish which genes could be producing the Rad52-YFP foci increase when the C17 plasmid was overexpressed, we did two different deletions of the genomic region cloned. The first deletion (C17ΔA) was done by digestion with XbaI endonuclease, that removed part of the *DIS3* ORF but also *SUF17* and *TSR4* genes. For the second deletion (C17ΔB), we digested with BamHI, eliminating *DIS3*, *TAT2*, *SUF17* and *TSR4* genes. None of these two new plasmids increased in a significant manner Rad52-YFP foci (3.3% and 2.6% cells with foci respectively compared with 2.2% for the wild type). C17ΔA and C17ΔAB deletions had in common the lack of *TSR4* and *DIS3*. Since *TSR4* gene was already truncated in C17, we decided to check *DIS3* overexpression as it is the most likely to be producing the phenotype. For that, *DIS3* entire gene, with its own promoter (500 pb upstream) and terminator regions (100 pb downstream), was cloned into YEp351, which is the same vector employed in the library. Rad52-YFP foci levels increased significatively with the overexpression of *DIS3* from this plasmid, with 5.9% of cells with foci compared to the 2.2% in the wild type. *DIS3* overexpression depicted similar levels of cells with Rad-YFP foci than the C17 plasmid, that presented a 5.2%. This suggest that *DIS3* overexpression was responsible of the DNA damage increase produced by C17 (Figure R3-A).

The plasmid C23 only contained two genes: *MGA1* encoding a protein similar to some heat shock transcription factors; and *RIE1*, a poorly characterized RNA-binding protein with three RNA-recognition motifs (RRM). Hence, we decided to clone the *RIE1* ORF in pYES2 plasmid because, unlike Mga1, Rie1 was an RNA-binding protein, which is the criterion that we established for the screening. *RIE1* overexpression showed a significant increase number in Rad52 foci (5.4%) compared with the wild type (2.1%), indicating that this gene overexpression increased damage (Figure R3-B).

In conclusion, the screening identified two different genes, *DIS3* and *RIE1*, both encoding RNA-binding proteins that when overexpressed reduced growth in a mutant

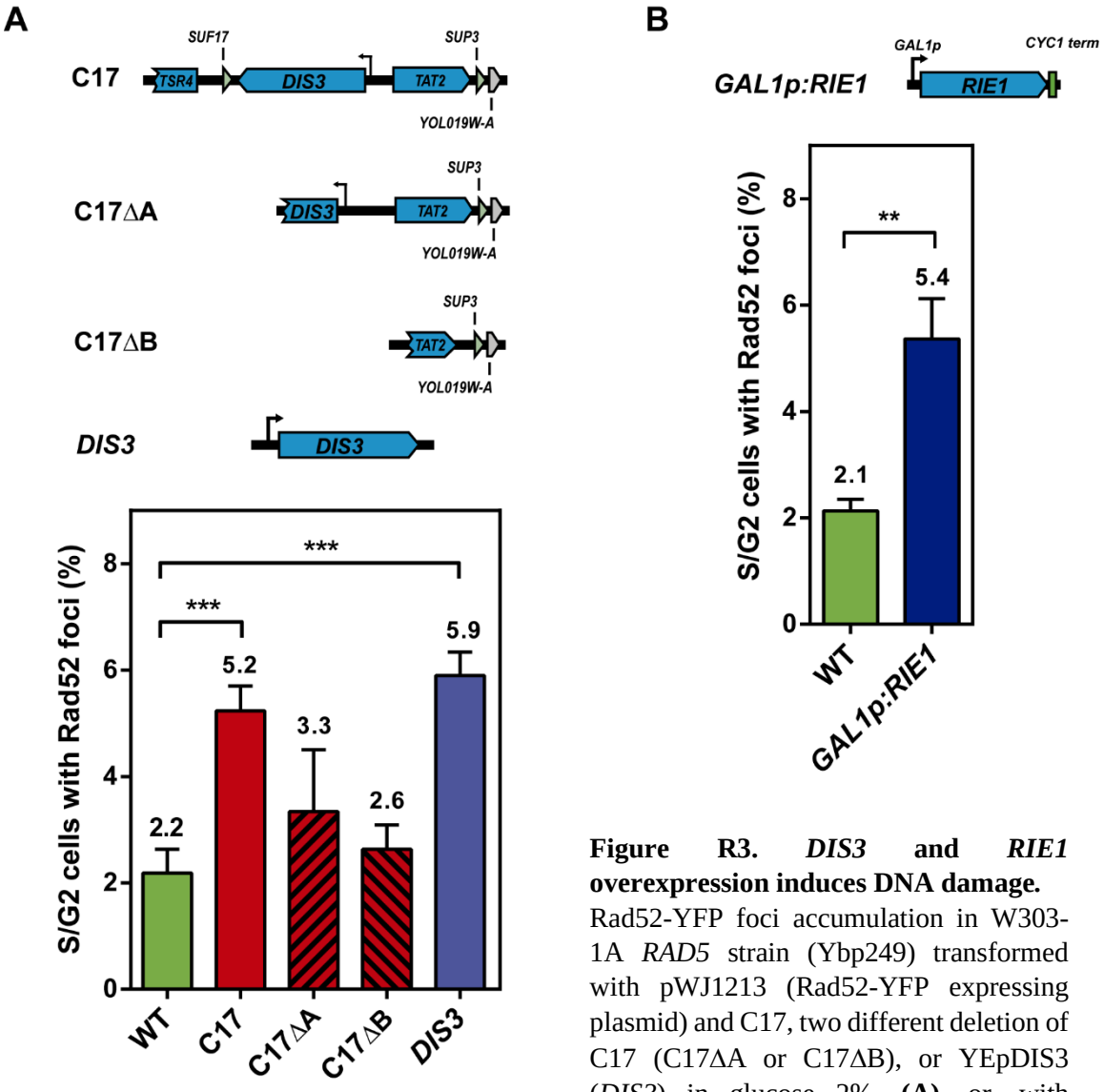


Figure R3. *DIS3* and *RIE1* overexpression induces DNA damage. Rad52-YFP foci accumulation in W303-1A *RAD5* strain (Ybp249) transformed with pWJ1213 (Rad52-YFP expressing plasmid) and C17, two different deletion of C17 (C17ΔA or C17ΔB), or YEpDIS3 (*DIS3*) in glucose 2% **(A)** or with pYES-RIE1 (*GAL1p:RIE1*) in galactose 2% **(B)**. Average and SD of at least three independent experiments are shown. *, p≤0,05; **, p≤0,01; ***, p≤0,005 (Student's t-test).

with increased R-loop formation. Moreover, their overexpression also induced DNA damage that, together with the decrease in *hpr1* strain growth, it could indicate a role in R-loop homeostasis in the cell.

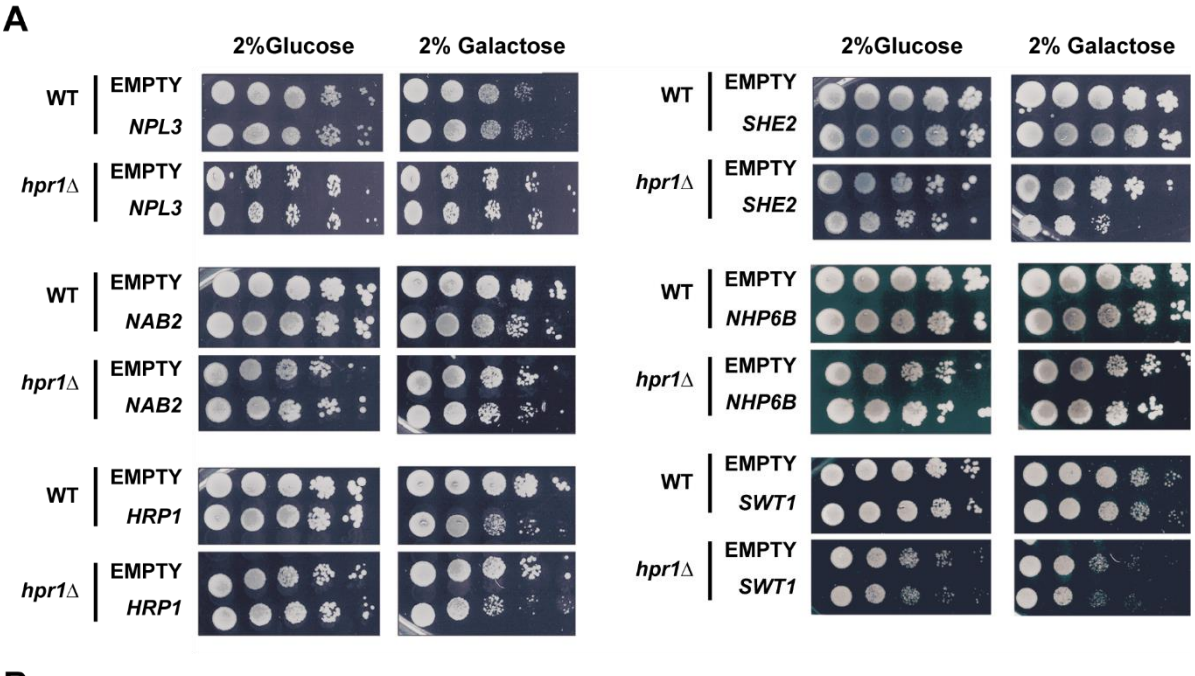
3.1.3. Direct screening of RNA-binding proteins that induce genomic instability upon overexpression

Since we did not know if all the yeast genes were represented in the library, we conducted in parallel a second screening to find other RNA-binding proteins that could affect R-loop metabolism and that were not obtained in the MW90 screening. We did that by overexpressing selected RNA-binding proteins that reduces cell viability when overexpressed, similarly to what was described for *YRA1Δi*. To select the candidates, we crossed entries for yeast genes with the Gene Ontology Term 'RNA binding' with genes with an overexpression phenotype of 'vegetative growth: inviable' using *Saccharomyces* Genome Database. From the 19 different genes obtained, including *YRA1*, we discarded those genes not directly related with mRNP formation or mRNA processing. The final list of genes selected included:

- *NAB2*, involved in the formation of export competent mRNPs (Anderson *et al.*, 1993).
- *NHP6B*, encoding a protein that interacts with chromatin modifiers (FACT, Swi/Snf or Spt6 among others) facilitating the formation of the preinitiation complex of RNA polymerase II (Paull *et al*., 1996).
- *NPL3*, a hnRNP with a role in transcription, mRNA splicing and transport (Bossie *et al.*, 1992).
- *HRP1*, component of cleavage factor I, required for pre-mRNA processing (Kessler *et al.*, 1997).
- *SHE2*, factor involved in localization of specific mRNAs to the bud tip. It interacts with DSIF complex (Long *et al.*, 2000).
- *SWT1*, an endoribonuclease that participates in mRNP quality control and associates with the nuclear pore complex (Röther *et al.*, 2006).

In order to find a growth reduction in a mutant strain that increases DNA:RNA hybrids, that could point to a R-loop metabolism defect by the excess of these proteins,

the selected genes were individually cloned under *GAL1* promoter control in pYES2 plasmid and overexpressed in wild type and *hpr1* cells, in a similar manner as we did for the MW90 library screening. At this point we decided to work with the haploid mutant of *hpr1* as the diploid did not seems to improve the fitness of the cells. Furthermore, the AID *HPR1-DG* system was not required when we could control the overexpression of the candidate genes adding galactose to the medium. We found that *HRP1* overexpression



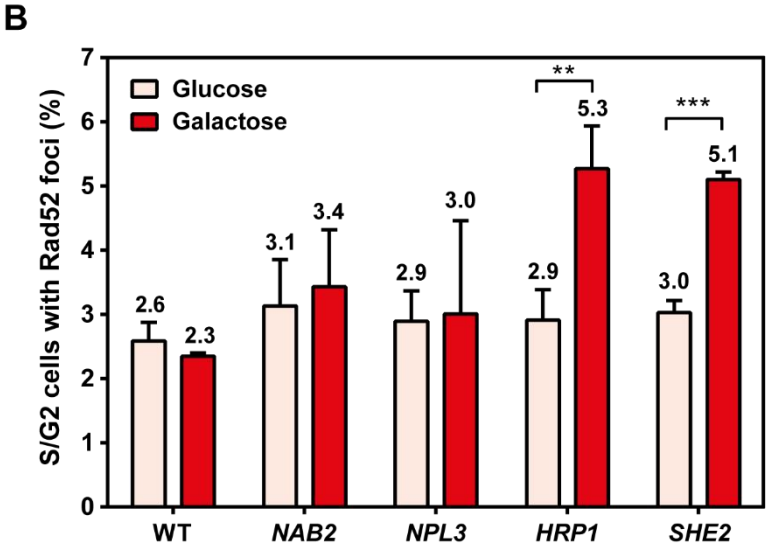


Figure R4. Search for RNA-binding proteins whose overexpression increases genetic instability.
(A) Drop test assay of *NAB2*, *NPL3*, *HRP1*, *SHE2*, *NHP6 and SWT1* genes overexpressed from pYES2 on W303-1A (WT) or SChY58a (*hpr1*Δ). The empty plasmid (EMPTY) was included as control. The indicated genes were overexpressed in 2% galactose growing conditions. **(B)** Rad52-YFP foci formation in W303-1A *RAD5* strain (Ybp249) transformed with pWJ1344 (Rad52-YFP expressing plasmid) and pYES2 vector with either *NAB2*, *NPL3*, *HRP1* or *SHE2* cloned or the empty plasmid (WT), grown in 2% glucose (repressed) or 2% galactose (overexpressed) medium. Average and SD of three independent experiments are shown. *, p≤0,05; **, p≤0,01; ***, p≤0,005 (Student's t-test).

caused a severe growth defect in both, wild type and *hpr1* mutant; *NAB2* overexpression produced a slight general reduction in growth in both strains, where *SHE2* overexpression only reduced growth in the *hpr1* mutant but not in the wild type. Finally, *NHP6B*, *NPL3* or *SWT1* overexpression had no effect neither in wild type nor in *hpr1* (Figure R4-A). The result showing that Npl3 excess did not reduce growth in *hpr1* could be in agreement with previous results reporting that *NPL3* overexpression partially supressed the hyperrecombination phenotype of the *hpr1* mutant (Santos-Pereira *et al.*, 2013), compensating the loss of Hpr1. This could suggest that Npl3 has a role in R-loop metabolism, even if this role acts in a different pathway than Hpr1 does.

Next, to determine if the overexpression of these genes leads to genomic instability, we checked the amount of Rad52-YFP foci exclusively for those candidates that decreased growth rate: *NAB2*, *HRP1* and *SHE2*, in the same way that we did with the screening of the MW90 library. *NPL3* was also included because its null mutant was previously described to increase DNA damage in an R-loop dependent manner (Santos-Pereira *et al.*, 2013). The results showed that both, *HRP1* (5.3%), and *SHE2* (5.1%) significantly increased the percentage of cells with Rad52-YFP foci compared to the wild type (2.6%), whereas *NAB2* and *NPL3* overexpression (3.4% and 3%, respectively) did not increase damage (Figure R4-B). The overexpression of the mentioned genes was confirmed by northern blot assay (Figure R5).

Considering these results, we focused the analysis in *HRP1* and *SHE2*, as the increase in damage caused by the overexpression of these RNA-binding proteins could be a consequence of an accumulation of R-loops.

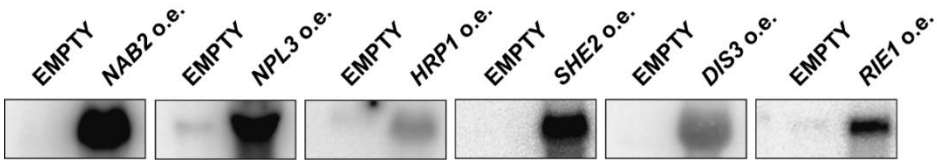


Figure R5. Overexpression control of the candidates by northern blot.
Northern blot assay with pYES2 (EMPTY), pYES-NAB2, pYES-NPL3, pYES-HRP1, pYES-SHE2, pYES-DIS3 or pYES-RIE1 after 2 h of transcription in 2% galactose medium in W303-1A. Specific probes for each mRNA were used.

3.1.4. Overexpression of *HRP1, SHE2, DIS3* and *RIE1* does not increase sensitivity to genotoxic agents

Given that *HRP1*, *SHE2*, *DIS3* and *RIE1* overexpression increased DNA damage, we wonder if the addition of genotoxic compounds that aggravate the damage could induce lethality. This could lead us to know with type of DNA damage is produced by the overexpression of these genes. To test that, we measured by drop test assay in plates with different genotoxic agents the growth of a wild type strain in which we overexpressed these four genes. We analysed the effect of hydroxyurea (HU), that reduces the pool of deoxynucleotides, affecting replication; UV light, that forms pyrimidine dimers, and Camptothecin (CPT), that poison topoisomerase I in the DNA. The results indicated that the overexpression of none of the candidates increased the sensitivity to any of the damaging agents beyond the effect in the wild type (Figure R6), thus, suggesting that the DNA damage that we previously measured could be produced in a different pathway of

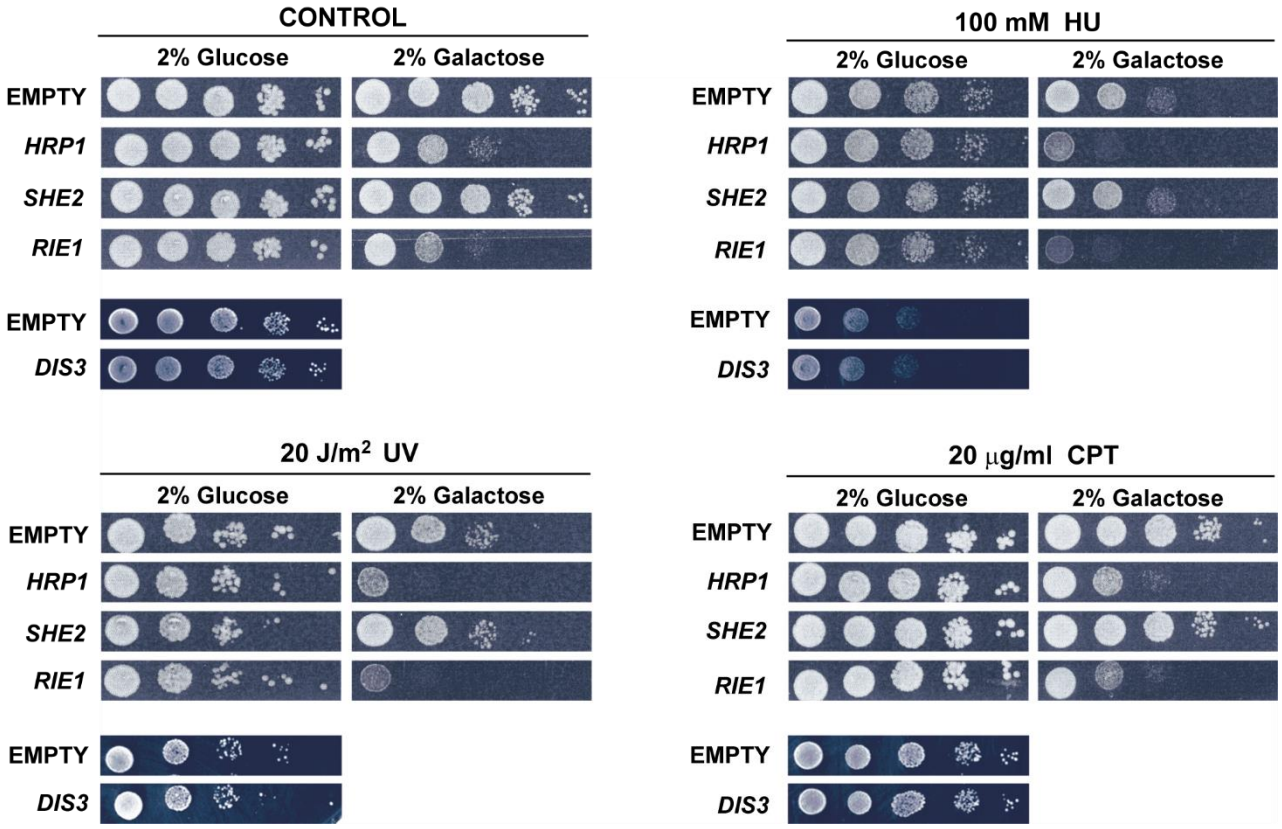


Figure R6. *SHE2, RIE1* or *DIS3* overexpression does not increase sensitivity to UV, HU or CPT. Drop test sensitivity assay of W303-1A without (glucose 2%) or with (galactose 2%) overexpression of *SHE2* or *RIE1* from pYES2 plasmid, or with overexpression of *DIS3* cloned in YEp351 multicopy plasmid in plates irradiated with 20 J/m^2 UV light, containing 100 mM hydroxyurea (HU) or 20 μg/ml Camptothecin (CPT). Plates were incubated for 3 days at 30ºC.

the induced by these genotoxic agents. A different possibility is that the damage that *HRP1*, *SHE2*, *DIS3* or *RIE1* overexpression produce may not be enough to distinguish a reduction in cell viability.

3.1.5. RNase H suppresses DNA damage induced by *DIS3, RIE1* and *SHE2* overexpression

To know if the increase in damage that we saw with the overexpression of *DIS3, RIE1, HRP1* or *SHE2* was R-loop dependent, we expressed *RNH1* to remove DNA:RNA hybrids. The endonuclease was cloned under *GAL1* promoter control to overexpress it at the same time that our candidate genes. As a control without *RNH1*, we use the empty plasmid, as repressing the expression with glucose would have affected the expression of the RNA-binding proteins as well. We found that the significative increase in the Rad52-YFP foci produced by *SHE2* and *DIS3* overexpression (6.1% for *SHE2* and 5.1% for *DIS3*) could be supressed with *RNH1* to wild type levels (to 3.0% with *SHE2* and 2.1% in *DIS3*). Rad52-YFP foci levels produced by *RIE1* overexpression (5.4%) could be partially but still significatively reduced with *RNH1* (3.9%). However, the increase in the Rad52-YFP foci produced by the overexpression of *HRP1* (5.8%) could not be supressed with *RNH1* (4.9%), pointing to a damage mechanism different to DNA:RNA hybrids (Figure R7-A).

As the increase in DNA damage seems to be caused by R-loops, we hypothesize that the overexpression of the candidate genes could reduce viability in a mutant that accumulates hybrids. To test this, we measured growth by a drop test assay overexpressing *DIS3, RIE1* and *SHE2* in *rnh1 rnh201* double mutant, that accumulates R-loops due to its inability to remove these structures, and in a wild type. In both, wild type and mutant, the overexpression of *HRP1* and *RIE1* produce a very sick growth phenotype that did not allow us to see a growth reduction. *DIS3* and *SHE2* overexpression did not affect growth neither in *rnh1 rnh201* nor in the wild type (Figure R7-B), suggesting that the levels of R-loops generated by the excess of these two proteins were not high enough to produce an impact in growth in the mutant.

Finally, taking in consideration all the results obtained, we decided to pursue the study of *DIS3, RIE1* and *SHE2*. The evidences obtained suggest that *HRP1* overexpression is causing a DNA damage by a mechanism not related with R-loops.

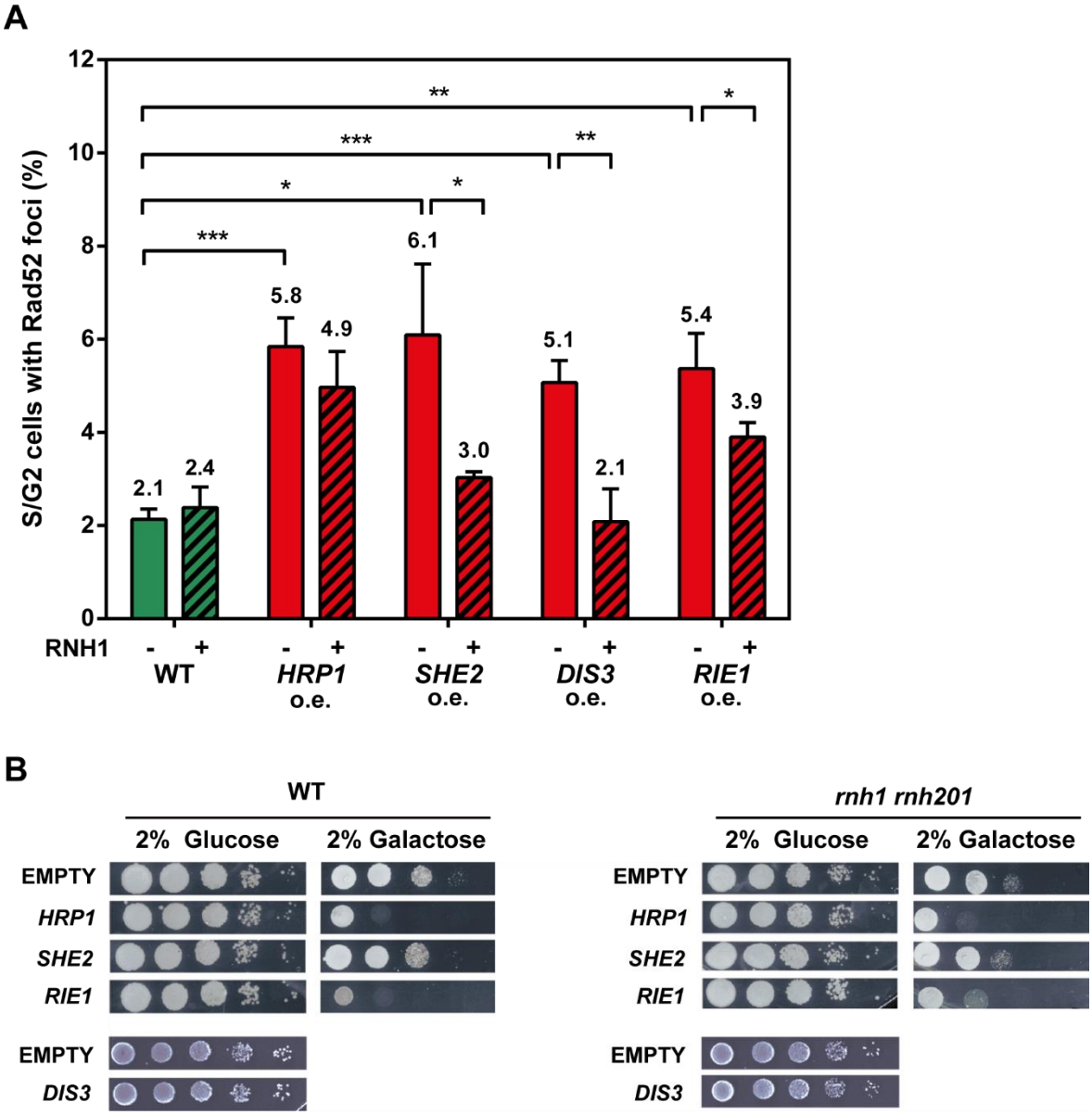


Figure R7. *RNH1* reduces Rad52 foci accumulation produced by *SHE2*, *DIS3* and *RIE1* overexpression.

(A) Rad52-YFP foci formation in W303-1A *RAD5* strain (Ybp249) transformed with pWJ1344, pYES2 vector containing either *HRP1*, *SHE2*, *RIE1* or the empty plasmid (WT) or YEpDIS3 (*DIS3* o.e.) and p313GAL1RNH1 or pRS313 empty vector for RNH1 expression. Cells were growth in 2% galactose medium. Average and SD of at least three independent experiments are shown. *, p≤0,05; **, p≤0,01; ***, p≤0,005 (Student's t-test). **(B)** Drop test assay of *HRP1*, *SHE2*, *RIE1* and *DIS3* genes overexpressed from pYES2 or YEp351 in W303 (WT) or RNH2-R (*rnh1 rnh2*01). The empty plasmid (EMPTY) was included as control.

3.1.6. *DIS3*, *SHE2* and *RIE1* overexpression did not increase recombination

Cells need to repair damage in order to maintain their genomic stability and an important part of these damages are repaired by homologous recombination (HR). Strains that accumulate damage usually have increased their recombination frequencies. Indeed, hyperrecombination was described in different mRNP mutants that increase damage but also with the overexpression of *YRA1Δi* (García-Rubio *et al.*, 2008; Gavaldá *et al.*, 2016). Since *DIS3*, *SHE2* and *RIE1* overexpression induced DNA damage, we wondered if their excess could also produce hyperrecombination. For that, we used the *GL-lacZ* chromosomic system to measure the levels of transcription-dependent recombination. The *GL-lacZ* system contains the *E. coli lacZ* gene that is a long, GC-rich sequence, prone to form R-loops, flanked by truncated direct repeats of the *LEU2* gene, that could restore a wild type *LEU2* gene by single-strand annealing (SSA) recombination (Figure M1-A and M1-B). Neither *SHE2* nor *DIS3* overexpression increased recombination frequencies significantly, but *DIS3* depicted a tendency to increase it, that was reduced to wild type levels by overexpressing *RNH1*. We also found that both, *RIE1* and *YRA1Δi*

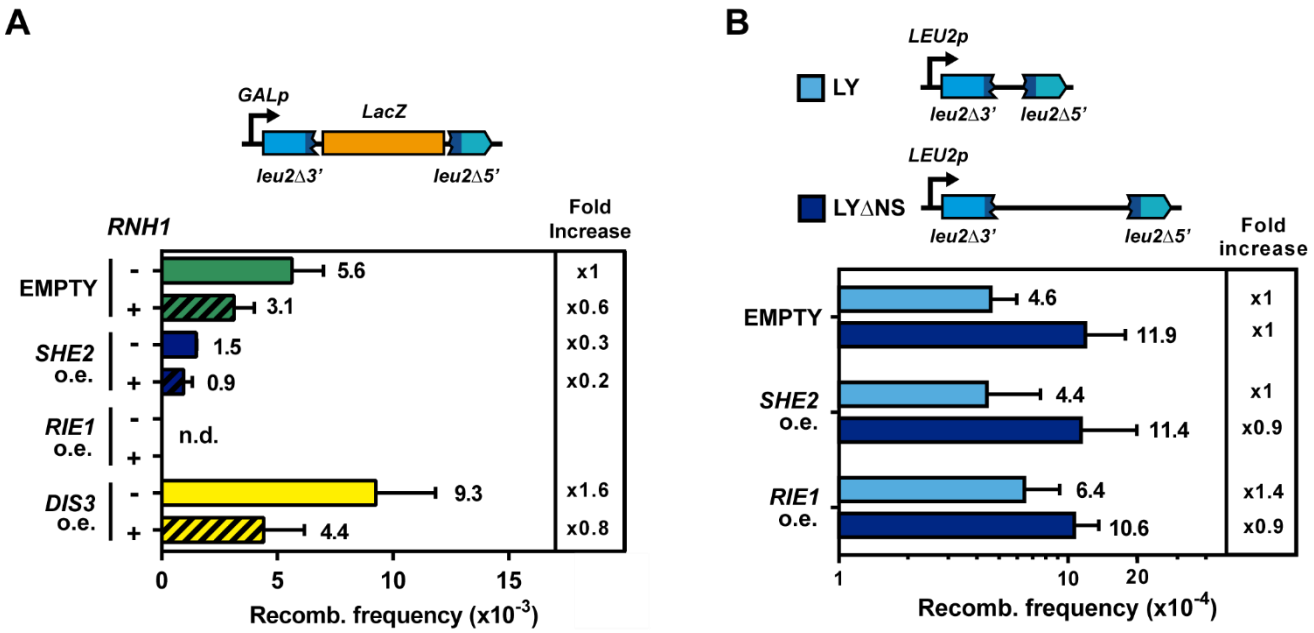


Figure R8. Recombination assay overexpressing *SHE2, RIE1* and *DIS3*.
(A) Recombination test in WGLZN strain carrying the *GL-lacZ* chromosomic recombination system, transformed with either *SHE2*, *RIE1* or *DIS3* overexpression plasmids or the empty pYES2 vector (EMPTY) and p413GALRNH1 (+ *RNH1*) or the empty vector pRS413 (- *RNH1*). **(B)** Recombination test in L or LYΔNS plasmids in W303-1A yeast overexpressing *SHE2* and *RIE1* from pYES2 plasmid or the empty control (EMPTY). In both experiments, cells where plated in galactose in order to overexpress the candidates and *RNH1*. Average and SD of at least three independent experiments are shown. 'n.d.': Not determined.

overexpression, had a huge negative impact in growth in the recombination system, that impeded us to measure recombination (Figure R8-A).

To confirm the effect of *RIE1* overexpression in recombination, we used the L and LYΔNS plasmid systems. They are based in the same truncated *LEU2* direct repeats, but containing a short spacer sequence in the case of the L system, or a long sequence derived from the YIp5 plasmid in the case of the LYΔNS (Figure M1-C). *SHE2* overexpression was also included in the assay to check the results obtained with the *GL-lacZ* and as a reference. None of them, *SHE2* or *RIE1* overexpression increased recombination beyond the wild type levels in any of the systems, L or LYΔNS (Figure R8-B).

Surprisingly, the obtained results indicated that the overexpression of none of the three candidates, *DIS3*, *SHE2* or *RIE1* increased recombination, apparently contradicting the increase seen in Rad52-YFP foci (that measure homologous recombination repair centres, as Rad52 is a central protein in this pathway). With *GL-lacZ*, L or LYΔNS systems we measured SSA recombination, which leads us to hypothesize that the damage that is generated by the excess of these genes needs to be repaired by a different pathway.

3.1.7. R-loops accumulate when *SHE2* and *RIE1* are overexpressed

Bearing in mind that *SHE2*, *RIE1* and *DIS3* overexpression increased DNA damage in a DNA:RNA hybrid-dependent manner, we addressed whether an excess of these factors increases the accumulation of R-loop directly by DNA:RNA Immunoprecipitation (DRIP). This assay determines the presence of R-loops in the genome using the S9.6 antibody, that specifically recognizes DNA:RNA hybrids. The genomic region selected were: *GCN4*, a constitutively highly expressed gene; *SPF1*, with lower transcription levels, and the rDNA *18S* gene. All of them have been previously reported to be prone to form R-loops, with a reported increase in mutants that accumulates hybrids (García-Benítez *et al.*, 2017). We also included *ASH1 E1* region for *SHE2* overexpression, as She2p has been described to directly interact with this sequence in the transcript (Shen *et al.*, 2010). We found that the overexpression of *SHE2* significatively increased R-loops in the ribosomal DNA and in the *ASH1* gene, but not in the other assessed loci. However, *RIE1* increased hybrids in *GCN4* and in the ribosomal *18S* gene, pointing to a more

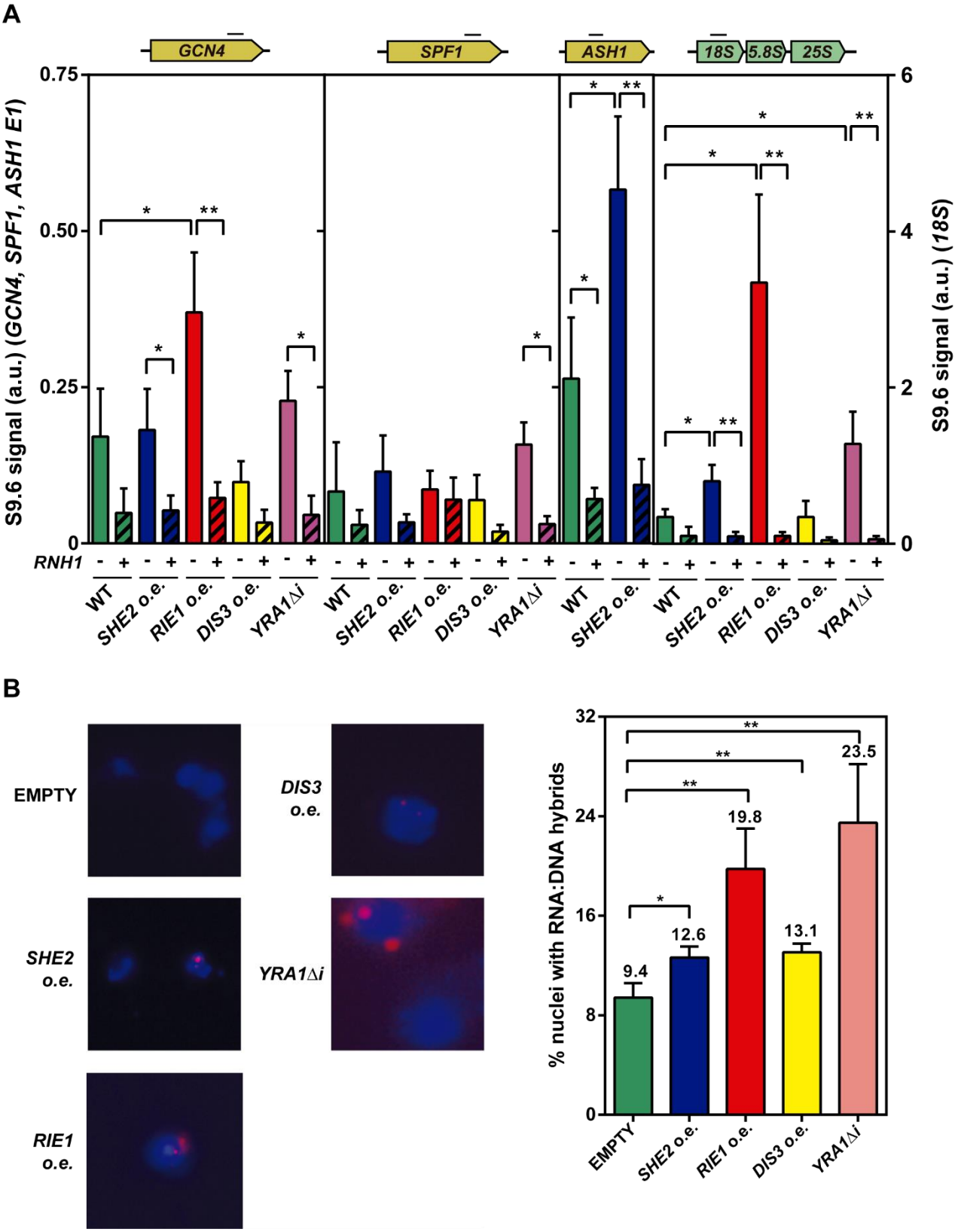


Figure R9. R-loops increase in cells overexpressing *SHE2* and *RIE1*.
(A) DNA:RNA immunoprecipitation (DRIP) with S9.6 antibody performed in W303-1A wild type cells overexpressing *SHE2*, *RIE1*, *DIS3* or *YRA1Δi* or the empty plasmid pYES2 (WT) at *GCN4*, *SPF1*, *ASH1* and *18S* genes. **(B)** Chromosome spreads S9.6 immunofluorescence performed in wild type yeast overexpressing *SHE2*, *RIE1*, *DIS3* or *YRA1*. Some representative pictures are represented on the left, with the total quantification in the right. Average and SD of at least three independent experiments are shown. *, p≤0,05; **, p≤0,01; (Student's t-test).

widespread R-loop accumulation. On the contrary, *DIS3* did not increase DNA:RNA hybrids in any of the genomic regions checked, not even in the ribosomal DNA where Dis3 is recruited to process the transcripts (Allmang *et al.*, 2000) (Figure R9-A).

To confirm the DRIP results and to have a more global vision of hybrid accumulation in the whole genome, we performed immunofluorescence experiments in chromosome spreads, using the S9.6 antibody that was immunodetected with a secondary Cy3- conjugated antibody. *RIE1* overexpression increased hybrids significantly (19.8% nuclei), supporting the more widespread effect in R-loop accumulation seen by DRIP. We also saw a slight, but significant, increment in hybrids with the overexpression of *SHE2* and *DIS3* (12.6% and 13.1% of nuclei, respectively). This suggest that an excess of Dis3 or She2 increment R-loops probably only at specific regions, the rDNA *18S* and *ASH1* in the case of She2. For Dis3, these regions are still to be determined. *YRA1* overexpression was included as a reference (Figure R9-B).

Overall, the DRIP and immunofluorescence data suggest that an excess of Rie1 induces the accumulation of DNA:RNA hybrids more globally than She2 or Dis3, that probably only affect specific regions. The increase in hybrids could explain the R-loop dependent increase in damage that we previously reported.

3.1.8. *DIS3* overexpression phenotype could be caused by exosome quenching

Dis3 is a catalytic subunit of the exosome with a central role in mRNA degradation and rRNA processing (Dziembowski *et al.*, 2007). One possible consequence of *DIS3* overexpression is that its overabundance could trigger its aggregation alone or together with other subunits of the complex, reducing the availability of a functional exosome. To test this hypothesis, we compared the effect of the overexpression and the mutation of *DIS3*. First, we check if inactivation of Dis3 using a thermosensitive allele (*dis3-ts*) and shifting the temperature to the non-permissive during 1 h could increase Rad52-YFP foci. We observed that the termosensitive *dis3-ts* mutant showed a significant increase (21.7%) of S/G2 cells with Rad52-YFP accumulation respect to the wild type without overexpression (10.4%). The increase was very similar to the effect that *DIS3* overexpression produces (20.4%). The increased damage in *dis3-ts* could be partially but still significantly reduced with *RNH1* overexpression (Figure R10-A), recapitulating the phenotypes observed for the overexpression.

In yeast, the absence of Dis3 leads to incorrectly processed mRNA accumulation and defective processing of the rRNA due to the inability of the exosome to degrade properly the RNAs, affecting mRNA decay rate (Davidson *et al.*, 2019; Milbury *et al.*, 2019) and accumulating rRNA intermediates (Allmang *et al.*, 2000). We wondered if *DIS3* overexpression could reproduce the same phenotype previously reported in the mutant. To achieve that, we analysed the effect of *DIS3* overexpression in rRNA processing by northern blot of 5.8S rRNA processing intermediates. We included a wild type strain with an empty vector, the YEpDIS3 overexpressing plasmid and the *dis3-ts* mutant grown at non-permissive conditions for 1 h. We found that both, *DIS3* overexpression and *dis3-ts* accumulate rRNA intermediate forms between the 7S and the 5.8S precursors, as previously described for the mutant (Schneider *et al.*, 2009). These intermediates could not be detected in the wild type strain, meaning that both, the lack and the overabundance of Dis3 leads to a rRNA processing defect (Figure R10-B).

Next, we performed a northern blot assay in which we transcribed the *GAL1* gene for 3 h and then we stop transcription by shifting the cultures to a glucose containing media, measuring the *GAL1* mRNA decay at different time points, as previously described (Rondón *et al.*, 2003). We found no differences in *GAL1* mRNA decay in either, overexpression or mutant conditions (Figure R10-C). We hypothesized that the absence of the expected phenotype in a wild type it is due because the 5'-3' and 3'-5' mRNA degradation pathways overlaps and, in order to find an increase in mRNA half-life, it is required to use a defective 5'-3' mutant like *dcp1* (Dziembowski *et al.*, 2000).

With the evidences that both, loss or excess of Dis3, increased damage in a similar manner, and both produced a defect in rRNA processing, we concluded that *DIS3* overexpression could be affecting the stoichiometry of the exosome complex, leading to its malfunction. Similarly to it was described for the *trf4* mutant (part of the TRAMP complex that interacts with the exosome for mRNA quality control), a reduction of a functional Dis3 could lead to accumulation of unstable and non-coding RNAs, that hybridize with the DNA contributing to R-loop formation (Gavaldá *et al.*, 2013). Furthermore, deregulation of cryptic unstable transcripts (CUTs) or antisense RNAs in exosome mutants could induce global transcriptional changes that have been reported to generate mitotic defects as a possible source of damage (Milbury *et al.*, 2019; Smith *et al.*, 2011).

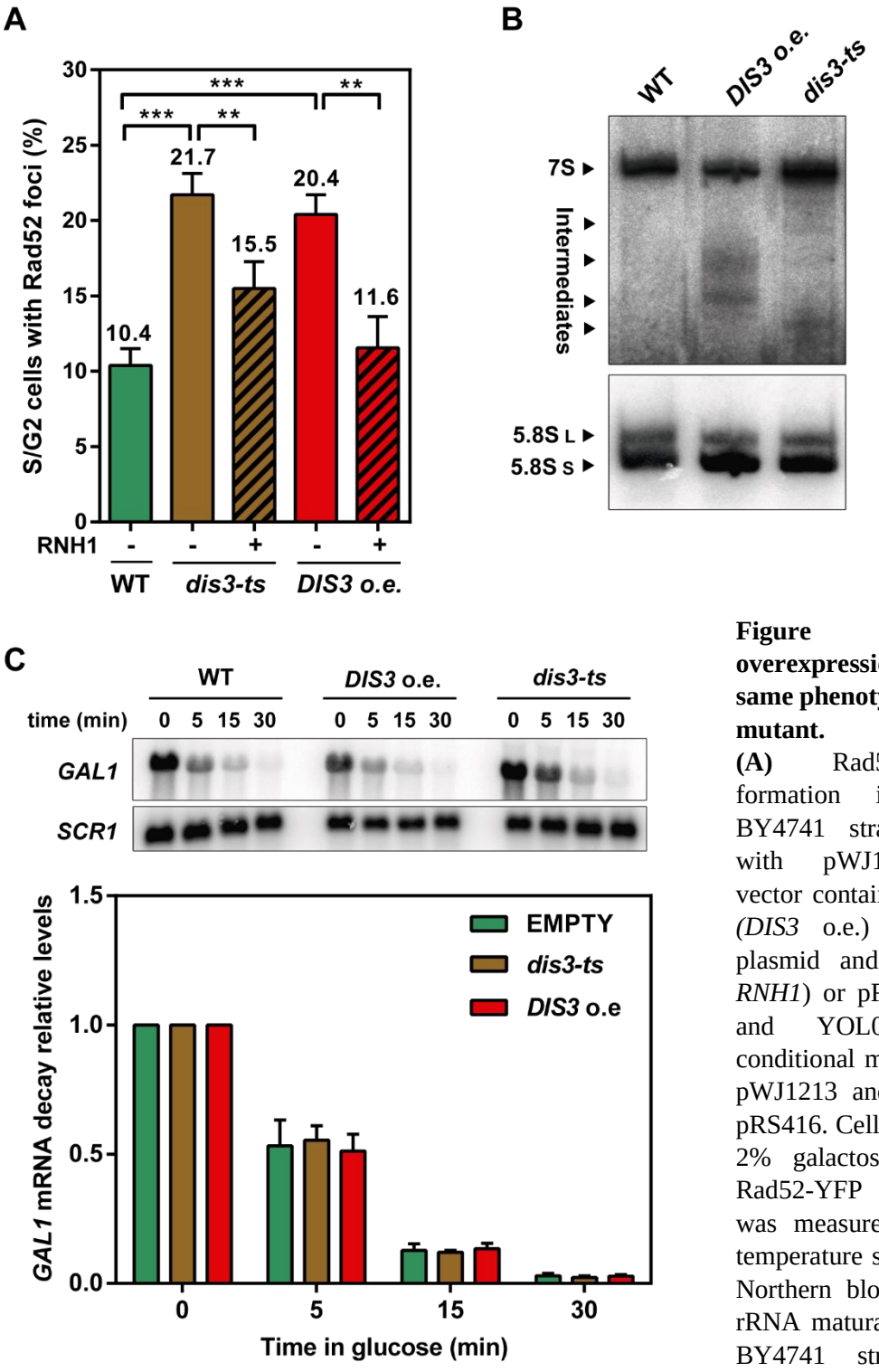


Figure R10. *DIS3* overexpression showed the same phenotype that the *dis3* mutant.

(A) Rad52-YFP foci formation in wild type BY4741 strain transformed with pWJ1213, YEp351 vector containing either *DIS3* (*DIS3* o.e.) or the empty plasmid and pGALRH1 (+ *RNH1*) or pRS416 (- *RNH1*) and YOL021C (*dis3-ts*) conditional mutant containing pWJ1213 and pGALRH1 or pRS416. Cells were growth in 2% galactose medium and Rad52-YFP foci formation was measured after 1 h of temperature shift at 37ºC. **(B)** Northern blot assay of 5.8S rRNA maturation products in BY4741 strain containing YEpDIS3 overexpressing plasmid or the empty one (WT) and the YOL021C (*dis3-ts*) mutant. The top panel including 7S was exposed overnight, while the bottom with the 5.8S was exposed 30 minutes. **(C)** mRNA *GAL1* decay northern blot after 1 hour of expression in galactose 2% at different times afterward washing and changing the cultures to glucose 2% containing media. The same strains that in the previous experiment were used. In both cases, the cultures were shifted to 37ºC for 1 hour to inactivate Dis3 in the thermosensitive mutant. Average and SD of three independent experiments are shown. **, p≤0,01; ***, p≤0,005 (Student's t-test).

3.1.9. *SHE2* and *RIE1* overexpression phenotypes differ from their respective mutants

Next, we decided to focus on studying the mechanism of the R-loop accumulation when either, RIE1 or SHE2 are overexpressed. First, we wanted to know if an excess of She2 or Rie1 is hampering their function either in the context of a complex in which they participate, as was the case for *DIS3* overexpression, or independently. She2p is part of the machinery that localizes specific mRNAs to the bud tip (Shen et al., 2010), while Rie1p is not described to form a complex. To compare the overexpression of these genes with their mutations, we measured the percentage of Rad52-YFP foci in wild type cells overproducing *RIE1* or *SHE2* and the null mutants of *rie1Δ* and *she2 Δ*. We found that, contrary of what we saw with *dis3-ts*, *she2Δ* or *rie1Δ* mutants did not increase Rad52-YFP foci (7.2% and 9% respect to the 7% of the wild type), in contrast to the overexpression of *SHE2* or *RIE1* (Figure R11).

Due to the role of She2 in mRNA transport, we decided to test if its overexpression could lead to non-specific binding to mRNAs, thus impairing globally mRNA export. To check that, fluorescence in situ hybridization (FISH) was performed using an oligo dT Cy3 fluorescent probe to detect total polyadenylated mRNA in the cells. The overexpression of *SHE2* did not show any effect of the mRNA distribution compared to the wild type, in contrast to the *mex67* mutant that presented a strong mRNA export

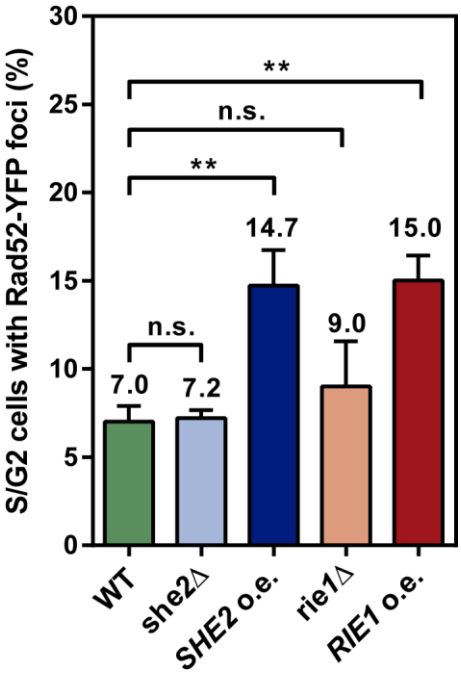


Figure R11. *she2Δ* and *rie1Δ* mutations do not increase Rad52-YFP foci.
Rad52-YFP foci formation in wild type BY4741 strain transformed with pWJ1213, pYES2 vector containing either *SHE2, RIE1* or the empty plasmid (WT) and YKL130C (*she2Δ*) or YR250C (*rie1Δ*) BY4741 mutant strains containing pWJ1213. Cells were growth in 2% galactose medium. Average and SD of at least three independent experiments are shown. **, p≤0,01; (Student's t-test).

defect with mRNA accumulation in the nucleus (Figure R12), as was previously reported (Estruch *et al.*, 2012).

Therefore, we concluded that, in the case of the overexpression of *SHE2* and *RIE1*, the genomic instability was not caused by interference with the processed in which they participate. Moreover, *SHE2* overexpression did not impair global mRNA export in the cell.

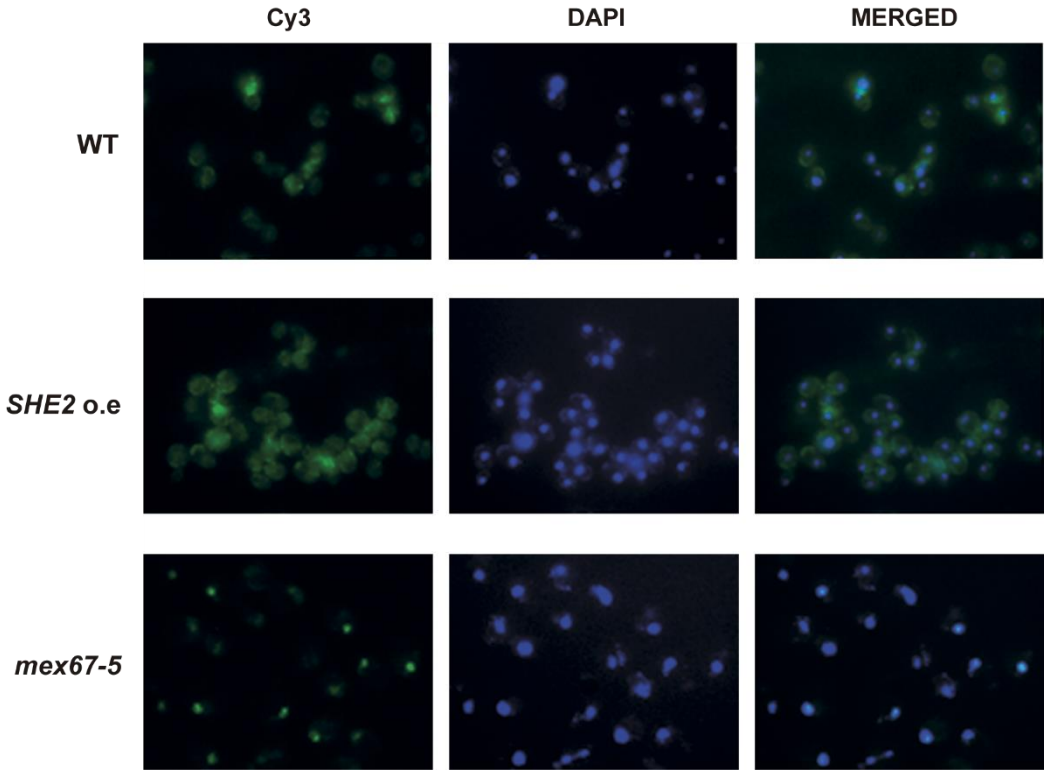


Figure R12. mRNA nuclear export is not affected by *SHE2* overexpression.
Localization of poly(A) RNA in W303-1 wild type overexpressing *SHE2* or transformed with the empty pYES2 vector (WT). WMC1-1A (*mex67–5*, nuclear export defective mutant) was included as positive control. RNA was detected by *in situ* hybridization with Cy3-labeled oligo(dT) (in green). Detection of the nuclei was determined by DAPI (blue) staining.

3.1.10. Overexpression of *RIE1* and *SHE2* does not affect global transcription

R-loops are proposed to have a role in transcription, pausing the RNA polymerase, especially at the promoter and during termination (Aguilera & García-Muse 2012). Beyond physiological levels, it has been suggested that excess of Yra1 stabilizes DNA:RNA hybrids and leads to transcription-dependent genome instability (García-Rubio *et al.*, 2018). Similarly to Yra1, She2 and Rie1 are RNA-binding proteins whose overexpression increase R-loops. We wondered if the increase in R-loop that they produce

could lead to transcription defects. As a first approach, we checked if the overexpression of these two genes could genetically interact with transcription mutants. To achieve that, we performed drop test assays with *spt4* and *dst1* mutants. Spt4 is a transcription-elongation factor, part of the DSIF complex, that regulates the processivity of RNA polymerase II (Hartzog *et al.*, 1998), while *DST1* encodes for TFIIS, a factor required to rescue backtracked RNAPIIs (Davies *et al.*, 1990). While *SHE2* overexpression did not present any differences in the mutants compared with the wild type, the *spt4* mutant partially supressed the growth defect observed by *RIE1* overexpression (Figure R13-A).

To confirm if these interaction with *RIE1* overexpression was specific of *spt4* mutant or whether it is extensive to other transcription factors, we also tested its overexpression in other transcription mutants: *spt16* (component of the FACS complex

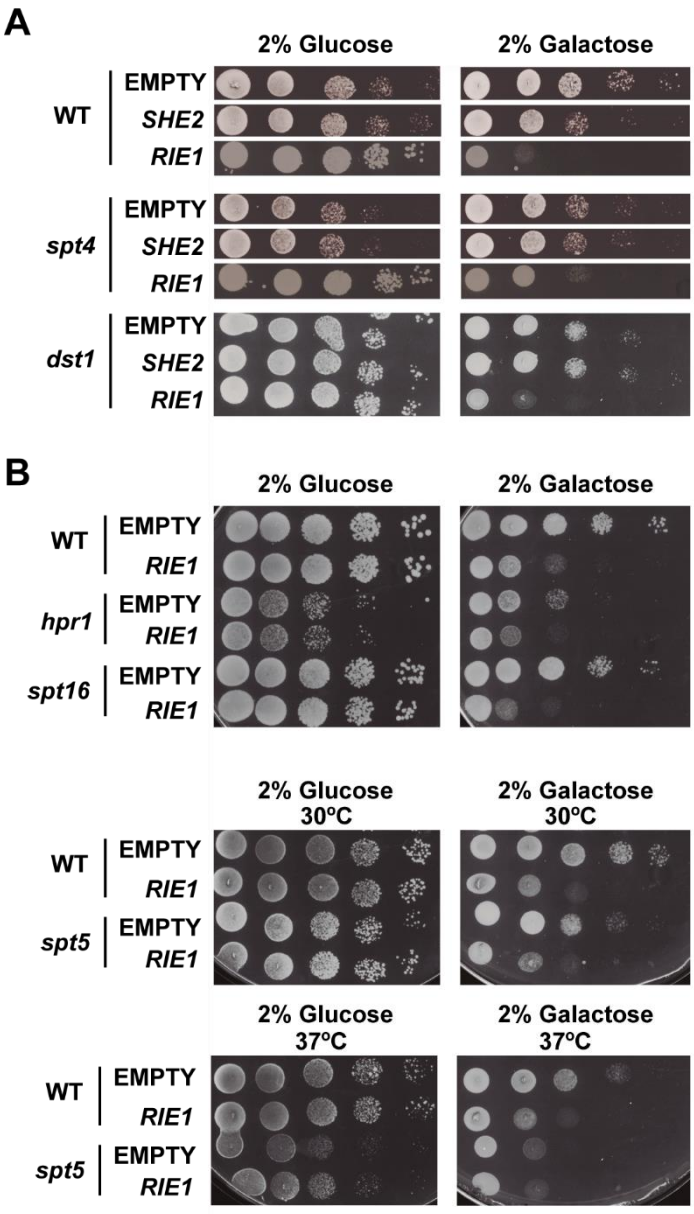


Figure R13. Spt4 transcription factor genetically interacts with *RIE1* overexpression.
(A) Drop test assay in YGR063c (*spt4*Δ), YGL043w (*dst1*Δ) or WT (BY4741) strains without (glucose 2%) or with (galactose 2%) *SHE2* or *RIE1* overexpression from pYES2 plasmid. The empty plasmid (EMPTY) was included as control. **(B)** Drop test assay performed with the same conditions as before, overexpressing *RIE1* or the empty plasmid control in W303 wild type (WT), SChY58a (*hpr1Δ*) DY8107 (*spt16*) and in the conditional mutant GHY94 (*spt5*). For spt5 the inactivation was achieve by a temperature shift to 37ºC.

that facilitates the access to DNA to the RNAPII), *spt5* (the other subunit of the DSIF complex) and *hpr1* (a mutant of the THO complex). We did not observe suppression of the growth defect induced by *RIE1* overexpression in any of these mutants (Figure R13-B). This could point to a genetic interaction between Rie1 and Spt4 or to the fact that Spt4 is required for the deleterious effects produced by *RIE1* overexpression.

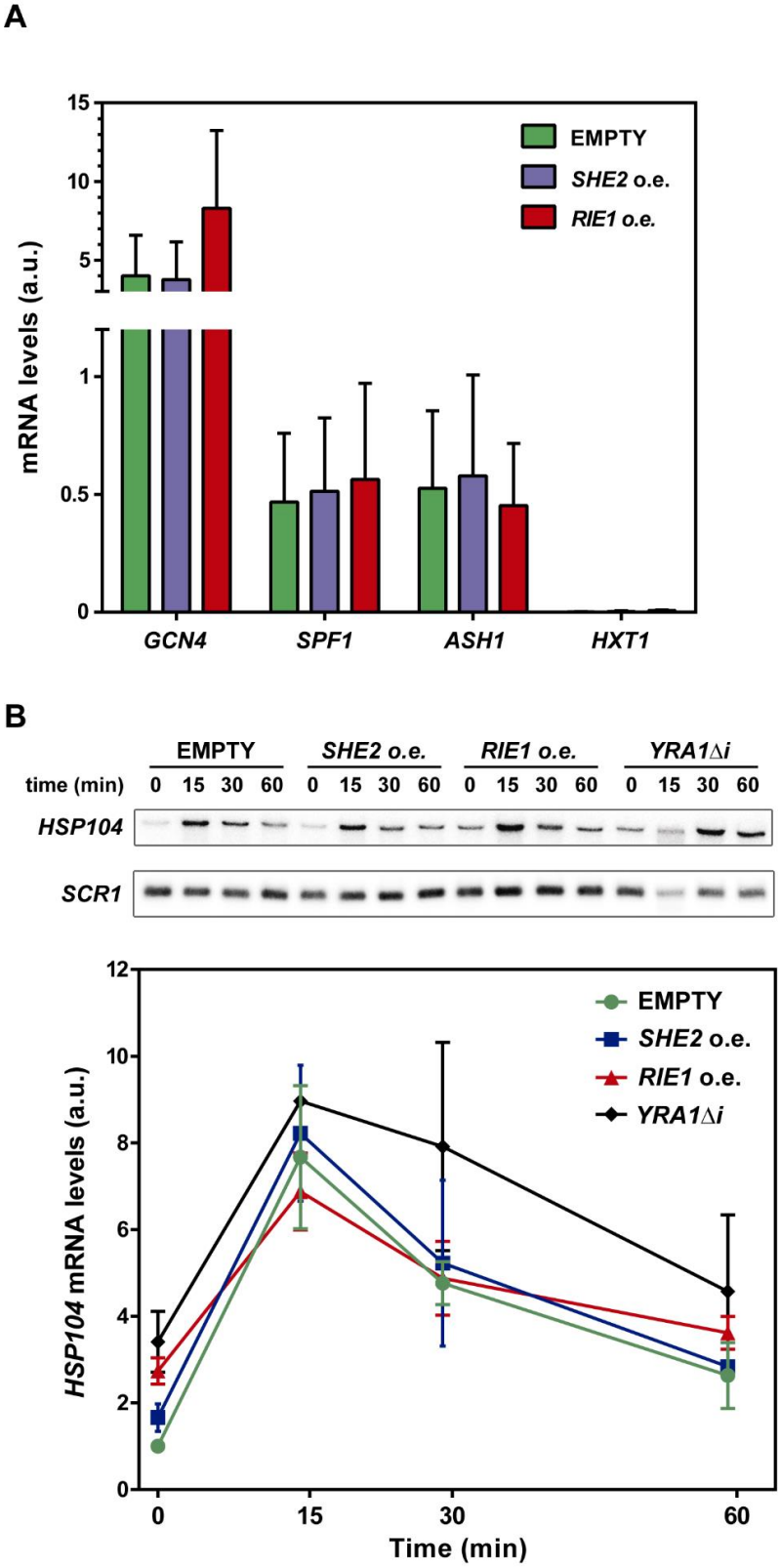


Figure R14. General transcription is not affected by *RIE1* or *SHE2* overexpression. **(A)** mRNA levels of *GCN4*, *SPF1* *ASH1* or *HXT1* (as a negative control) in W303-1A strain overexpressing *SHE2*, *RIE1* or none of them (EMPTY). The mRNA levels were measured by RT-qPCR. **(B)** *HSP104* mRNA induction measured by northern blot assay in W303-1A yeast overexpressing *SHE2*, *RIE1* or *YRA1Δi* at the indicated times after the heat-shock at 37ºC. Average and SD of at least three independent experiments are shown.

Next, we checked if the increased levels of She2 or Rie1 affect the transcription of the selected genes that has been previously analysed for R-loop accumulation: *GCN4*, *SPF1* and *ASH1* (as one of the characterized genes that specifically requires She2 to localize its transcripts to the bud tip). *HTX1* was also included as a negative control, as it is repressed in galactose containing media. We extracted RNA after overnight overexpression of *SHE2* and *RIE1* and quantified the transcripts by RT-qPCR. None of the candidates affected significantly the mRNA levels of any of these targets (Figure R14-A).

Finally, to investigate in more detail a possible defect in transcription elongation, we decided to measure at short times the transcripts of an inducible gene, as the mRNA levels of a constitutive gene depends not only in the transcription but also in its decay rate. To do that, we measured by northern blot assay the transcripts of the heat-shock inducible *HSP104* gene at different time points, overexpressing *SHE2* or *RIE1* overnight at 26ºC and then inducing *HSP104* transcription by shifting the cultures to 37ºC for a short time. The northern blot did not show differences between the empty plasmid control or the strains overexpressing *SHE2* or *RIE1*. *YRA1* overexpression was included as a reference, showing a small but not significative increase in the mRNA levels of *HSP104* could be observed (Figure R14-B).

We concluded that the overexpression of either, *SHE2* or *RIE1* did not affect the transcription of the genes that accumulate hybrids, probably because they do not interfere directly with the transcription itself.

3.1.11. *RIE1* enters to the nucleus when overexpressed, while *SHE2* is recruited to chromatin

The overexpression of both, *SHE2* and *RIE1* increased R-loop and induced DNA damage. We wonder to know if She2 or Rie1 proteins are recruited to the genes that accumulates DNA:RNA hybrids, or if the increase in R-loops is produced indirectly. To assess that, we fused both proteins to a YFP tag, that allow us not only to immunoprecipitate them but also to see the localization of the proteins *in vivo*. The fused products were cloned under *GAL1* promoter control. The recruitment of She2-YFP or Rie1-YFP to the chromatin was assessed by chromatin immunoprecipitation (ChIP) with an anti-GFP antibody after 3h of overexpression in galactose. We studied localization to the genes that

accumulated R-loops with excess of these two proteins: *GCN4*, *SPF1*, *E1* and *E3 ASH1* zipcode regions, the ribosomal gene *18S* and *HXT1* as a negative control of transcription.

We found that overexpressed She2-YFP signal at all the genes analysed was higher than the strain that that express YFP epitope alone. This increase was statistically significant in *GCN4* and *ASH1*, even do the tendency to recruit it could still be observed in all the genes but *HXT1*. This suggested that She2-YFP was recruited to chromatin in a transcription-dependent manner, in agreement with previous data showing that She2 travels with the elongating form of the RNA polymerase II (Shen *et al.*, 2010). Contrarily, Rie1-YFP overexpression did not show an increase over the YFP epitope control. This could indicate that Rie1 protein did not interact directly with the chromatin (Figure R15).

We also checked the localization of the two proteins *in vivo* after 3 h of induction in overexpressing conditions (2% of galactose) or reducing the expression level adding glucose to the medium (2% galactose + 0.10% glucose). She2-YFP protein was localized in the cytoplasm and nucleus, with a stronger nuclear signal, in both conditions,

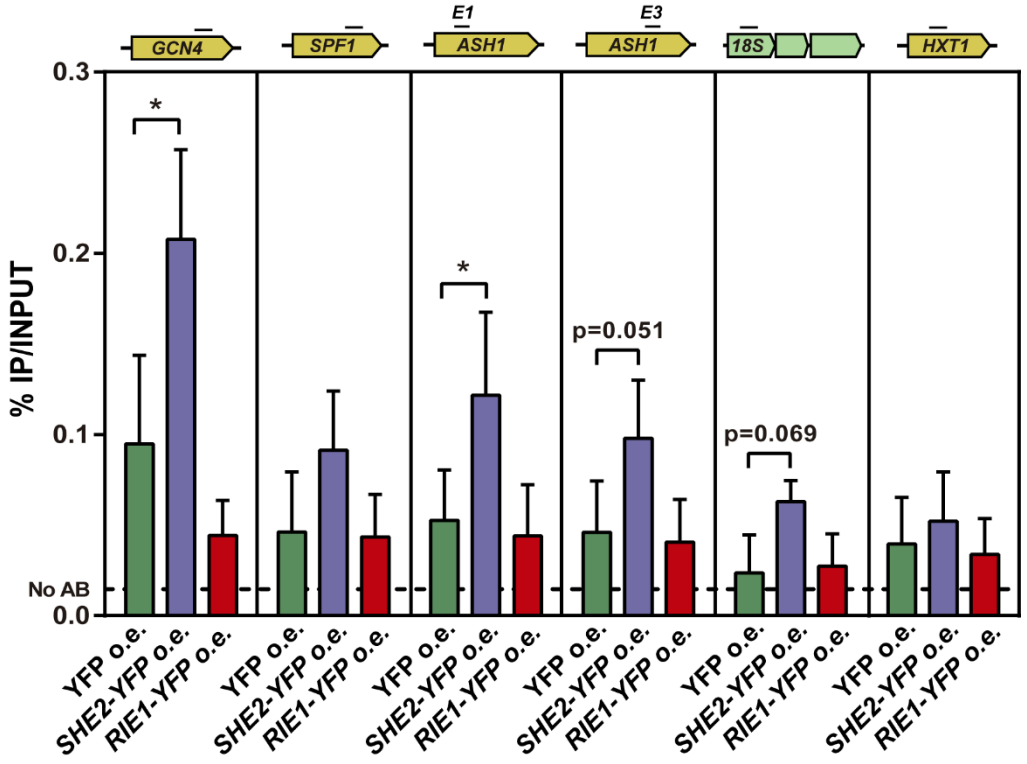


Figure R15. Overexpressed She2p is recruited to chromatin, while Rie1p is not.
Recruitment of overexpressed She2-YFP, Rie1-YFP or YFP in a wild-type background measured by ChIP using an anti-GFP antibody in *GCN4*, *SPF1*, *ASH1 E1* and *ASH1 E3* zipcode regions, *18S* rDNA or *HXT1* negative control. Dashed line indicates the no-antibody threshold. Average and SD of at least three independent experiments are shown. *, p≤0,05; (Student's t-test).

overexpressed (97% of cells with nuclear She2-YFP signal) or with its expression reduced (98.1%), corresponding with it has been described for the endogenous levels of the protein. With 0.10% of glucose, Rie1-YFP only could be observed in the cytoplasm (3.9% of cell with nuclear signal), but in overexpressing conditions some cells presented signal into the nucleus in wild type (19.6%) and *hpr1* mutant (34%). As Rie1 abundance increases in replicative stress conditions (Tkach *et al.*, 2012), we wondered if the addition of HU to the medium could also affect the localization of the overexpressed protein. The amount of Rie1-YFP in the nucleus upon 100mM HU treatment was similar to the localization when overexpressed in both, wild type (22.1%) and *hrp1* mutant (36.2%), so the replicative stress itself did not induce the shift of the protein to the nucleus, and it may be just an expression regulatory mechanism for the gene at endogenous levels (Figure R16-A).

With these results, we could conclude that the overexpressed She2 protein accumulates in the nucleus in most of the cells, as reported when it is expressed at endogenous levels. Even if its distribution did not change, She2p excess in the nucleus is recruited to chromatin, producing an accumulation of DNA:RNA hybrids at certain genomic regions and DNA damage. Rie1, on the contrary, appeared excluded from the chromatin, but it is localized in the nucleus when it is overexpressed, increasing R-loops and DNA damage. Rie1 does not have any nuclear localization sequence (NLS), so its presence in the nucleus should require an adaptor protein, or maybe its overexpression leads to a leaky entry. It would be interesting to introduce an NLS into Rie1 in order to test if its localization to the nucleus is enough to reproduce the overexpression phenotype.

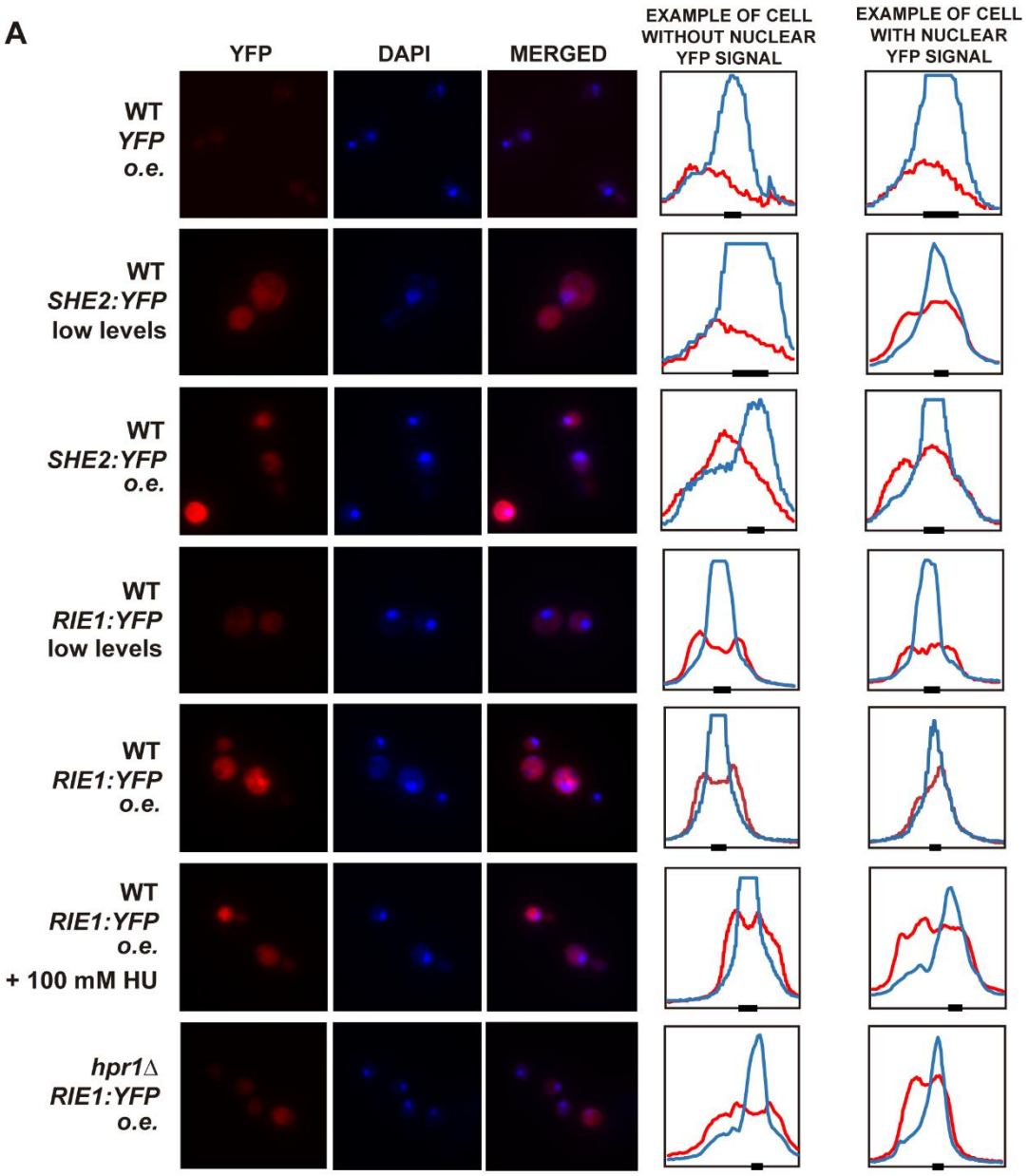


Figure R16. Overexpressed Rie1p localized to the nucleus.

(A) She2-YFP, Rie1-YFP and YFP control localization after 3 hours of overexpression in 2% galactose (overexpressed) or expressed at lower levels in 2% galactose, 0.10% glucose medium (low levels) in W303-1A wild type or HPBAR1-R (*hpr1Δ*). DNA was stained with DAPI. A representative image of each condition is showed (right), with an illustrative example of the quantification of one single cell with or without YFP nuclear signal in each case (left). DAPI is marked in blue, while YFP signal is in red. The nucleus position corresponds to the x-axis black box. These graphs were generated with LasX software. **(B)** Quantification of at least 100 cells per sample represented.

3.2. Effect of R-loops and ssDNA damage on the elongating RNAPII

The fact that the RNA polymerase II requires an intact DNA molecule as a template implies that DNA damage will affect transcription. Indeed, different groups have addressed the consequences of UV light-induced damage, that causes a shutdown of transcription and the loss of RNAPII processivity (Gaillard *et al.*, 2007; Jansen *et al.*, 2002); or the effect of double-strand breaks (DSBs), that results in RNAPII stalling by chromatin condensation followed by RNAPII removal (Pankotai *et al.*, 2012; Shanbhag *et al.*, 2010). However, little is known on how other possible sources of DNA damage affects the transcription and what is the fate of an RNAPII in those situations. In this context, we decided to focus on how non-B DNA structures as R-loops or how single-strand breaks (SSB) affect transcription and the consequences on the genome.

3.2.1. Generation of the *GAL1p:LYS2* transcription system

First, we designed a molecular system that would allow us to study the effect of SSBs and transcription-associated non-canonical DNA structures as R-loops over RNAPII under conditions in which we could control transcription. We selected the yeast *LYS2* gene, since it is a long (4.18 Kb) non-essential gene, that does not accumulate RNAPII at any specific region, as was reported by ChIP-seq with an Rpb3 specific antibody (Gómez-González *et al.*, 2011), a feature that suggests the absence of intrinsic structures that may affect RNAPII elongation. However, the endogenous transcription level of *LYS2* is low even in the absence of lysine, as measured by RNAPII ChIP using 8WG16 Rpb1 specific antibody (Figure R17-A). In order to increase the transcription in an inducible manner, the *LYS2* promoter was replaced by the *GAL1* promoter, whose expression could be induced adding galactose and silenced with glucose, generating a yeast strain with the *GAL1p:LYS2* system integrated in its genome. To select integrative colonies, a *NATnt2* resistance cassette was also included upstream of the *GAL1* promoter (Figure R17-B). Positive colonies were grown in nourseothricin-containing medium and integration was checked by PCR (Figure R17-C).

In order to test if the expression of *LYS2* responds to galactose in the new strain generated, we streaked it in medium without lysine and with either galactose or glucose. As expected, in the absence of lysine, the strain grew in plates with 2% galactose but it was unable to grow in 2% glucose plates, verifying that the *GAL1* promoter was not leaky. The wild-type control strain, with endogenous *LYS2* gene, grew in both carbon sources (Figure R17-D). In order to determine the level of transcription of *GAL1p:LYS2*, we

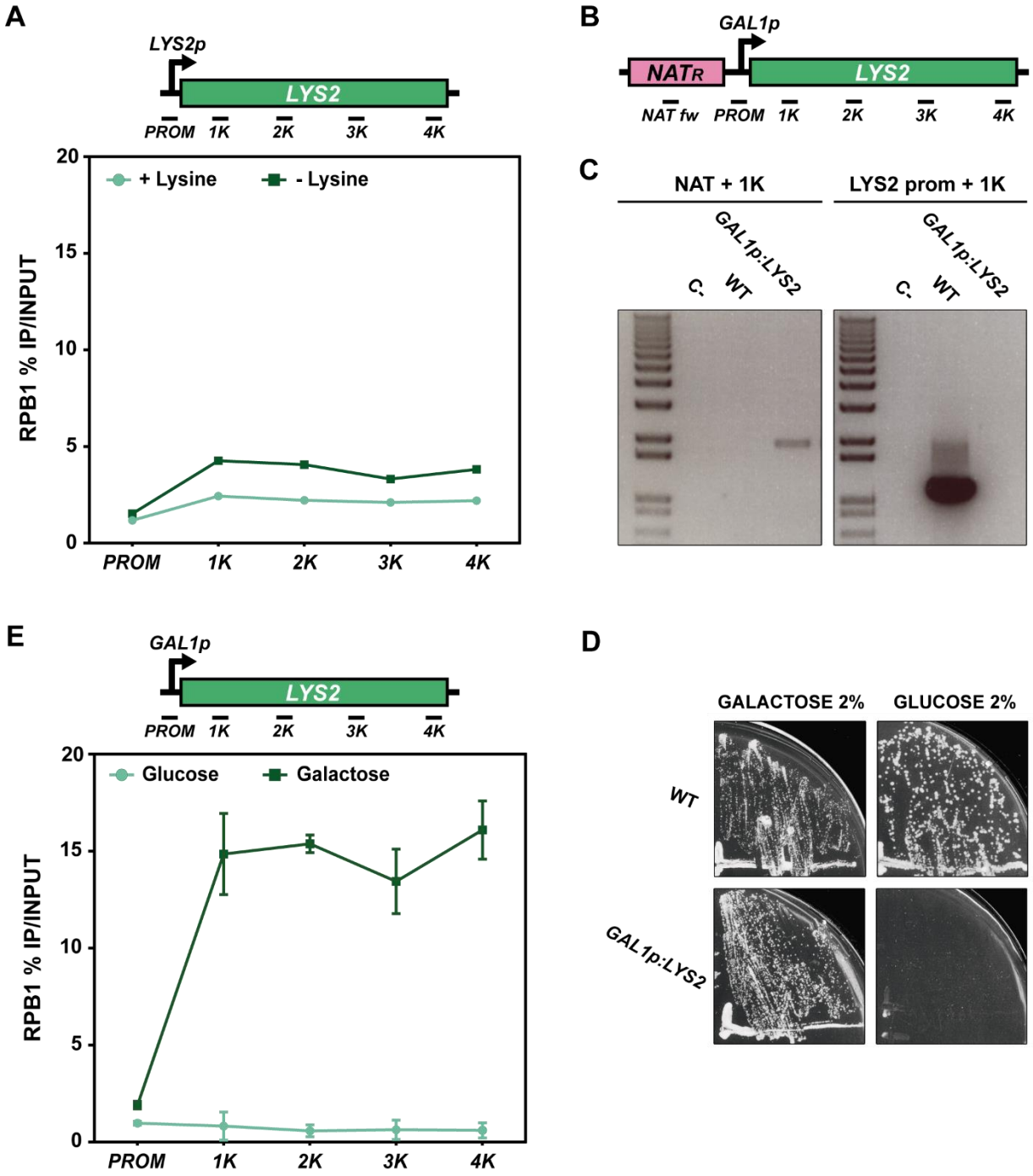


Figure R17. Design of the *GAL1p:LYS2* system.
(A) RNAPII ChIP profile in a W303-1A wild type strain growth with or without lysine. **(B)** Draft of the *GAL1p:LYS2* system with the *LYS2* promoter replaced by a *NATnt2-GAL1p* cassette. Regions amplified by qPCR are depicted. **(C)** PCR from genomic DNA of the *GAL1p:LYS2* strain to verify the integration of the cassette (NAT + 1K) replacing the endogenous promoter (LYS2 prom + 1K) was amplified from W303-1A (WT) and GLY-2D (*GAL1p:LYS2).* A negative control without DNA (C-) was also included. **(D)** GLY-2D (*GAL1p:LYS2)* and W303-1A strains were grown in plates without lysine and with either 2% galactose (transcription induced) or 2% glucose (transcription represed). **(E)** RNAPII profile at the *GAL1p:LYS2* gene measured by ChIP in GLY-2D (*GAL1p:LYS2)* strain growth in 2% galactose or 2% glucose medium. Average and SD of three independent experiments are shown.

carried out a RNAPII ChIP. We detected no RNAPII in *LYS2* in 2% glucose containing media and a high RNAPII occupancy after 1 hour of 2% galactose induction. Moreover, this high induction of transcription did not seem to affect elongation as we observe a flat RNAPII profile in the *GAL1:LYS2* gene (Figure R17-E).

3.2.2. Generation of an R-loop accumulating system: *GAL1p:LYS2:Sμ350*

To study the effect of the DNA:RNA hybrids on the elongating RNAPII, a sequence that is prone to form R-loops, *Sμ350*, was cloned in *GAL1p:LYS2*. *Sμ350* is a short (350 bp) sequence from immunoglobulin class-switch region of murine B-cells that contains 14 repetitions of a conserved GC-rich motif in tandem. The sequence has been previously reported to form R-loops and G-quadruplex on the opposite strand both, *in vivo* and *in vitro* (Ruiz *et al.*, 2011; Tornaletti *et al.*, 2008; Duquette *et al.*, 2004). G-quadruplex are helical non-canonical DNA structures formed by guanine tetramers that may favour the formation of R-loops in the opposite strand and may regulate transcription, as reported in the promoter of the c-Myc gene (Simonsson *et al.*, 1998).

In order to integrate *Sμ350* sequence into *GAL1p:LYS2*, we used the Cas9 system. First, we amplified the *Sμ350* sequence from the plasmid pRS413-SF introducing 50 pb of homology with *LYS2* at both sides of the resulting cassette. Next, we cloned a specific guide RNA (gRNA) in pML104 plasmid, that already contains the *cas9* gene. This gRNA directs the Cas9 enzyme to the position 2942 of the *LYS2* ORF. Finally, we co-transformed the *GAL1p:LYS2* strain with the pML104 plasmid (that express Cas9 and the gRNA) and the cassette containing the *Sμ350* sequence with the *LYS2* homology. In the cell, Cas9 induced a DSB directed by the gRNA that was repaired by homologous recombination with the *Sμ350* cassette, integrating it in the designed position of *LYS2* gene (Figure R18-A). The transformants were growth in plates without uracil to maintain the selection of the Cas9-expressing plasmid. Only the cells that integrate the *Sμ350* cassette will lose the gRNA target site and will be able to grow. This provides a selection method without introducing any additional marker. The obtained strains, containing the *GAL1p:LYS2:Sμ350* system, were analysed by PCR, obtaining 7 positive colonies that were sequenced to verify *Sμ350* integration. We keep working with the clone GLSd-1B, that did not contain further mutations (Figure R18-B).

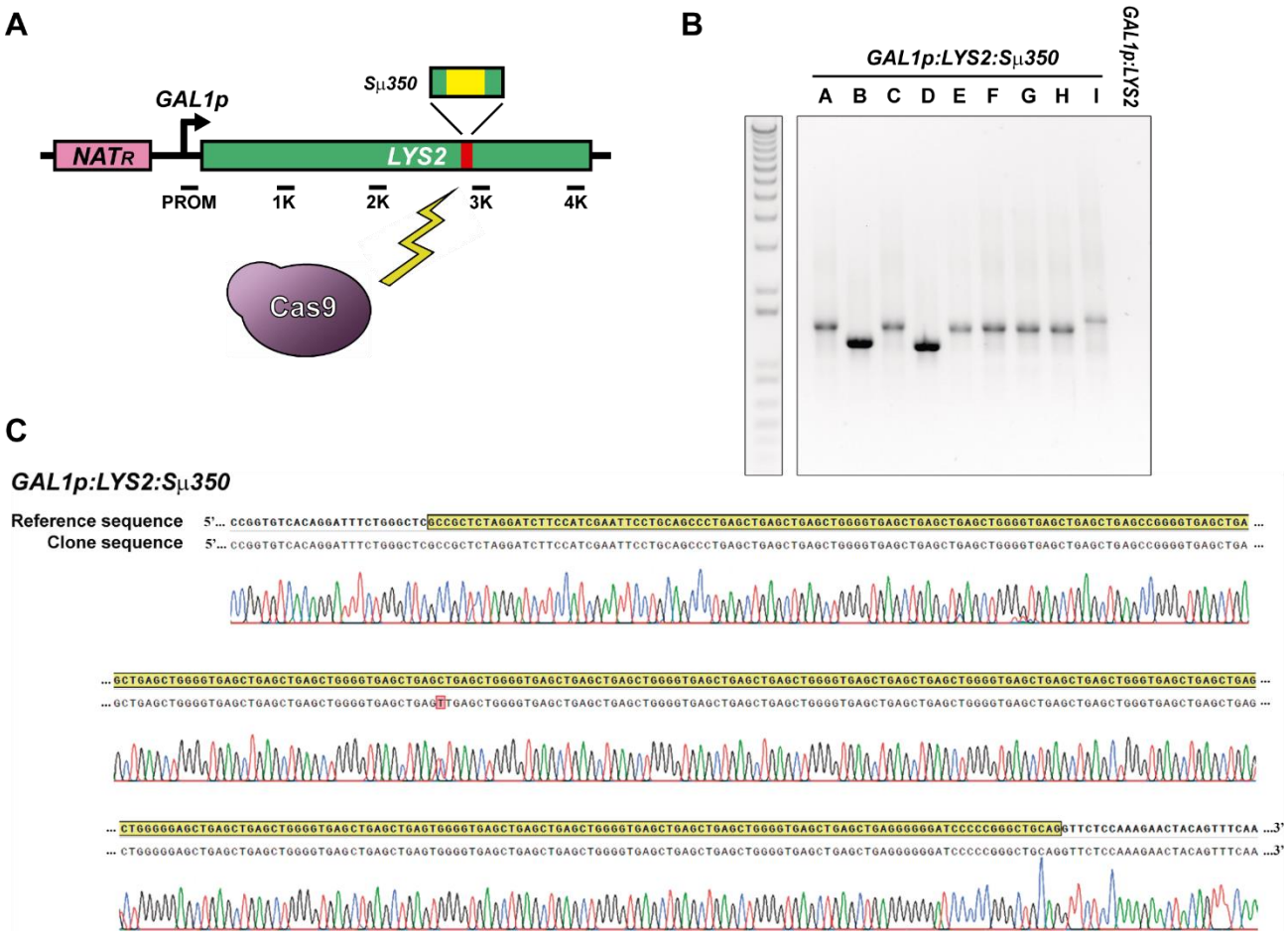


Figure R18. Generation of the *GAL1p:LYS2:Sμ350* system.
(A) *GAL1p:LYS2* system in which a DNA fragment containing *Sμ350* sequence was introduced by Cas9-directed homologous recombination. **(B)** PCR from genomic DNA from different *GAL1p:LYS2:Sμ350* clones to verify the integration of the *Sμ350* cassette, using 'Sm350 compr' and 'LYS2 4K rv' primers. GLY-2D (*GAL1p:LYS2*) was included as control. **(C)** Aligned map of the *Sμ350* inserted in the *LYS2* gene, obtained by sequencing (bottom) of the GLSd-1B clone (*GAL1p:LYS2:Sμ350-H)* and the reference sequence (top).

3.2.3. The *GAL1p:LYS2:Sμ350* system accumulates DNA:RNA hybrids

To check if the *GAL1p:LYS2:Sμ350* system accumulates R-loops as intended, we performed a DNA:RNA hybrid immunoprecipitation experiment (DRIP) in *GAL1p:LYS2* and *GAL1p:LYS2:Sμ350* strains either in wild type, *hpr1* (Figure R19-A) or *rnh1 rnh201* (Figure R19-B) mutant backgrounds. We choose these mutants because both have been previously reported to accumulate R-loops: *hpr1*, due to a defect in mRNP biogenesis (Huertas & Aguilera 2003); and the double mutant *rnh1 rnh201* that is unable to eliminate these structures (Wahba *et al.*, 2011). We observed a significative increase in R-loops in the *LYS2* region that contains the *Sμ350* sequence in *GAL1p:LYS2:Sμ350* compared to the same region in the *GAL1p:LYS2* system. This R-loop accumulation was detected in

all the strains: wild type, *hpr1* and *rnh1 rnh201*. Moreover, the increase could not be appreciated in the 5' region of the *LYS2* gene, indicating that R-loop increment was specific of the *Sμ350*. A control region that is not transcribed (an intergenic region of the

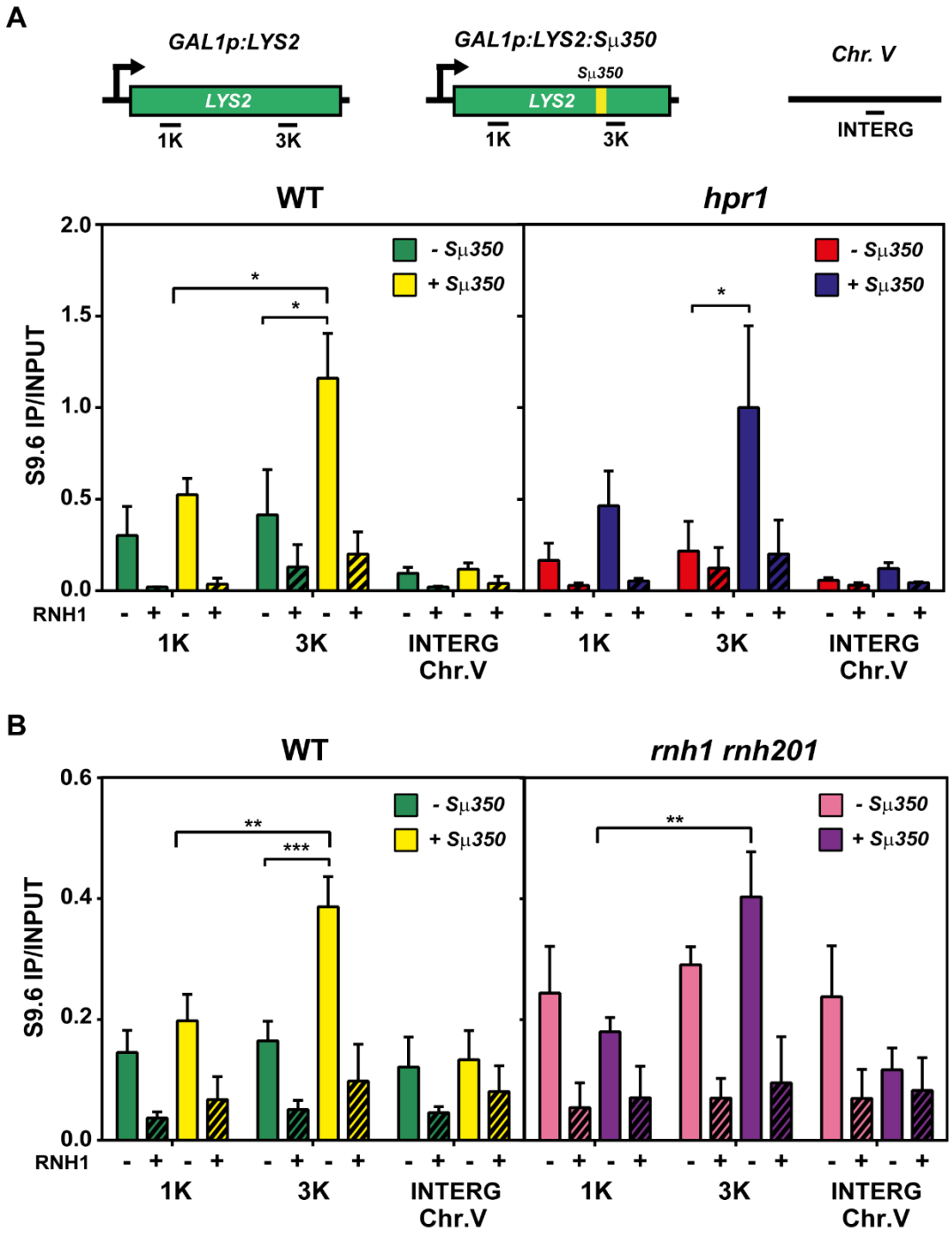


Figure R19. R-loops accumulates in the *GAL1p:LYS2:Sμ350* system.
DNA:RNA immunoprecipitation (DRIP) with S9.6 antibody performed in *GAL1p:LYS2* (- *Sμ350*) or *GAL1p:LYS2:Sμ350* (+ *Sμ350*) systems in wild type (WT), *hpr1* **(A)** or *rnh1 rnh201* **(B)** strains after 2 hours of transcription induction. A region upstream of the *Sμ350* site (1K) and other adjacent to the *Sμ350* site (3K) were tested. The chromosome V intergenic region (INTERG Chr.V) was included as a no-transcription control. Average and SD of at least three independent experiments are shown. *, p≤0,05; **, p≤0,01; ***, p≤0,005 (Student's t-test).

chromosome V) was included. The level of RNA:DNA hybrids detected was similar in all the strains analysed suggesting that maybe the *Sμ350* sequence itself is enough to induce a strong accumulation of R-loops. In any case, the DRIP experiment demonstrated that the *GAL1p:LYS2:Sμ350* system accumulates DNA:RNA hybrids specifically in the region containing the *Sμ350* sequence in the transcribed *LYS2* gene and this increment in R-loops was high enough to be detected even in a wild type background.

3.2.4. Steady-state RNAPII profile does not change in *GAL1p:LYS2:Sμ350*

First, to study how R-loops could challenge transcription, we measured RNAPII distribution through *LYS2* gene by ChIP in the systems without (*GAL1p:LYS2*) or with *Sμ350* (*GAL1p:LYS2:Sμ350*) in a wild-type strain or *hpr1* and *rnh1 rnh201* mutants. After inducing transcription for 2 h by addition of galactose to the media, the RNAPII level was measured using the 8WG16 antibody. Transcription was induced in G1 arrested cells to avoid transcription-replication conflicts that could interfere with the result. We detected a slight, but significative, increase in RNAPII level in the *hpr1* mutant at the region upstream to *Sμ350* in *GAL1p:LYS2:Sμ350* compared to the same region in the *GAL1p:LYS2* system (Figure R20-A). Moreover, the *rnh1 rnh201* mutant carrying the *Sμ350* sequence depicted higher level of RNAPII along the entire gene, although this increase was not significant (Figure R20-B). These results suggested that R-loop formation could affect RNAPII elongation, but this is difficult to detect measuring steady-state transcription. We also found that the total amount of RNA polymerase II was lower in the *hpr1* mutant than in the wild type, regardless of the presence of *Sμ350*. This result suggests that the *hpr1* mutant presents elongation defects that are R-loop independent but that are aggravated by these structures probably causing the accumulation of RNAPII detected. It would be interesting to repeat the experiment overexpressing RNase H to check the implication of RNA:DNA hybrids in RNAPII accumulation in *hpr1* mutant.

3.2.5. RNA polymerase II elongation rate is reduced on the *GAL1p:LYS2:Sμ350* in an R-loop dependent manner

As mentioned earlier, it is difficult to detect defects in elongation measuring steady-state RNAPII as the mechanisms to remove blocked polymerases could be masking them.

Therefore, we decided to examine the progression of the RNA polymerases immediately after transcription is stopped, allowing us to measure the elongation rate that could not be evaluated during a steady-state transcription. For that, we induced *LYS2* transcription for

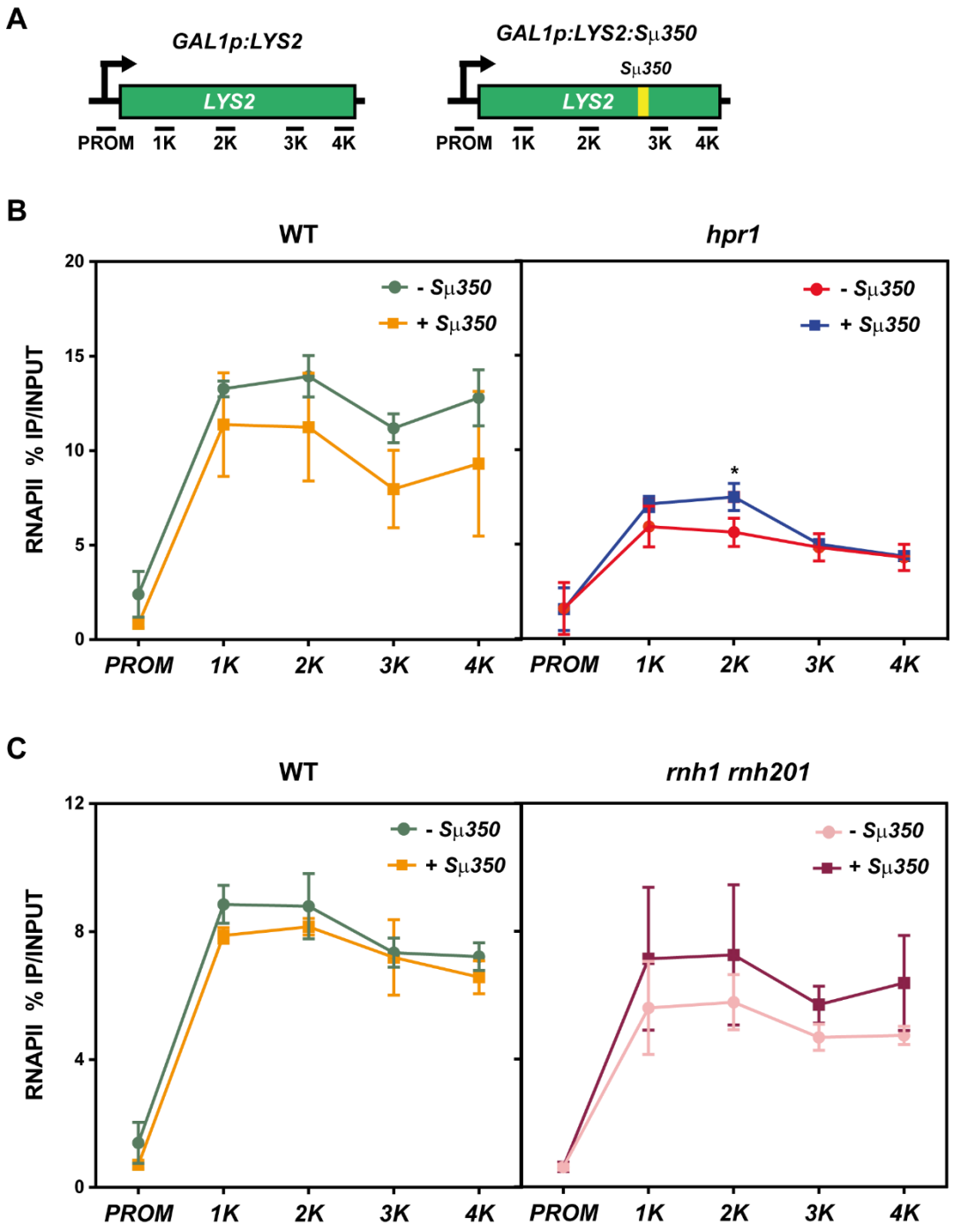


Figure R20. RNAPII profile in the *GAL1p:LYS2:Sμ350* system.
RNAPII profiles in the *LYS2* gene measured by Rpb1 ChIP in *GAL1p:LYS2* (- *Sμ350*) or *GAL1p:LYS2:Sμ350* (+ *Sμ350*) systems, in wild type (WT), *hpr1* **(B)** or *rnh1 rnh201* **(C)** strains after 2 hours of transcription induction. Regions analysed are depicted **(A)**. Average and SD of three independent experiments are shown. *, p≤0,05 (Student's t-test).

2 h in G1 arrested cells, to avoid transcription-replication conflicts, and then we switched it off by adding glucose to the medium. Chromatin was crosslinked at very short time points after glucose addition (0, 2, 4, 6 and 8 min, respectively). Wild type and *hpr1* strains containing either *GAL1p:LYS2* or *GAL1p:LYS2:Sμ350* system were analysed. We found an accumulation of RNAPII in the strains carrying the *Sμ350* sequence that was significant at very short times (2 min after shutting down transcription). This built up of RNAPII could be appreciated in the wild type strain but was higher in the *hpr1* mutant. Also, the increment in RNAPII was higher downstream the *Sμ350* sequence (Figure R21). The RNAPII profile in the *GAL1p:LYS2:Sμ350* system suggests an elongation defect in the regions adjacent to *Sμ350* sequence. Interestingly, the RNAPII level did not decrease over time in a distance proportional manner, as we would expected that it will be reduced first at the 5' of the gene and finally in the 3' as RNAPII elongates through the gene. Instead of that, we detected a general RNAPII drop after its accumulation that was similar throughout the *LYS2* regions. This result leads us to think that maybe transcription could be prematurely terminating, or the RNAPII was being removed from the DNA before it reached the 3'-end of the *LYS2* gene.

In order to probe if the reduction in the elongation rate reported was a direct consequence of the presence of R-loops, the ChIP experiment was repeated at 2 min after transcription shutdown, but overexpressing *RNH1* from the *GAL1* promoter, to impede the accumulation of R-loops in *Sμ350*. In these conditions, we observed that the accumulation of RNAPII detected in the wild type and *hpr1* strains with the *GAL1p:LYS2::Sμ350* system decreased in the presence of RNase H, although it is only statistically significant in the *hpr1* mutant (Figure R22).

From these experiments we concluded that RNA polymerase II elongation rate was affected specifically by the *Sμ350* region in a DNA:RNA-hybrid dependent manner.

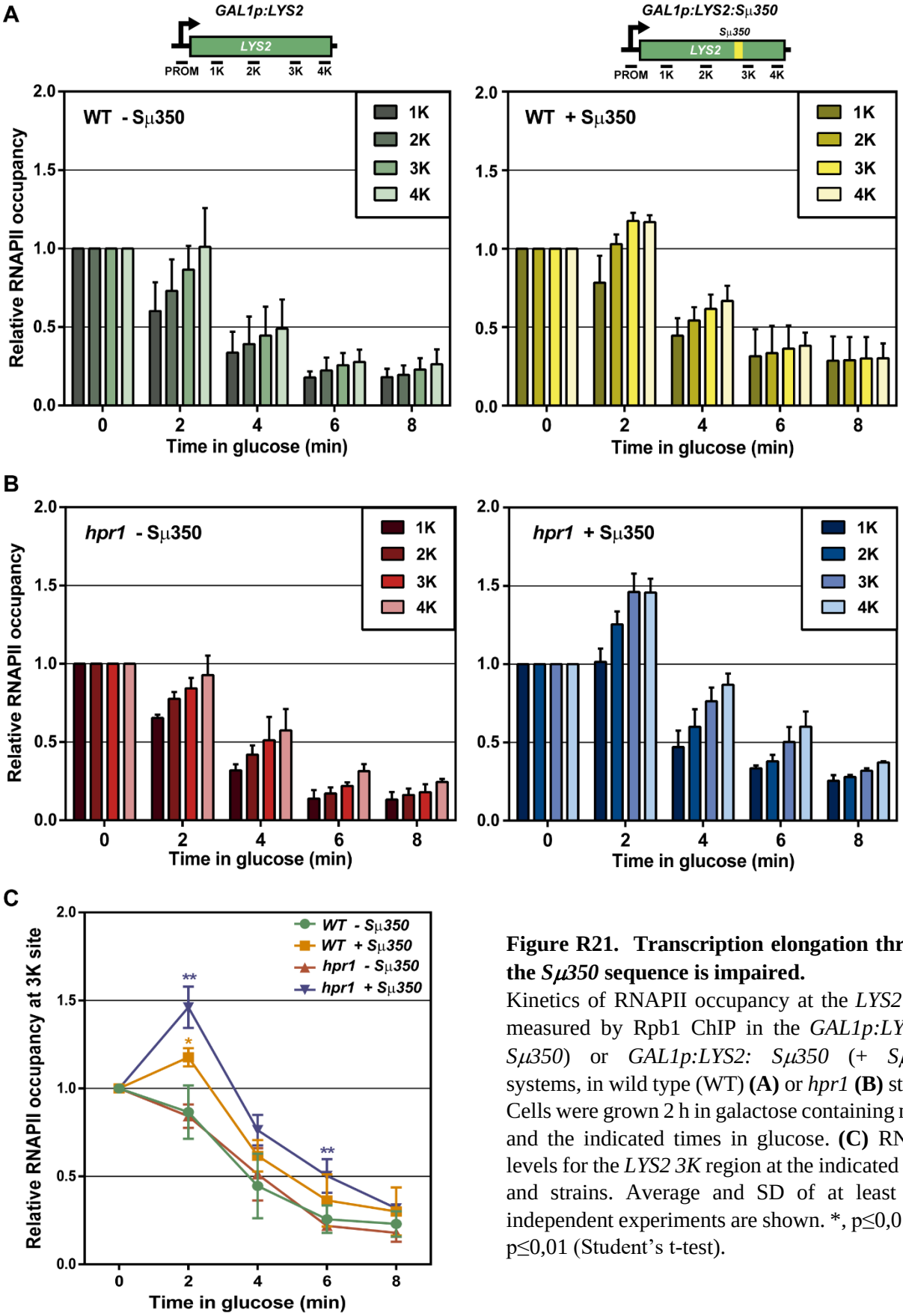


Figure R21. Transcription elongation through the *Sμ350* sequence is impaired.
Kinetics of RNAPII occupancy at the *LYS2* gene measured by Rpb1 ChIP in the *GAL1p:LYS2* (-*Sμ350*) or *GAL1p:LYS2: Sμ350* (+ *Sμ350*) systems, in wild type (WT) **(A)** or *hpr1* **(B)** strains. Cells were grown 2 h in galactose containing media and the indicated times in glucose. **(C)** RNAPII levels for the *LYS2 3K* region at the indicated times and strains. Average and SD of at least three independent experiments are shown. *, p≤0,05; **, p≤0,01 (Student's t-test).

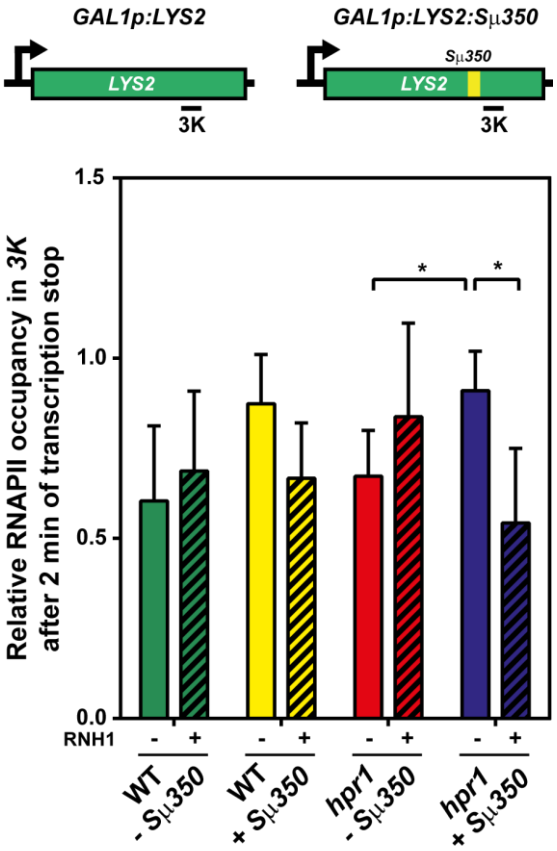


Figure R22. RNAPII accumulation in *Sμ350* is supressed with *RNH1* overexpression. RNAPII occupancy in the *LYS2 3K* region measured by Rpb1 ChIP in the *GAL1p:LYS2* (- *Sμ350*) or *GAL1p:LYS2: Sμ350* (+ *Sμ350*) systems, in wild type (WT) or *hpr1* strains transformed with pGALRH1 (+ *RNH1*) or pRS416 (- *RNH1*) after 2 minutes of transcription shutdown by addition of 2% glucose to the medium. Average and SD of at least three independent experiments are shown. *, p≤0,05 (Student's t-test).

3.2.6. *LYS2* mRNA level decreases on *GAL1p:LYS2:Sμ350*

The flat RNAPII profile observed at the later time-points of the kinetic suggests that stalled polymerases do not reach the 3' end of the gene. They could be being removed from the DNA, either by a premature termination process or by degradation. In both cases, the RNA generated would be truncated and consequently full-length transcripts would be reduced. To check whether this is indeed the case, we analysed the impact *Sμ350* insertion in the *LYS2* mRNA by northern blot assay. For that, we induced transcription of the *GAL1p:LYS2* and *GAL1p:LYS2:Sμ350* systems for 1 h in wild type and *hpr1* backgrounds. We found that, in both strains, the total amount of *LYS2* mRNA was significantly reduced when the *Sμ350* sequence was present (Figure R23), supporting the idea of a RNAPII removal or degradation as a consequence of the R-loop accumulation. Additionally, the mRNAs level in the *hpr1* strain were lower than in the wild type evening the absence of *Sμ350*, in agreement with previous data (Chávez *et al.*, 2001) and the previously mention idea of an R-loop independent elongation defect.

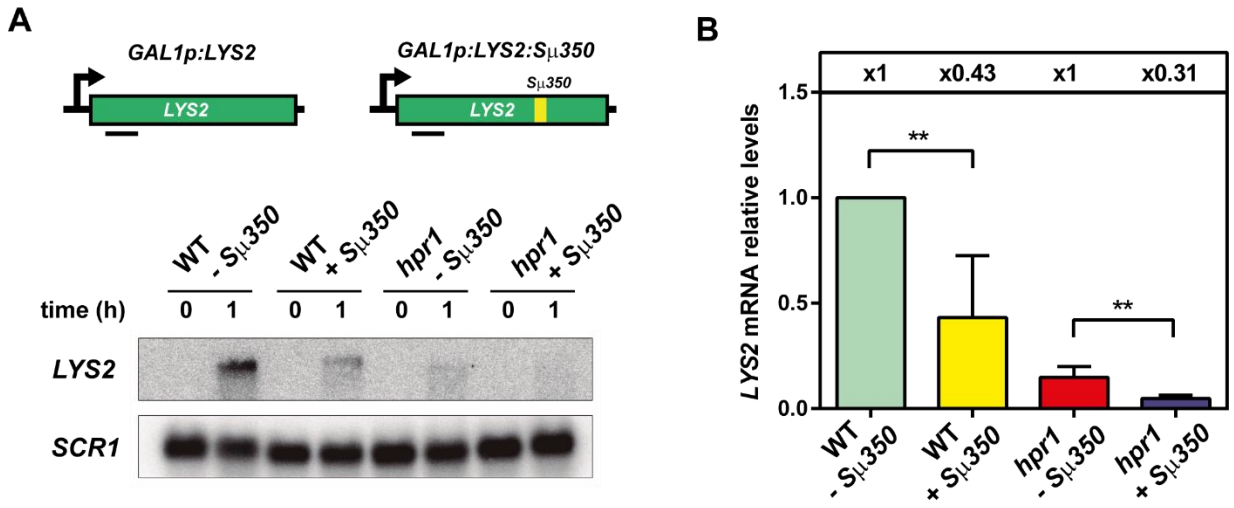


Figure R23. Transcription of *Sμ350* sequence reduced *LYS2* mRNA levels.
mRNA levels of *LYS2* measured by northern blot assay after 1 hour of transcription induction in *GAL1p:LYS2* (- *Sμ350*) or *GAL1p:LYS2:Sμ350* (+ *Sμ350*) systems, in wild type (WT) or *hpr1* mutant strains. **(A)** Representative image of one of the northern blots performed. *SCR1* mRNA was measured as loading control. The position the *LYS2* probe is represented. **(B)** Quantification of the *LYS2* transcript with (+ *Sμ350*) or without (- *Sμ350*) the *Sμ350* sequence in wild type (WT) or *hpr1* mutant strains. Average and SD of five independent experiments are shown. **, p≤0,01 (Student's t-test).

3.2.7. Stalled RNAPII in *GAL1p:LYS2:Sμ350* is not removed by Nrd1-dependent termination

To gain further insight in the mechanism disassembling stalled RNAPIIs in the *GAL1p:LYS2:Sμ350* system we assessed whether they could be prematurely terminating by a Nrd1-dependent pathway. Nrd1 is part of the Nab3-Nrd1-Sen1 complex (NNS) that terminates unspecific intergenic transcription of cryptic unstable transcripts (CUTs) (Arigo *et al.*, 2006). We focused in this termination process because it takes place in the gene body and it dependents on the recognition of a 4 nt signal in contrast to the canonical poly(A)–dependent pathway. Thus, we decided to check if Nrd1 was recruited specifically to the *GAL1p:LYS2:Sμ350* system. To do so, Nrd1 was tagged with 3xHA epitopes in the wild type strain with either *GAL1p:LYS2* or *GAL1p:LYS2:Sμ350* system integrated, to perform Nrd1 ChIP in the same conditions in which we detected the RNAPII accumulation. We checked the region immediately downstream of *Sμ350* (*3K*), a region upstream (*1K*) and as a positive control, a gene whose termination was described to be Nrd1-dependent: *SNR47* (Steinmetz *et al.*, 2001). The results did not show any difference in the recruitment of Nrd1 between the system with or without *Sμ350* (Figure R24). Moreover, the levels of Nrd1 were significantly lower in any of the *LYS2* regions

analysed compared to the *SNR47* control, Therefore, it seems that Nrd1 is not being specifically recruited to the *Sμ350* containing system. However, to rule out Nrd1 role in RNAPII removal it would be interesting to check the RNAPII profile in a *nrd1* mutant.

3.2.8. RNAPII stalled in the *GAL1p:LYS2:Sμ350* system does not change its CTD phosphorylation state

The carboxy-terminal domain (CTD) of the Rpb1 subunit of the RNAPII modifies its phosphorylation pattern thoughout the different stages of the transcription cycle but also when the RNAPII is stalled or blocked (Woudstra *et al.*, 2002; Shanbhag *et al.*, 2010; Pankotai *et al.*, 2012). This prompted us to examine if a particular phosphorylated form of the RNAPII CTD is accumulated at *Sμ350* as a way to determine the state of the transcription. To do that, we performed a ChIP assay at the *GAL1p:LYS2* and *GAL1p:LYS2:Sμ350* systems in a *hpr1* mutant, after 2 min of transcription shutdown in

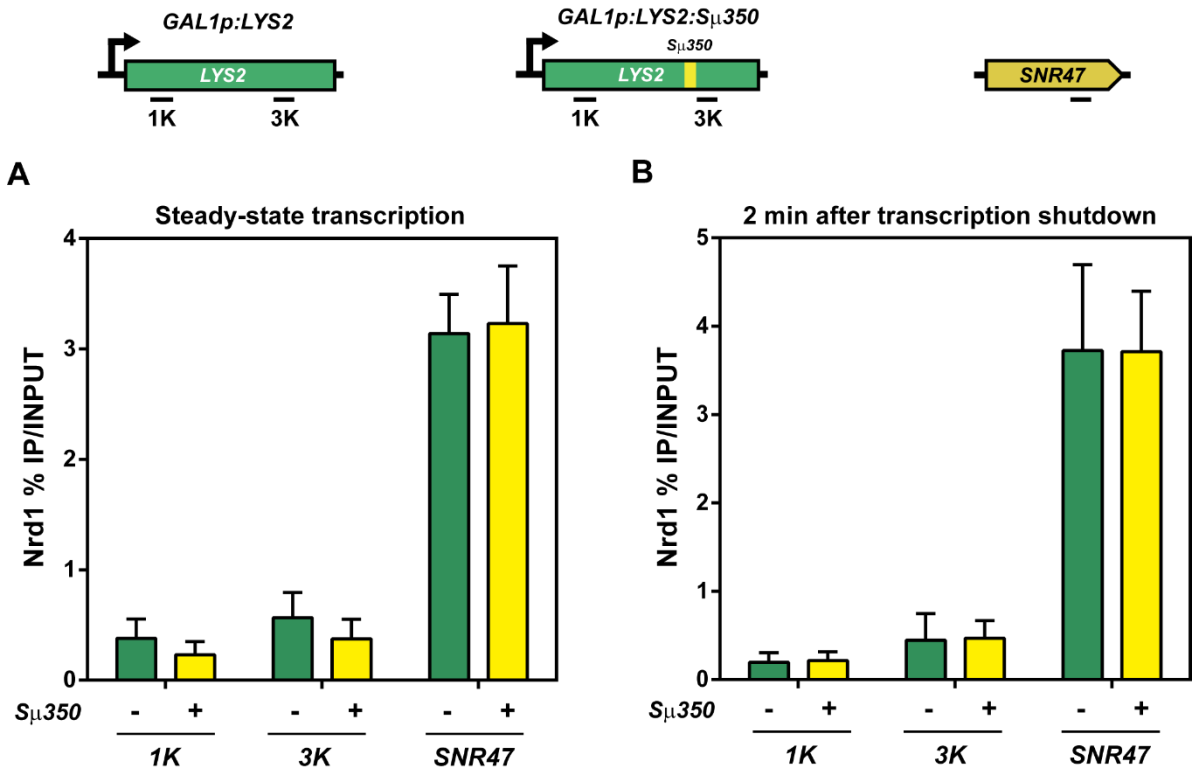


Figure R24. Nrd1 does not localize to the *GAL1p:LYS2:Sμ350* system.
Nrd1-HA localization measured by ChIP in the *GAL1p:LYS2* (- *Sμ350*) or *GAL1p:LYS2:Sμ350* (+ *Sμ350*) systems in a wild-type strain after 2 hours of transcription (Steady-state transcription) or 2 min of transcription shutdown, measured in the *LYS2* 1K and 3K regions. SNR47 was included as positive control. Average and SD of three independent experiments are shown.

G1 arrested cells. We used specific antibodies for the different phosphorylated forms of Rpb1 CTD: Ser5-P (3E8) (Figure R25-A); Ser2-P (3E10) (Figure R25-B); Tyr1-P (3D12) (Figure R25-C) and Thr4-P (6D7) (Figure R25-D). Each of these phosphorylations marks a different state of the RNA polymerase II: Initiation (Ser5-P), elongation (Ser2-P and Thr4-P) or termination (Tyr1-P) (Heidermann *et al.*, 2013). The results obtained indicate that none of these phosphorylated forms of RNAPII CTD were significantly different in

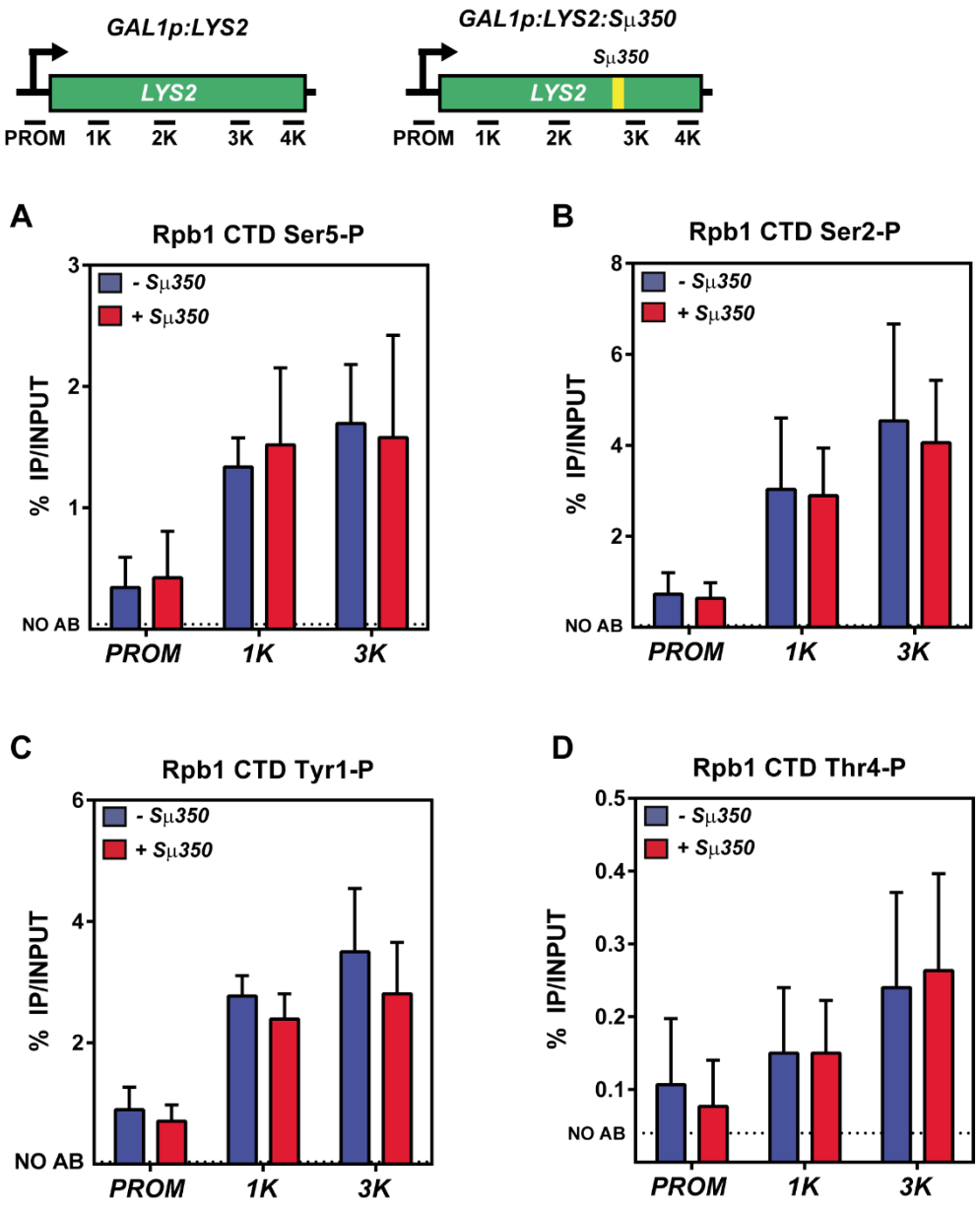


Figure R25. CTD phosphorylation does not change in the *GAL1p:LYS2:Sμ350* system.
RNAPII CTD phosphorylated forms measured by ChIP in the *GAL1p:LYS2* (- *Sμ350*) or *GAL1p:LYS2:Sμ350* (+ *Sμ350*) systems in an *hpr1* strain after 2 min of transcription shutdown. Ser5-P **(A)**, Ser2-P **(B)**, Tyr1-P **(C)** and Thr4-P **(D)** levels were measured in the *GAL1* promoter (*PROM*), *LYS2 1K* and *3K* regions. Average and SD of at least three independent experiments are shown.

the *GAL1p:LYS2:Sμ350* respect to the *GAL1p:LYS2* system. Therefore, we concluded that the phosphorylation state of RNAPII CTD did not change in the RNAPII stalled by R-loops in the studied conditions, suggesting that the RNAPII block maybe needs to extend in time in order to induce changes in the CTD phosphorylation.

In brief , we showed that co-transcriptionally formed R-loops are able to transiently stall RNAPII, reducing the elongation rate. Stalled RNAPII is probably quickly removed, as we do not detect a built up of RNAPII at steady-state transcription conditions and part of the RNAPII do not reach the 3'-end of the gene. We proposed that a premature termination or degradation of RNAPII could explain these results. Although the NNS complex is not specifically recruited to *Sμ350,* further studies to determine its implication are needed. Moreover, the role of alternative termination pathways or poly-ubiquitylation-dependent removal of RNAPII needs to be pursued.

3.2.9. Generation of transcriptional system with an inducible single-stranded break: *GAL1p:LYS2:FRT*

In order to induce a single-stranded break (SSB) at a specific locus, we used the Flp-nick system, in which a mutated form of the flipase recombinase (FlpH305L) binds to a specific flipase recognition target site (FRT), mediating the cut of one of the DNA strands and remains covalently bound to the DNA during the process (Nielsen *et al.*, 2009). The SSB with a covalently bound protein produced by the Flp-nick system mimics the damage produced by a poisoned topoisomerase I. To generate an inducible SSB at the *LYS2* gene, we amplified by PCR two different cassettes containing the FRT sequence from the pTINVFRT-1 plasmid, flanked by 50 bp sequences homologue to the *LYS2* region in which we want to insert it. The FRT sequence was amplified in a direct or an inverted orientation to insert it in the template or in the non-template strand of the *LYS2* gene. These two cassettes were independently introduced in the *GAL1p:LYS2* system by Cas9-induced homologous recombination by co-transforming one of the cassettes and the pML104 plasmid, as we previously did for cloning *Sμ350* into *GAL1p:LYS2*. To direct the integration to the position 2942 of the *LYS2* ORF we employed the same gRNA. We obtained two systems: one with the FRT in the template strand, *GAL1p:LYS2:FRTt*; and another with the FRT in the non-template strand, *GAL1p:LYS2:FRTnt* (Figure R26-A). Positive transformants were selected, checked by PCR and sequenced to rule out the

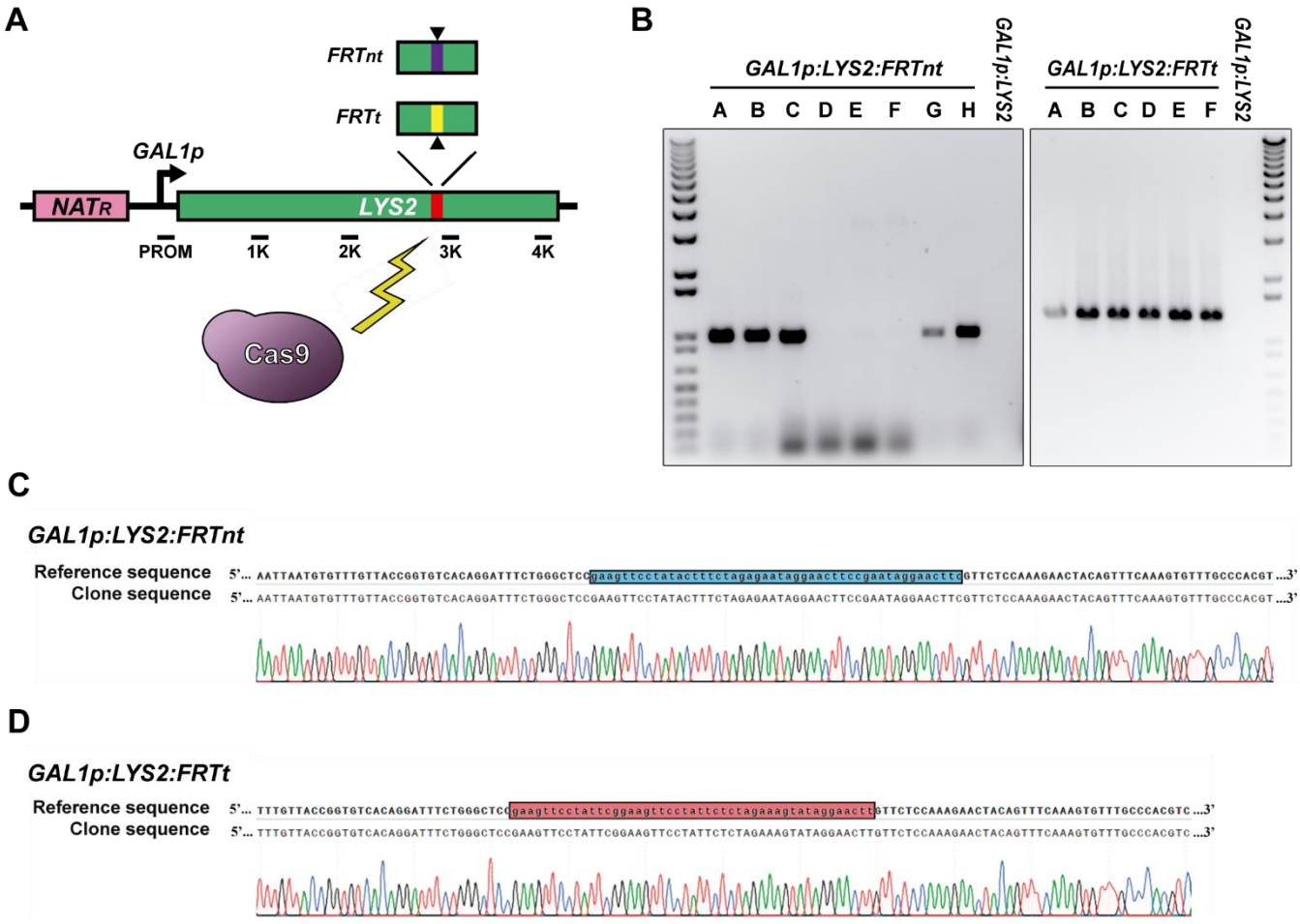


Figure R26. Generation of the *GAL1p:LYS2:FRT* systems.
(A) *GAL1p:LYS2* system in which a DNA fragment containing the FRT sequence in both orientations was introduced by induction of Cas9-directed homologous recombination. **(B)** PCR from genomic DNA from different *GAL1p:LYS2:FRTnt* clones using 'FRT compr' and 'LYS2 4K rv' primers; or *GAL1p:LYS2:FRTt* clones using 'FRT compr' and 'LYS2 3K fw' primers to verify the integration of the FRT cassette. GLY-2D (*GAL1p:LYS2*) was included as control. **(C)** Aligned map of the FRTnt sequence inserted in the *LYS2* gene, obtained by sequencing (bottom) of the GLFc-2B clone (*GAL1p:LYS2:FRTnt-)* and the reference sequence (top). **(D)** Aligned map of the FRTt sequence inserted in the *LYS2* gene, obtained by sequencing (bottom) of the GLFt-3C clone (*GAL1p:LYS2:FRTt-)* and the reference sequence (top).

presence of additional mutations (Figure R26-B). The flipase recombinase (FlpH305L) was overexpressed from a galactose-inducible plasmid (pBIS-GALkFLP) or a doxycycline repressible vector (pCM190-FLP).

3.2.10. Quantification of the flipase recombinase-induced SSB on the *GAL1p:LYS2:FRT* systems

To test if FlpH305L was able to produce a nick in the target sequences inserted in the *LYS2* gene, we performed a southern blot assay in two different conditions: native, that

sensed only DSB; and alkaline, that detected both, SSB and DSB. We used the *GAL1p:LYS2*, *GAL1p:LYS2:FRTt* and *GAL1p:LYS2:FRTnt* strains. *FLPH305L* and *LYS2* were expressed for 3 h in galactose on G1 arrested cells (to avoid SSB conversion into DSB during replication) before genomic DNA isolation. Genomic DNA was digested with *Cla*I and *Pvu*I, generating a specific DNA fragment of 3.12 kb that could be seen as a single band in the gel if FlpH305L did not cut, or as two bands, of 1.98 and 1.15 kb respectively if the DNA was cut by FlpH305L. Our data showed that FlpH305L cut the DNA in both, *GAL1p:LYS2:FRTnt* and *GAL1p:LYS2:FRTt*, while the *GAL1p:LYS2* system showed no damage. A great proportion of the DNA damage at the *GAL1p:LYS2:FRTnt* and *GAL1p:LYS2:FRTt* systems were SSB, that could be perceived as the difference between the alkaline respect of the native gel. The percentage of molecules with SSBs was 3.24% for the *GAL1p:LYS2:FRTt* system and 2.01% for the *GAL1p:LYS2:FRTnt*, with no detectable level of SSB in the *GAL1p:LYS2* system (Figure R27).

We conclude that the Flp-nick system induced SSBs during *LYS2* transcription in both, the *GAL1p:LYS2:FRTt* and *GAL1p:LYS2:FRTnt* strains, as we were able to measure the product of the SSB, although the percentage of the population cut was low.

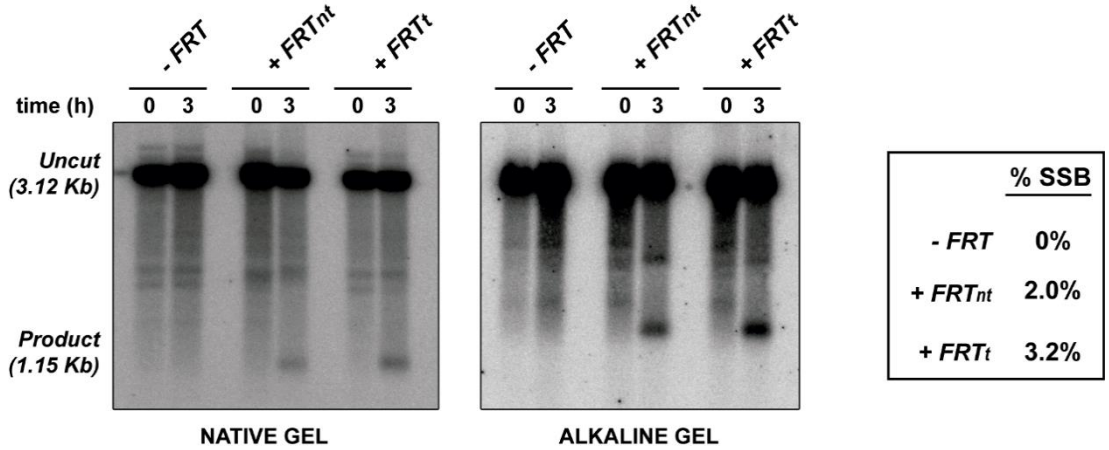


Figure R27. FlpH305L induced SSBs in the *GAL1p:LYS2:FRTt* and *GAL1p:LYS2:FRTnt* systems. Southern blot assay in native or alkaline conditions in the *GAL1p:LYS2*, *GAL1p:LYS2:FRTnt* and *GAL1p:LYS2:FRTt* strains overexpressing FlpH305L from the pBIS-GALkFLP plasmid. Genomic DNA was digested with *Cla*I and *Pvu*I and a *LYS2* specific probe was used.

3.2.11. RNAPII accumulates in the *GAL1p:LYS2:FRTt* system preferentially upstream of the SSB site

Next, to study the fate of an RNAPII that encounters an SSB, we measured the RNAPII profile in the *LYS2* gene by ChIP using the 8WG16 antibody. The experiment was carried out using the *GAL1p:LYS2, GAL1p:LYS2:FRTt* and *GAL1p:LYS2:FRTnt* strains transformed either with *FLPH305L* overexpressing plasmid or the empty vector. Cells were arrested in G1 to avoid DSB production by replication. Transcription of *LYS2* and *FLPH305L* were simultaneously induced for 3 h before sample collection. When *FLPH305L* was overexpressed, the RNAPII level increased in the *GAL1p:LYS2:FRTt* system compared to *GAL1p:LYS2*. This RNAPII accumulation was observed throughout the *LYS2* ORF, being significant in the region upstream of the FRT site. The *GAL1p:LYS2:FRTnt* system, however, presented the same profile than *GAL1p:LYS2*, without RNAPII accumulation (Figure R28-A). This increase in RNAPII was specific of the *LYS2* ORF in the *GAL1p:LYS2:FRTt* system as it was not observed at the promoter or in other genes like *GCN4* (Figure R28-B). In the absence of *FLPH305L* expression, all the strains showed the same RNAPII profile (Figure R28-C). Thus, RNAPII accumulation is caused by the flipase recombinase nicking template DNA.

This data revealed a defect in transcription induced by SSBs that may stall RNA polymerase II. We could expect that, if the SSB is blocking the progression of the RNAPII through the DNA, it will accumulate exclusively upstream of the damage. However, our observations indicated that although it accumulates preferentially upstream it also increases downstream the nick. The presence RNAPII downstream could be caused by either nicks also affecting elongation downstream or nicks initiating pervasive transcription as has been recently shown for DSBs (Victor *et al.*, 2019). To differentiate both, we decided to uncouple the expression of the *LYS2* gene and the *FLPH305L* to induce the nick before *LYS2* expression. To perform the assay, we overexpressed *FLPH305L* from the *Tet* promoter during 3 h in G1 synchronized cells before *LYS2* induction with galactose. As we wanted to know the effect of the SSB in the first rounds of transcription to minimize RNAPII occupancy at the 3' end region of *LYS2* before the recombinase induces the nick, we measured RNAPII levels by ChIP 10 minutes after *LYS2* induction in a region upstream (*2K*) and downstream (*4K*) of the FRT site. We found in the *GAL1p:LYS2:FRTt* strain a significant increase of RNAPII in the region upstream the SSB (2K), but not in the region downstream (4K), while the *GAL1p:LYS2:FRTnt*

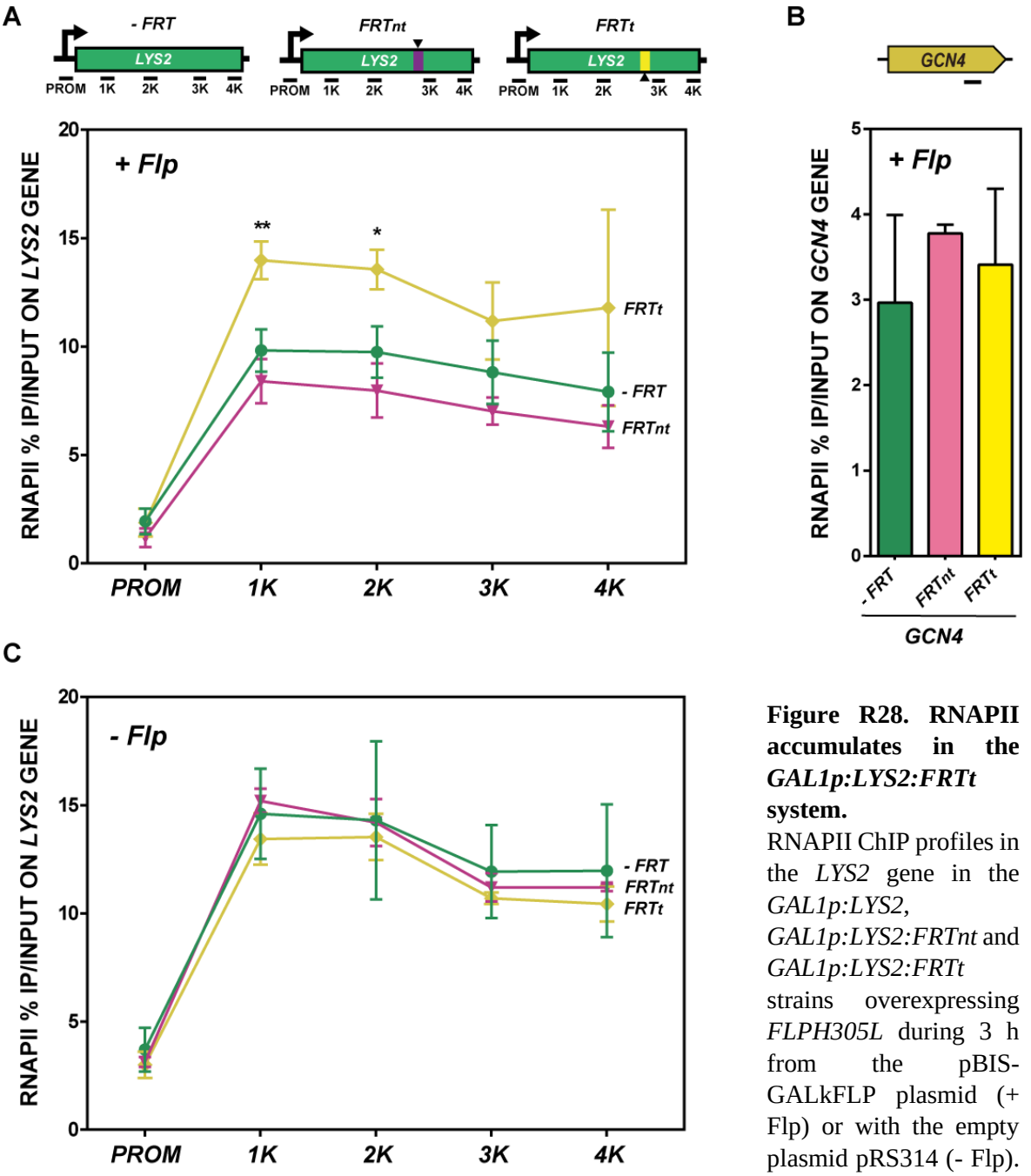


Figure R28. RNAPII accumulates in the *GAL1p:LYS2:FRTt* system.
RNAPII ChIP profiles in the *LYS2* gene in the *GAL1p:LYS2*, *GAL1p:LYS2:FRTnt* and *GAL1p:LYS2:FRTt* strains overexpressing *FLPH305L* during 3 h from the pBIS-GALkFLP plasmid (+ Flp) or with the empty plasmid pRS314 (- Flp). GCN4 gene was analysed as a control. Average and SD of at least three independent experiments are shown. *, p≤0,05; **, p≤0,01 (Student's t-test).

strain presented a similar RNAPII profile to the *GAL1p:LYS2* system (Figure R29). This experiment confirmed the result obtained in the steady-state RNAPII ChIP and suggests that a nick in the template DNA affects transcription of the approaching RNAPII but it may also affect the polymerases placed downstream of the damage.

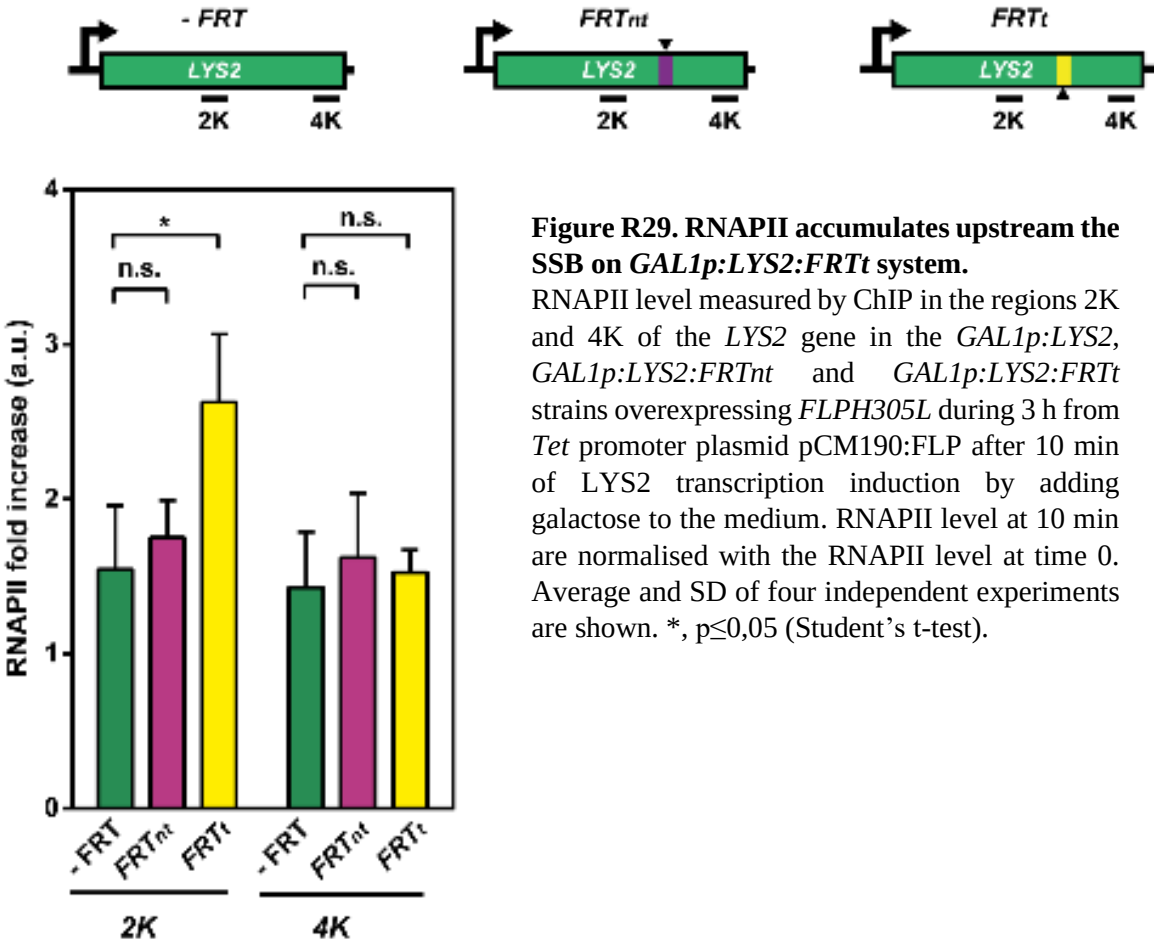


Figure R29. RNAPII accumulates upstream the SSB on *GAL1p:LYS2:FRTt* system.
RNAPII level measured by ChIP in the regions 2K and 4K of the *LYS2* gene in the *GAL1p:LYS2*, *GAL1p:LYS2:FRTnt* and *GAL1p:LYS2:FRTt* strains overexpressing *FLPH305L* during 3 h from *Tet* promoter plasmid pCM190:FLP after 10 min of LYS2 transcription induction by adding galactose to the medium. RNAPII level at 10 min are normalised with the RNAPII level at time 0. Average and SD of four independent experiments are shown. *, p≤0,05 (Student's t-test).

3.2.12. *LYS2* mRNA level decreases specifically in the *GAL1p:LYS2:FRTt* strain

We reasoned that there are two possible fates for a RNAPII stalled by a SSB in the template strand: either transcription could be quickly resumed, in which case the mRNA level should not be affected in steady-state conditions, or if the blockage is more persistent or the RNAPII is being removed, *LYS2* transcripts will decrease. In order to test these possibilities, we measured *LYS2* mRNA in *GAL1p:LYS2:FRTt* and *GAL1p:LYS2:FRTnt* strains transformed with a plasmid overexpressing *FLPH305L* from the *tet* promoter, or an empty vector. We overexpressed *FLPH305L* during 3 h and *LYS2* gene during the last 1 hour before collecting the samples. *LYS2* mRNA level was measured by northern blot assay. We detected a significant 0.65-fold reduction in the *LYS2* mRNA of the *GAL1p:LYS2:FRTt* system with the induction of *FLPH305L* specifically. In contrast, the amount of *LYS2* mRNA in the *GAL1p:LYS2:FRTnt* system did not change with the expression of *FLPH305L* (Figure R30), thus indicating that a SSB generated specifically in the template strand of the DNA reduces *LYS2* transcription probably by stalling RNAPII. This result is in agreement with previous *in vitro* studies that showed a reduction

in the amount of transcript when SSBs localized in the template strand of the DNA (Zhou & Doetsch 1993; Kathe *et al.*, 2004; Neil *et al.*, 2012).

To summarize, using the *GAL1p:LYS2:FRTt* and *GAL1p:LYS2:FRTnt* systems we were able to determine that a SSB produced in the template strand of the DNA, but not in the non-template, is able to stall RNAPII upstream of the damage site. This RNAPII stalling resulted in a *LYS2* mRNA reduction, suggesting that the blocked RNAPII did not reach the 3'-end of the gene probably because it is being removed. Further analysis is required to determine the mechanisms involved in this process.

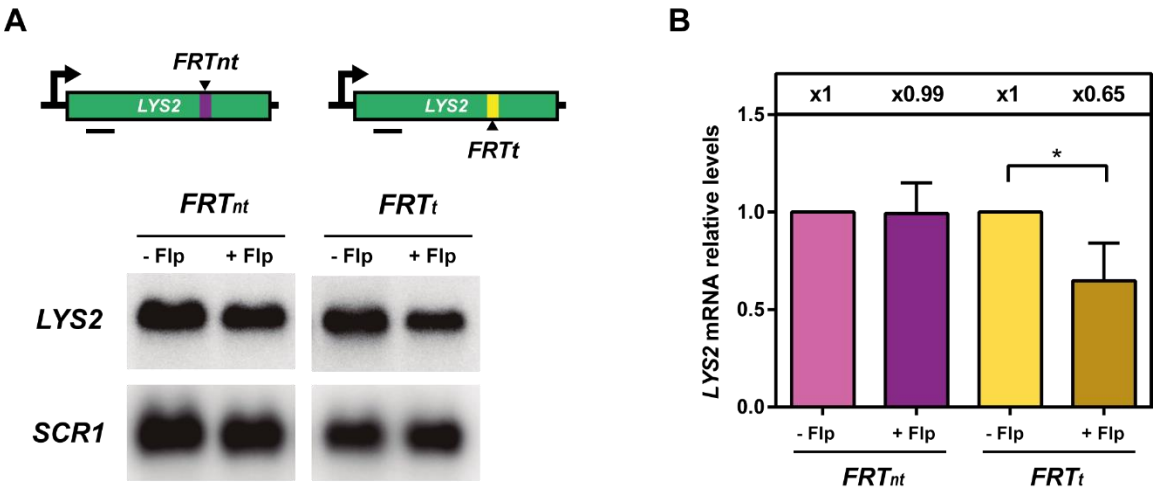


Figure R30. Induction of a SSB in the *GAL1p:LYS2:FRTt* system reduces *LYS2* mRNA level. mRNA level of *LYS2* measured by northern blot assay after 3 h of *FLPH305L* overexpression and 1 h of *LYS2* transcription induction in the *GAL1p:LYS2:FRTnt* (FRTnt) and *GAL1p:LYS2:FRTt* (FRTt) strains transformed with either, pCM:FLP (+ Flp) or pRS314 (- Flp) plasmids. **(A)** Representative image of one of the northern blots performed. *SCR1* mRNA was measured as loading control. The position the *LYS2* probe is represented. **(C)** Quantification of the *LYS2* transcript with (+ *Flp*) or without (- *Flp*) FLPH305L expression in *GAL1p:LYS2:FRTnt* (FRTnt) or *GAL1p:LYS2:FRTt* (FRTt) strains. Average and SD of three independent experiments are shown. *, p≤0,05 (Student's t-test).

3.2.13. Generation of *GAL1p:LYS2* diploid recombination systems

Using the different systems based on the *GAL1p:LYS2* construct that we have described we were able to measure the outcomes of transcription confronting R-loops and SSBs, as described. Next, we adapted these systems to analyse whether DNA damage caused by R-loops or SSBs could be repaired by recombination and to determine the role of transcription in this process. The new adapted systems would allow us to identify the factors involved in resolving the RNAPII accumulation and elongation defects induced by R-loops and SSBs by measuring the effect on recombination in different transcription

and DNA repair mutants. To the date, most of the genetic recombination systems employed in yeast measured single-strand annealing (SSA), a specific homology-dependent pathway of double-strand break (DSB) repair that works between two direct-repeat sequences. Here, we decided to focus our study in the other major HR pathways: synthesis-dependent strand annealing (SDSA), recombination between homologs and break-induced replication (BIR), that relies on a strand invasion in the homolog chromosome (as in diploid cells). For this, we generated diploid strains with either the *GAL1p:LYS2:Sμ350*, *GAL1p:LYS2:FRTt* or *GAL1p:LYS2:FRTnt* fusion inserted in one *LYS2* locus, and a point mutation in the other one. In these strains we could control transcription of the *LYS2* gene with the *GAL1* promoter and induce formation of DNA:RNA hybrids or SSBs. All three constructs introduce premature stop-codons either in the *Sμ350* or the FRT sequence that would make their Lys2 protein non-functional. Thereby, the only manner to restore a functional *LYS2* gene without any premature stop-codon would be by an HR event between both mutated alleles (Figure R31).

To generate these new strains we used the Cas9 system. First, we introduced a single-base deletion in the position 3705 of the *LYS2* ORF in a wild-type haploid strain.

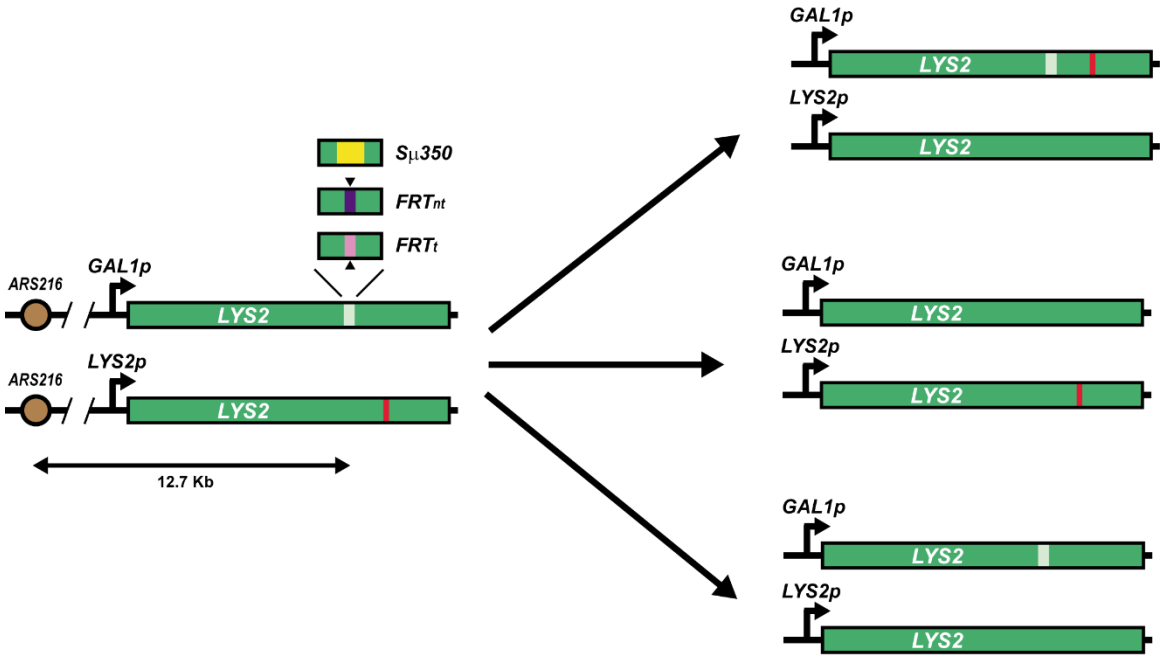


Figure R31. Generation of the *GAL1p:LYS2* diploid recombination systems.
Scheme of the *GAL1p:LYS2* diploid strains with either, the *GAL1p:LYS2:Sμ350* (*Sμ350*), *GAL1p:LYS2:FRTnt* (FRTnt) or *GAL1p:LYS2:FRTt* (FRTt) systems in one chromosomal *LYS2* locus and a *lys2* with a point mutation in the other. The possible outcomes of HR that would generate a lysine prototrophic yeast are represented.

Cas9 endonuclease was expressed from the pML104 plasmid that contained a gRNA that directed Cas9 to the indicated position of the *LYS2* gene. The single-nucleotide deletion was produced by error prone repair of the DSB induced by Cas9. Positive colonies were isolated, replicated in medium without lysine in order to identify *lys2* mutants, and finally sequenced to confirm the presence of the mutation. The selected strain (YLYS2-6) contains a single T deletion that generated an early stop codon at the desired position. Next, we crossed the strains carrying *GAL1p:LYS2:Sμ350, GAL1p:LYS2:FRTt* or *GAL1p:LYS2:FRTnt* systems with YLYS2-6, selecting the diploids in medium with nourseothricin and α-factor.

The generated diploid strains allow us to measure non-SSA recombination with the homolog chromosome in different conditions of transcription or DNA damage. Consequently, using these strains, we could measure the effect of R-loops and SSBs in different mutants to figure out the mechanism to resolve the elongation defects reported.

3.2.13.1. R-loops increase recombination between homologous chromosomes

Previous studies reported that mutants that accumulate R-loops, such as *hpr1*, increase SSA. This increase in recombination was produced as a consequence of transcription-replication conflicts associated with transcription elongation deficiencies (Prado *et al.*, 1997; Gómez-González & Aguilera 2009). In addition, we suggested in this work that R-loops could stall RNAPII although very transiently. To determine if the accumulation of R-loops generates DNA damage that could be repaired by SSA-independent recombination, we used the diploid system containing *GAL1p:LYS2:Sμ350*. We measured the recombination level with or without transcription of *LYS2* gene by adding galactose or glucose to the medium. We performed the experiments with and without *RNH1* overexpression from *Tet* promoter to discriminate between transcription and R-loop accumulation as causal effects of any difference observed. With these conditions, we detected a low although significant increase in recombination frequencies (from $4.8x10^{-6}$ to $16.2x10^{-6}$) with respect to the non-transcribed control. However, this difference was lost when *RNH1* was overexpressed, indicating, therefore, a clear impact of the presence of R-loops on recombination. *RNH1* overexpression produced a slight, non-significant, increase in recombination frequency in the absence of transcription at *LYS2* ($7.2x10^{-6}$), probably due to the stress induced by the overexpression of this

ribonuclease (Figure R32). Using this genetic assay, we have verified therefore that R-loops are at least partially responsible of an increase in DNA damage that can be repaired by HR.

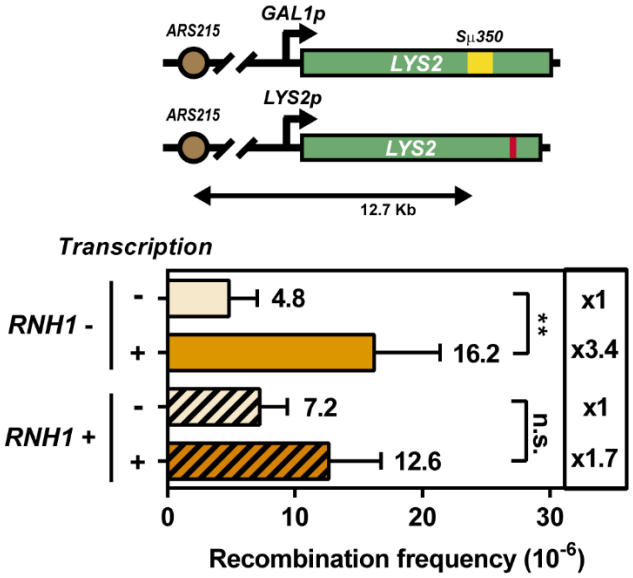


Figure R32. R-loop formation induces recombination on *GAL1p:LYS2:Sμ350* system.
Recombination analysis in the *GAL1p:LYS2:Sμ350* diploid system transformed with pCM189-RNH1 (*RNH1* +) or pCM189 empty vector (*RNH1* -), performed in plates with 2% glucose (transcription -) or 2% galactose (transcription +). Average and SD of four independent experiments are shown. **, p≤0,01 (Student's t-test).

3.2.13.2. FRT-induced damage increases homologous recombination independently of the strand

We showed that a nick in the template DNA strand leads to transcriptional conflicts that resulted in RNAPII stalling. Now, using the *GAL1p:LYS2:FRTt* and *GAL1p:LYS2:FRTnt* genetic systems we could assess whether transcription could affect the repair of the damage induced by overexpression of the mutated FLP nuclease FLPH3045L. For this, we performed recombination assays after inducing FLPH3045L for 5.5 h from the *Tet* promoter with or without transcription of *LYS2* gene culturing yeast cells in galactose or glucose liquid medium, respectively. Induction of *FLPH3045L* increased the recombination frequency in 3 orders of magnitude in both systems, *GAL1p:LYS2:FRTt* and *GAL1p:LYS2:FRTnt*, with no significant differences between them (Figure R33). This suggests that recombination is probably triggered by SSBs converting into DSBs by replication and not by the stalling of RNAPII observed specifically in the *GAL1p:LYS2:FRTt*. Interestingly, when *LYS2* was transcribed, the recombination frequency diminished one order of magnitude in both systems (Figure R33). Further analysis would be required to determine the cause of this decrease in recombination, but we one possibility could be that the *FLPH3045L* ability to recognize its target sequence

could be hindered by transcription. Another possibility would be that transcription could impede the repair of the damage by HR. Furthermore, the recombination frequency induced by *FLPH3045L* expression was much higher than that produced by R-loops, consistent with the idea that transient R-loops themselves are a poor source of DNA damage that probably requires either their stabilization by chromatin modification or by DNA:RNA hybrid-binding protein, or an associated impairment of transcription that could facilitate transcription-replication collisions to increase damage.

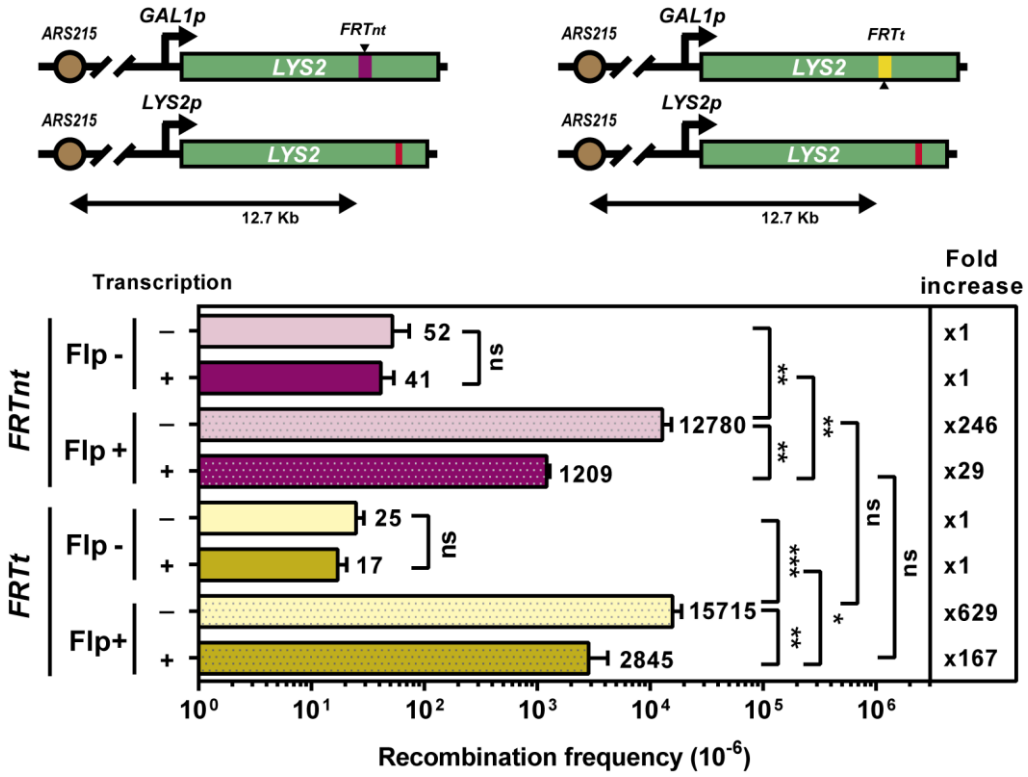


Figure R33. A SSB induces hyperrecombination in both *GAL1p:LYS2:FRT* systems.
Recombination analysis in the *GAL1p:LYS2:FRTnt* (FRTnt) and *GAL1p:LYS2:FRTt* (FRTt) diploid systems transformed with pCM190:FLP (Flp +) or pCM190 empty vector (Flp -), performed in liquid medium with 2% glucose (transcription -) or 2% galactose (transcription +). *FLPH104L* was overexpressed for 5.5 h by removing doxycycline from the media. Average and SD of three independent experiments are shown. *, p≤0,05; **, p≤0,01; ***, p≤0,005 (Student's t-test).

4. DISCUSSION

4.1. Excess of Dis3, Rie1 and She2 RNA-binding proteins induce R-loop dependent genome instability

Yra1 is a component of the mRNP that, when it is in excess, not only binds to RNA but also to DNA:RNA hybrids accumulating at naturally R-loop-forming regions. This interaction increases the amount of R-loop structures in the cell, probably by preventing their resolution (García-Rubio *et al.*, 2018; Rondón & Aguilera 2019). Consistently with the effect in the genome originated by R-loops, Yra1 overexpression reduces the global level of replication and causes R-loop dependent genome instability (Gavaldá *et al.*, 2016). Considering those consequences, it is not surprising that Yra1 expression is tightly regulated by the cell. We have conducted a screening to assess if the overabundance of other RNA-binding protein could also lead to R-loop dependent genome instability. We identified four RBP-coding genes whose overexpression showed an increase in DNA damage: *DIS3*, *HRP1*, *SHE2* and *RIE1* (Figures R3 and R4), yet only in *DIS3*, *SHE2* and *RIE1* it is R-loop dependent. Considering this result in the view of previous studies in mRNP mutants that also produce genome instability (Domínguez-Sánchez *et al.*, 2011; González-Aguilera *et al.*, 2011; Santos-Pereira *et al.*, 2013), our findings suggest that both, the lack and overabundance of certain RNA-binding proteins could potentially destabilize mRNP biogenesis, accumulating R-loops and threatening genome integrity.

HRP1 overexpression, on the contrary, does not induce R-loop dependent DNA damage (Figure R7-A). Although it is not clear whether *hrp1* mutation enhances R-loops, it has been clearly shown that 3'-end processing of the nascent RNA prevents the formation of these structures (Stirling *et al.*, 2012). Thus, it is possible that the excess of Hrp1 does not cause the same effect than its absence or that it affects other processes that prevent DNA damage. In this line, *hrp1-5* mutant, but also mutation in other cleavage factors as *rna14-1* or *rna15-1*, are defective in TCR, probably because RNAPII is not removed at an adequate rate, indicating a dysfunction of DDR (Gaillard & Aguilera 2014) and suggesting that Hrp1 deregulation may impair DNA repair and thus it accumulates DNA damage independently of R-loops, as we observe here.

Surprisingly, the DNA damage produced by *DIS3*, *RIE1* or *SHE2* overexpression was not associated with an increase in recombination (Figure R8), that is an usual outcome from DSBs, and in this sense the overexpression of *DIS3*, *RIE1* or *SHE2* differs from what is described for *YRA1*, that shows a strong R-loop-mediated

hyperrecombination (Gavaldá *et al.*, 2016). However, this is not the only case of increased Rad52 foci not linked to higher recombination level. An example of this is the *mlp1Δ* mutant that impairs tethering of the transcription machinery with the nuclear pore. This mutant only increases recombination in presence of the activated-induced cytidine deaminase (AID), that changes cytidine into uridine, but *mlp1* increases DNA damage and R-loop accumulation even in the absence of AID (García-Benítez *et al.*, 2017). This suggests that changes in mRNP biosynthesis factors could generate different phenotypes attending to their role in the whole process.

A possible explanation for the discrepancy between our candidates and Yra1 is that the damage generated by *DIS3*, *RIE1* or *SHE2* overexpression could be preferentially repaired by a pathway different from HR, like NHEJ. An excess of Yra1 also reduces growth in a *rad53* S-phase checkpoint mutant background or in combination with *rad51* and *rad52* recombination mutants (Gavaldá *et al.*, 2016), suggesting that the DNA damage generated by *YRA1* overexpression may be primarily repaired by HR. It would be interesting to know whether this is the case for the other RBPs overexpressed or whether they are lethal in combination with mutants of the NHEJ pathway. Alternatively, DNA damage generated by *DIS3*, *RIE1* or *SHE2* overexpression could be remaining unrepaired leading to plasmid loss or could be not sufficient to alter the level of recombination in the systems assayed. In agreement with this idea, we could not detect synthetic lethality in *rnh1 rnh201* mutant (Figure R7-B) or increased sensibility to genotoxic agents (Figure R6). The fact that cells have developed a tightly regulated mechanism to control *YRA1* expression (Rodríguez-Navarro *et al.*, 2002) reinforces the hypothesis that probably a Yra1 excess has a stronger effect in R-loop metabolism than the overproduction of Dis3, Rie1 or She2.

4.1.1. Role of *DIS3* overexpression in genomic stability

Dis3 is part of the exosome, a 10-subunit complex with nuclease activity that was first identified for its function in 5.8S rRNA 3'-end processing (Mitchell *et al.*, 1996) but also participates in RNA turnover (Chelebowski *et al.*, 2013). Functionally we can differentiate nine structural subunits that form the core complex to which the two functional subunits, Dis3 and Rrp6, interact. Rrp6 is an exclusively nuclear accessory component of the exosome with 3'-5' exonuclease activity, while Dis3 is both nuclear

and cytoplasmic and contains not only exo but also endonuclease activity (Schneider *et al.*, 2009; Dziembowski *et al.*, 2000; Liu *et al.*, 2006). The exosome associates with different co-factors that facilitate or direct its function, being TRAMP the main one (LaCava *et al.*, 2005). TRAMP is formed by the RNA helicase Mtr4, the RNA binding protein Air1/2 and a poly(A) polymerase Trf4/5, that adds a short poly(A) tail to the RNAs to facilitate the degradation of highly structured RNAs by the exosome.

DIS3 overexpression increases DNA damage in a *RNH1*-sensitive manner (Figure R7), in agreement with the slight increase in DNA:RNA hybrids detected by immunofluorescence (Figure R9). Recently, it has been reported also by IF that *dis3-ts* conditional mutant increases R-loops (Millbury *et al.*, 2019). Moreover, the *dis3-ts* mutant also shows an increase in DNA damage that is partially supressed by RNase H, similarly to *DIS3* overexpression (Figure R10-A). This led us to think that maybe both situations were similar. Indeed, we observed that *DIS3* overexpression reproduced the accumulation of rRNA intermediates previously described for the mutant (Figure R10-B) (Millbury *et al.*, 2019; Allmang *et al.*, 2000; Mitchell *et al.*, 1996). Thus, we conclude that the phenotypes that we observed overexpressing *DIS3* are probably caused by a loss of function of the whole exosome probably due to Dis3 aggregation, alone or with other subunits of the complex, in agreement with the role of Dis3 in stabilizing the exosome core complex (Dziembowski *et al.*, 2000).

Dis3 is not the only component of the exosome linked to R-loops metabolism. Cells lacking Rrp6 show increased transcription-associated hyperrecombination and chromosomal instability (Luna *et al.*, 2005; Wahba *et al.*, 2011). These phenotypes are also extensive to TRAMP complex mutants. Mutations in *AIR1* or *TRF4* genes generate chromosomal instability (Wahba *et al.*, 2011) and at least *trf4*Δ causes R-loop dependent hyperrecombination (Gavaldá *et al.*, 2013). One possible mechanistic explanation is that global RNA metabolism deregulation could produce accumulation of aberrant and ncRNAs that would hybridize with the DNA contributing to R-loop formation and leading to genome instability. Indeed, depletion of Dis3, but also mutations in its exonuclease activity, presented an extensive transcription deregulation, accumulation of unstable transcripts and ncRNAs and defects in heterochromatin formation (Chan *et al.*, 2014; Millbury *et al.*, 2019; Davidson *et al.*, 2019; Murakami *et al.*, 2007). Therefore, it is possible that the accumulation of unscheduled ncRNA or non-degraded unstable transcripts would favour RNA hybridization to DNA. If this is the case, the increase in

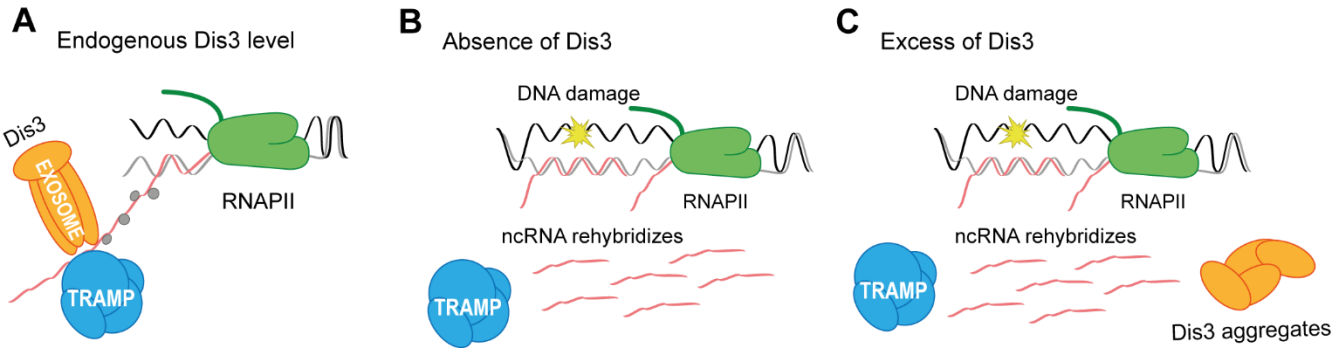


Figure D1. Proposed mechanism for *DIS3* overexpression DNA damage.
(A) At endogenous levels, Dis3 is part of the exosome complex and contributes to degrade non-coding aberrant RNAs together with the TRAMP complex. Lack of functional Dis3 by either its inhibition in a conditional mutant **(B)** or by an excess that may lead to protein aggregation **(C)**, disrupts some exosome functions, accumulating non-coding and aberrant RNAs that could potentially re-hybridize with the template DNA strand, producing an unspecific increase in DNA:RNA hybrids in all the genome and the decondensation of the centromeric regions of the chromosomes.

DNA:RNA hybrids would not appear in genes that tend to accumulate R-loops naturally, but instead, in ncRNAs genes. This could explain why overexpression of *DIS3* presents only a slight but still significant, increase in DNA:RNA hybrids measured by IF that could not be detected in previously reported R-loop accumulating genes by DRIP (Figure D1). In future analyses, it would be convenient to specifically check by DRIP ncRNA regions to test this hypothesis.

An alternative mechanism could be related with changes in chromatin derived from the absence of a functional exosome. Although *S. cerevisiae* lacks the RNAi machinery, heterochromatin regions, like telomeres and rDNA, are transcribed and those heterochromatic RNAs are involved in silencing (Wyers *et al.*, 2005; Vasiljeva *et al.*, 2008). Importantly, heterochromatic RNAs turnover is critical for silencing in a process mediated by the exosome and aided by TRAMP (Houseley *et al.*, 2007; Bühler *et al.*, 2007; Murakami *et al.*, 2007; Vasiljeva *et al.*, 2008). Thus, mutations in the exosome or TRAMP induce changes on chromatin condensation as a consequence of transcription deregulation (Coy & Vasiljeva 2010; Murakami *et al.*, 2007). In *S. cerevisiae* the stabilization of heterochromatic RNAs produced by the lack of the Rrp6 or Rrp4 exosome subunits affects the establishment of heterochromatin (Vasiljeva *et al.*, 2008). Similarly, *trf4* mutants, that are not able to properly link NNS termination machinery with the exosome, or the *nrd1* mutant itself present heterochromatic transcripts accumulation (Houseley *et al.*, 2007; Vasiljeva *et al.*, 2008). The role of TRAMP and the exosome is conserved in organisms that have the RNAi pathway like *S. pombe* or *Drosophila*

melanogaster (Bühler *et al.*, 2007; Eberle *et al.*, 2015). Although the silencing mechanism is unknown, recent studies in *Drosophila* suggest that stabilization of the heterochromatic RNAs competes with chromatin for the interaction of silencing factors like HP1 disrupting the packaging of the heterochromatin (Eberle *et al.*, 2015).

We have recently shown that chromatin plays a crucial role in regulating R-loop formation, as mutations in histone H3 and H4 tails or in chromatin remodelling complexes that renders a more open chromatin facilitates R-loop accumulation (Garcia-Pichardo *et al.*, 2017; Salas-Armentero *et al.*, 2017).Therefore, defects in the exosome activity, by an excess of Dis3 or also by its absence, would lead to deregulation of transcription and the stabilization of heterochromatic transcripts, precluding heterochromatin formation and favouring R-loop formation, as in the case of the *trf4*, *mtr4*, *rrp6* or *rrp4* mutants (Davidson *et al.*, 2019; Gavaldá *et al.*, 2013; Vasiljeva *et al.*, 2008; Houseley *et al.*, 2007).

4.1.2. *SHE2* overexpression contributes to R-loop formation and produces DNA damage

She2 is an RNA-binding protein that co-transcriptionally recognizes specific mRNAs by binding to localization elements (or zipcodes) to guide the transcripts through the nucleolus to a specific localization in the cytoplasm (Böhl *et al.*, 2000; Long *et al.*, 2000; Du *et al.*, 2008; Shen *et al.*, 2010). When She2 recognizes a localization element in the RNA, it binds to it via conformation selection and helps with mRNA folding. This interaction is not very selective until She2 recruits She3 and the type V myosin Myo4 in the cytoplasm, stabilizing the complex. (Böhl *et al.*, 2000; Edelmann *et al.*, 2017). Upon Myo4 interaction, the transcript is transported through actin filaments to the bud tip (Takizawa & Vale 2000).

Contrarily of what we showed for *DIS3*, the phenotype of *SHE2* overexpression differs from its null mutant, as the absence of this protein did not increase DNA damage (Figure R11), suggesting that an excess of this protein may have an specific role inducing genomic instability. Moreover, overabundance of She2 induces R-loop accumulation as seen by IF, that could be specifically detected by DRIP in *ASH1* and *18S* rDNA genes (Figure R9). *ASH1* mRNA contains four localization elements: E1, E2A, E2B and E3 that are redundant, as any of them is sufficient to localize the *ASH1* transcript to the bud tip. These elements are predicted to fold into stem-loop secondary structures that conform the

zipcodes recognized by She2 (Chartrand *et al.*, 1999). Even when *18S* rRNA does not present any of these localization elements, it is described to fold into stem-loop secondary structures that are required for its correct processing (Sharma *et al.*, 1999), potentially sharing structural similarities with the localization elements recognized by She2. This suggests that She2 could have a direct role in R-loop metabolism in genes with zipcode-like structures. Indeed, *ASH1* is not the only target for She2, at least 24 mRNAs have been associated with She2-mediated localization (Takizawa *et al.*, 2000; Shepard *et al.*, 2003). Interestingly, She2 binds each localization element with different affinity (Long *et al.*, 2000). Thus, stoichiometry of She2 and its targets is altered, probably allowing She2 interaction with low-affinity sequences that would not contact at physiological levels, like the ribosomal RNA. In addition, considering the ability of She2 to interact with dsRNA, it is possible that the changes in the stoichiometry of She2 could facilitate its interaction with other double-stranded nucleic acids such as the DNA:RNA hybrids. If this is the case, She2 could be stabilizing or facilitating R-loop formation, possibly explaining the increase observed (Figure R9). Therefore, it would be interesting to check the whether She2 binds DNA:RNA hybrids, *in vitro* by Electrophoretic Mobility Shift Assay (EMSA) and *in vivo* checking whether She2 recruitment to *ASH1* or rDNA is RNase H sensitive.

Although She2 interacts with a small group of mRNAs, it is recruited to all RNAPII-transcribed genes through its interaction with Spt4 at endogenous level (Shen *et al.*, 2010) and when overexpressed (Figure 15). This, together with the changes in stoichiometry caused by the overexpression previously discussed, could promote She2 interaction with new targets. She2 action seems to be restricted to the genes whose transcripts have zipcode-like structures, as RNA:DNA hybrids are only detected in *ASH1* and rDNA by DRIP (Figure R9). Considering all the data, we hypothesised that the overabundance of She2 in the cell changes its stoichiometry favouring its interaction with DNA:RNA hybrids, as the composition of an R-loop is not very different to the double-stranded RNA structures that She2 naturally recognizes. This interaction does not seem to occur in the usual R-loop accumulating genes, but instead is promoted in genes whose transcripts tend to fold in stem-loop secondary structures, probably because they facilitate the contact of She2 with the RNA (Figure D2).

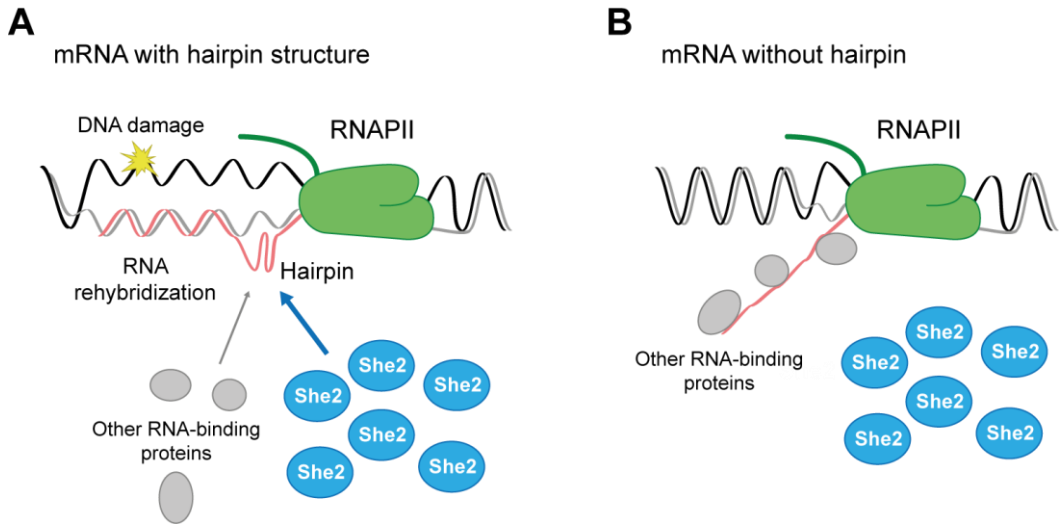


Figure D2. Model for genome instability caused by *SHE2* overexpression.
She2 is a non-highly restrictive RNA-binding factor that strongly depends in the stoichiometry between its substrates and the amount of protein to correctly interact with their target mRNAs. When She2 is in excess, the balance between other RNA-binding proteins and She2 breaks, and She2 is probably able to bind RNA hairpins that are similar to the stem-loops localization elements that it recognizes, stabilizing them and favouring the RNA re-hybridization into the DNA, or maybe affecting elongation and facilitating R-loop formation in those regions. This could lead to DNA damage by ssDNA exposure or perhaps by transcription-replication collisions **(A)**. When no RNA hairpins are present in the nascent RNA, overabundant She2 is probably not able to bind them and induce genomic instability **(B)**.

4.1.3. *RIE1* overexpression changes the distribution of its protein

Rie1/Ygr250c was first identified within the frame of EUROFAN project as an RNA-binding protein whose deletion did not produce any alteration in growth (Sartori *et al.*, 2000; Feroli *et al.*, 1997). Later on, it was reported that Rie1 localizes to stress granules with as Hrp1, Gbp2 or Pab1 among other proteins (Buchan *et al.*, 2008) and that it contains three RNA-recognition motifs (RRMs). Rie1 RRM domains present homology with other RNA-processing proteins as Pap1 or Hrp1, some of them with reported genomic instability phenotypes as Ist3, Nsr1, Npl3 or Rna15 (Schmidlin *et al.*, 2008; Santos-Pereira *et al.*, 2013; Gaillard & Aguilera 2014; M. San Martín–Alonso personal communication) (Figure D3-A). Interestingly, Hrp1 was also identified in our screening, but the DNA damage caused by its overexpression was not R-loop dependent (Figure R7). Hrp1 is part of the cleavage factor complex I (CFI) and can be found in both, nucleus and cytoplasm as part of the mRNP (Kessler *et al.*, 1997) while Pab1, is a poly(A)-binding protein that interacts with CFI and controls the length of the 3' poly(A) tail. Pab1 is

described to associate with Rie1 through domain P that is also required for RNA deadenylation (Richardson *et al.*, 2012).

We identified in our screening that *RIE1* overexpression could have a role in R-loop metabolism. An excess of Rie1 induces DNA damage, that could be partially supressed by RNase H (Figure R7), and it produces a severe growth inhibition in the cell (Figures R6, R7 and R13). In contrast, *rie1Δ* mutant does not reproduce these phenotypes, neither the increase in DNA damage (Figure R11) nor the growth defect (Sartori *et al.*, 2000). Therefore, the excess of Rie1 affects the cell in a different manner that its absence.

Although Rie1 was described as cytoplasmic in a screening of a collection of GFP-tagged genes expressed at physiological level (Huh *et al.*, 2003) we have shown that when it is overexpressed, Rie1 shuttles to the nucleus (Figure R16). Rie1 nuclear localization was even higher in the R-loop accumulating mutant *hpr1* (Figure R16). In this line, we have observed that mRNPs isolated from an hpr1 mutant contain more Rie1 than wild-type mRNPs (A. Rondón, personal communication). This together with the increase in R-loops (Figure R9) and R-loop dependent DNA damage (Figure R7-A) that we observe upon Rie1 overexpression lead us to propose that an excess of this protein produce its inappropriate entry into the nucleus, where it may interfere with RNA metabolism inducing R-loop accumulation. We can envision two alternative mechanism: either by altering the function other RNA-processing proteins like Pab1 directly interacting with them; or perhaps competing with other proteins with similar RRM domains (Pab1, Npl3...) for mRNA interaction. Both situations could affect a proper mRNP formation generating R-loops (Figure D3-B). Interestingly, mutation in *NPL3* or *NSR1* that encode for proteins with similar RRM domains increases R-loop in the cell (Santos-Pereira *et al.*, 2014; M. San Martín–Alonso personal communication). Thus, Rie1 entrance in the nucleus could be inducing R-loops by sequestering some of these proteins or altering their function. It would be interesting to check whether artificially directing Rie1 to the nucleus by inserting a nuclear localization signal in it would recapitulate some of the phenotypes described for the overexpression, including R-loop formation. To gain further insight into the mechanism that lead to genome instability we should check whether nuclear Rie1 could alter the recruitment of Pab1 and other proteins to the mRNP or whether it can directly associate with DNA:RNA hybrids.

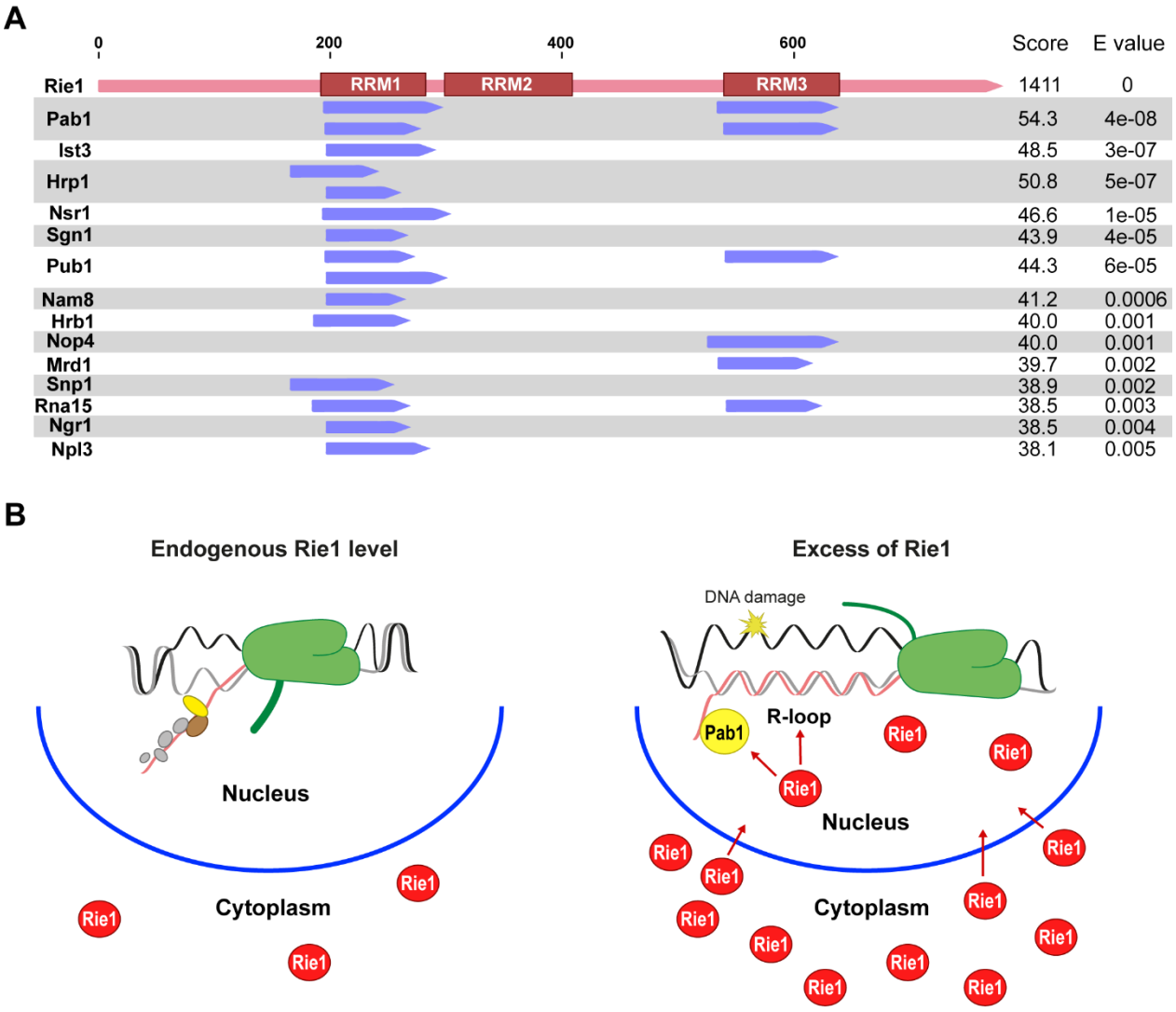


Figure D3. Proposed mechanism for genomic instability produced by *RIE* overexpression.
(A) Alignment of Rie1, showing homology in the RRM domains with other nuclear or cytoplasmic RNA-binding proteins that have a role in RNA processing like Pab1, Ist3, Hrp1, Nsr1, Rna15 or Npl3. **(B)** In our current model we hypothesize that, at endogenous level, Rie1 is mainly at the cytoplasm (left). However, when *RIE1* is overexpressed, it enters into the nucleus and it interferes with RNA metabolism by direct binding with the mRNA or the DNA:RNA hybrids through its 3 RRM domains, or maybe by blocking other RNA-processing proteins as Pab1.

4.1.4. *DIS3, SHE2 and RIE1* genomic instability mechanisms differ from *YRA1*

To sum up, we think that the mechanism for R-loop mediated genome instability caused by *YRA1*, *DIS3*, *RIE1* or *SHE2* overexpression is probably different. On one hand, Yra1 overexpression accumulates R-loop in all the regions assayed while She2 and Dis3 seems to be restricted to specific ones. Moreover, the level of DNA:RNA hybrids detected in each situation also differs, probably explaining why *YRA1* overexpression induces

hyperrecombination while excess of the other factors do not. We concluded that, even when an excess of other RNA-binding proteins as Rie1, She2 or Dis3 could induce genomic instability by increasing DNA damage in an R-loop dependent manner, the mechanism is probably different to *YRA1* overexpression and related to their function in mRNP biogenesis. Rie1 artefactual nuclear localization alters the mRNP composition, while an excess of She2 accumulates R-loops specifically in genes that encode RNAs with stem-loop motifs, differently to Yra1, that it is recruited to any R-loop accumulating gene, and Dis3 by altering exosome function either increases RNA availability or chromatin accessibility. On the contrary, Yra1 stabilizes previously formed R-loops (García-Rubio *et al.*, 2018).

4.2. R-loops generate transient RNAPII stalling

Although the majority of the DNA in the cell is in a B form, it is also able to assume non-B forms as G-quadruplexes (G4) or R-loops at specific regions or under certain conditions. A G4 consists in staked groups of four guanines interacting with each other through a cyclic hydrogen-bound, forming a four-stranded helical structure (Rhodes & Lipps 2015; Maizels & Gray 2013; Sen & Gilbert 1988). Considering that G4 forms in G-rich DNA strands and RNA:DNA hybrids are favoured in C-rich strands both structures could be present at the same DNA region forming a G-loop.

We were interested in assessing whether RNAPII could transcribe through R-loops. Previous *in vitro* evidence showed that transcription by T7 polymerase of a plasmid with 15 repeats of 20 bp from murine immunoglobulin heavy chain switch region Sμ cloned formed G-loops and these structures reduced the amount of total transcript and leaded to the presence of shorter mRNAs (Sen & Gilbert 1988). The result was reproduced with the mammalian RNAPII, but in this case transcription arrest could only be detected after several rounds of transcription (Tornaletti *et al.*, 2008). In addition, transcription of covalently closed circular DNA templates by T3 RNA polymerase is at least partly inhibited by R-loop formation (Bentin *et al.*, 2005) and *in vitro* transcription assay with yeast whole cell extracts of a pre-formed DNA:RNA hybrid template showed a reduction in the amount of transcript generated (Tous & Aguilera 2007), suggesting defects in transcription produced by the presence of these structures in the DNA.

Despite these *in vitro* results, there is little evidence to date of the effect of this non-B DNA structures *in vivo*. The relevance of these *in vitro* experiments needs to be evaluated, since only 0.4–8.9% of potential non-B-forming DNA sequences actually fold into detectable secondary structures *in vivo* (Kouzine *et al.*, 2017), suggesting that *in vitro* conditions did not recapitulate the *in vivo* environment as cells contain different mechanisms to impede and prevent their formation. We have generated a molecular system in the budding yeast genome to study the effect of R-loops *in vivo* during transcription, focusing on the fate of RNAPII. The system generated contains the *LYS2* gene under a strong and inducible promoter, interrupted by a 350 bp fragment of Ig Sμ sequence (*GAL1:LYS2:Sμ350*). The Sμ mammalian Ig heavy chain switch (S) region introduced consists of repetitive G-rich sequences. Formation of DNA:RNA hybrids and G-quadruplexes in Sμ350 and Sμ1050 regions transcribed *in vitro* was previously shown by electron microscopy (Duquette *et al.*, 2004). Indeed, we confirmed by DRIP that the *GAL1:LYS2:Sμ350* system accumulated R-loops when transcribed (Figure R19). This increase in R-loops could be detected in a wild-type strain and, strikingly, it was not enhanced by *hpr1* or *rnh1 rnh201* mutations, that eliminate systems to prevent or remove DNA:RNA hybrids (Huertas & Aguilera 2003; Wahba *et al.*, 2016), suggesting that probably Sμ350 sequence could be enough to induce a high level of R-loops by itself (Figure R19). The reported accumulation of R-loops, consistent with other yeast DRIP assays, was much lower than the level achieved during the *in vitro* experiments for the same sequence (where 42% of the molecules forming G-loops were measured) (Duquette *et al.*, 2004). This low level of R-loops seems not to be enough to cause changes in RNAPII profile in the wild type *GAL1:LYS2:Sμ350* in a steady-state analysis. However, a slight but significant RNAPII accumulation could be measured upstream the *Sμ350* sequence in the *hpr1* mutant (Figure R20), or if we analyse RNAPII elongation rate by measuring the progression of the last round of transcription minutes after switching it off (Figure R21).

We found that *GAL1:LYS2:Sμ350* accumulates RNAPII in both, wild type and *hpr1* backgrounds, being the effect higher in the mutant, supporting the idea of an elongation defect (Figure R21). This data agrees with previous Run-on and ChIP experiments in the entire Sμ region of B cells from Ung-/- mice (in order to avoid class-switch recombination), that showed higher RNAPII signal in the regions flanking the G-rich sequence in the presence of AID, suggesting that Sμ sequence affects transcription

and the presence of non-B DNA structures retards elongation of the polymerases (Rajagopal *et al.*, 2009). Conversely, in our system the RNAPII accumulation reported is not AID-dependent and it is produced by a shorter sequence, which may reproduce more faithfully the effect of a normal R-loop in the cell, as opposed to the RNAPII block produced by the Ig Sμ entire sequence (5 Kb) that takes place during class-switch recombination (Rajagopal *et al.*, 2009). Moreover, the change in RNAPII elongation rate that we reported could be supressed by RNase H expression, indicating that it is a direct consequence of R-loop accumulation (Figure R22). These findings suggest that physiological R-loops are able to stall RNAPII, leading to transcriptional elongation defects.

Finally, the kinetic of the RNAPII profile in the *GAL1:LYS2:Sμ350* system shows that RNAPII level decreases equally throughout the *LYS2* gene over time. This suggests that not all the RNAPII molecules are reaching the end of the gene. Supporting this idea, the amount of transcripts in *GAL1:LYS2:Sμ350* system was significantly lower than in the system without *Sμ350* (*GAL1:LYS2*), indicating that the incomplete transcription of *LYS2* gene is producing unstable mRNAs (Figure R23). Therefore, it seems that stalled RNAPII may being removed from the DNA. There are two alternative ways of eliminating the polymerase, through termination or degradation. We initially examined if transcription could be prematurely terminating. The two mayor termination mechanisms are the poly(A)-dependent or the NNS (Nrd1-Nab3-Sen1) mediated pathway. NNS termination usually takes place inside the gene body to degrade incorrectly processed or cryptic transcripts, and it is more likely to occur in the studied conditions. Thus, we analysed by ChIP the level of Nrd1 protein at *GAL1:LYS2:Sμ350* system in the conditions in which we observed the RNAPII peak, but no specific recruitment of this termination factor could be detected (Figure R24). Considering that Nrd1 is recruited to the system, even though it is not preferentially interacting with the Sμ region, it would be interesting to check the RNAPII profile in the absence of the NNS complex to rule out its implication in disengaging stalled RNAPII. Also, to get further insight into the process we should analyse whether the poly(A)-dependent termination, coupled to the uncapping of the transcript, could occurring in the Sμ region (Jimeno-González *et al.*, 2010).

An alternative possibility is that stalled RNAPII is being removed by degradation. In human cells, NEDD4 and TFIIS localize to the 5' of the Sμ sequence, indicating that stalled RNAPII may undergo backtracking and/or ubiquitination-mediated

destabilization (Sun *et al.*, 2013). During transcriptional stress, NEDD4/Rsp5, an E3 ligase, promotes degradation of RNAPII by the proteasome. Depletion of human NEDD4 increases RNAPII level in the Sμ sequence and the amount of transcript, pointing to an ubiquitin-mediated RNAPII removal in these conditions (Sun *et al.*, 2013). In yeast, Def1 mediates RNAPII degradation under transcriptional stress, being recruited to the Rpb1 monoubiquitylated by Rsp5 to allow the recruitment of the Elongin-Cullin E3 ligase complex (Elc1, Ela1, Cul3, Rbx1) (Wilson *et al.*, 2013). Whether this ubiquitylation mechanism is unique for the Ig Sμ region and directed by AID or if it takes place in any RNAPII stalled as a consequence of R-loop formation, needs to be elucidated, but having into account that ubiquitylation-dependent RNAPII removal mediated by Rsp5/NEDD4 is also produced in response to DSB-induced transcriptional repression and during UV light damage (Woudstra *et al.*, 2002; Pankotai *et al.*, 2012), this could suggest that

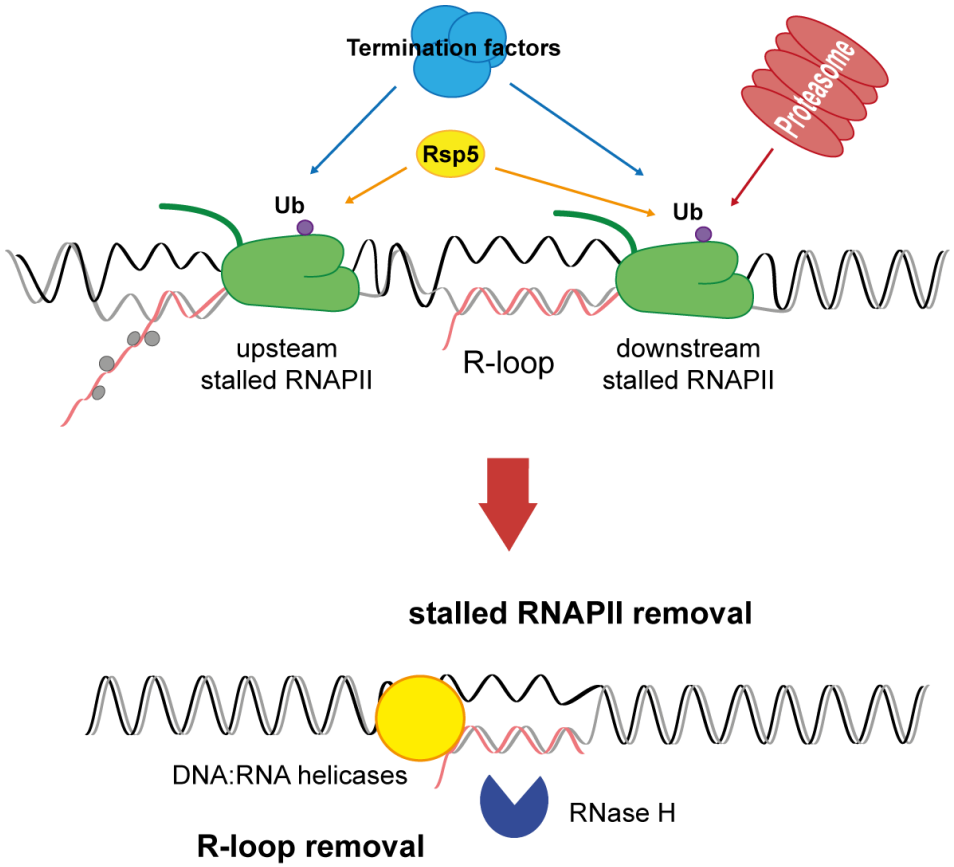


Figure D4. R-loops stall RNAPII and mediates its removal.
R-loop is able to stall RNAPII downstream the forming sequence. Upstream RNAPIIs also have their elongation impeded, probably by torsional stress. Stalled RNAPII are being removed from the chromatin, maybe by a premature termination mechanism or perhaps by ubiquitylation mediated by Rsp5 and proteasome degradation. This would allow to remove more efficiently the R-loops and to resume transcription by subsequent RNAPII molecules.

degradation has general role in resolution of RNAPII blockages. We would like to check the ubiquitination state of the stalled RNAPII or the recruitment of Rsp5 and Def1 to the R-loop accumulating sequence to determine if the RNAPII stalled as a consequence of an R-loop in the *GAL1:LYS2:Sμ350* system could be ubiquitinated and degraded.

To conclude, we have generated a new system that allow us to measure *in vivo* RNAPII response to the accumulation of DNA:RNA hybrids. We propose that in these conditions, RNAPII stalls at the R-loop, as reported by the increase in RNAPII levels. This stall appears to be transient and rapidly resolved by the cell, probably by removing RNAPII, as suggested by the quick drop in RNAPII level after its accumulation and by the reduction in the amount of transcript (Figure D4). If this removal is produced by premature termination or by RNAPII degradation still need to be elucidated, but NNS-dependent pathway does not seem to be involved.

4.3. SSBs in the template strand impair transcription

Single-strand breaks (SSBs) are one of the most common sources of damage in the cell, consisting in disruptions of the DNA backbone in one strand, that are usually accompanied by the loss of a single nucleotide and/or damage of the 3′ and 5′ end of the break. One of the sources of SSBs is the abortive activity of topoisomerase I endonuclease. Topoisomerase I (Top1) creates a cleave complex that induces a transient nick in the DNA in order to reduce torsional stress during transcription or replication. In some cases, an incomplete reaction that does not resolve the SSB or some compounds, like camptothecin that interfere with the process, leaves Top1 covalently bounded to the nicked DNA (Pommier *et al.*, 2003). We can reproduce a situation similar to a Top1-bound SSB with the Flp-nick system that employs the flipase recombinase mutant FlpH305L to induce SSBs irreversible bounded by the flipase at a specific genomic sequence: the flipase recognition target sequence (FRT) (Nielsen *et al.*, 2009).

To study *in vivo* the effect of SSBs over RNAPII-mediated transcription, we generated two molecular systems in which we could control transcription of the *LYS2* gene that contains an FRT site, each system in a different orientation, to induce SSBs in the template or non-template strand of the DNA using the Flp-nick system: *GAL1:LYS2:FRTt* and *GAL1:LYS2:FRTnt.* The observation that RNAPII accumulates in the system that induces SSB in the template strand (*GAL1:LYS2:FRTt*) but not in the

system with the SSB in the non-template strand (*GAL1:LYS2:FRTnt*) (Figure R28) strongly suggests that the integrity of the template strand is crucial during transcription. In agreement with our results, *in vitro* transcription of a nicked DNA template with SP6 and T7 RNA polymerases and run-off analysis of RNA polymerase II from HeLa nuclear extracts resulted in the production of truncated transcripts (Zhou & Doetsch 1993; Kathe *et al.*, 2004; Neil *et al.*, 2012). In the *in vitro* assays, 100% of the template DNA molecules contain SSBs, and RNAPII did not have the required factors to resolve the damage. In these conditions, the inferred percentage of RNAPII stalled could reach the 90% (Kathe *et al.*, 2004). Contrarily, in our *in vivo* assay, we detected RNAPII accumulation during steady-state transcription when only a fraction (2.0-3.2%) of the analysed population contains SSBs (Figure R27), what probably implies a strong inhibitory effect of this type of damage. Moreover, when we induced transcription after FlpH304L expression and we measured RNAPII elongation during the first rounds of transcription, we detected that RNAPII stalled preferentially in the region upstream the SSB specifically in the *GAL1:LYS2:FRTt* system, that produces SSB in the template strand, while the RNAPII level downstream the SSB is similar in *GAL1:LYS2:FRTt*, *GAL1:LYS2:FRTnt or GAL1:LYS2* systems (Figure R29), demonstrating that the presence of a single SSB in the template strand blocks RNA polymerase II preferentially upstream of the SSB. However, when transcription of the damaged template extends in time, RNAPII accumulates not only upstream but also downstream of the SSB (Figure R28-A). This could be a consequence of chromatin changes triggered by the stalled RNAPIIs, that occurs during the repair of the nick and that may affect even the RNAPIIs that have already pass the damage and placed downstream of it. Although little is known on how SSBs alter the chromatin, repair of different lesion like UV damage or DSBs is linked to chromatin modification, where the NER machinery recruits remodelers, histone methyltransferases that set up H3K9 methylation (Ayrapetov *et al.*, 2014), Polycomb group (PcG) transcriptional repressors, DNA-PK, ATM/Tel1 or ATR/Mec1, all of them mediating the ubiquitylation of histone H2A (Vissers *et al.*, 2012; Matsuzaki *et al.*, 2008; Heyer *et al.*, 2010), that are marks of repressive chromatin. It would be interesting to check the state of the chromatin in the vicinity of the SSB to understand the mechanism of RNAPII stalling.

Finally, we showed that *LYS2* transcript level was reduced by induction of FlpH305L when a SSB is produced specifically in the template strand, while no differences in the amount of mRNA were observed if the SSB is localized in the non-template (Figure R30), suggesting that stalled RNAPII does not reach the 3'-end of the gene. We hypothesize that stalled RNAPIIs may be removed or eliminated from the DNA in order to repair the damage. In the case of covalently bound Top1-DNA complexes, that is similar to Flp-nick induced damage, it has been proposed that they could be repaired by transcription-coupled repair (TCR), as the transcription inhibitors 5,6-dichloro-1-β-D-ribofuranosyl benzimidazole (DRB) and α-amanitin block degradation of Top1 by the proteasome, in which transcription is recovered after Top1-DNA complexes are degraded (Desai *et al.*, 2003). Indeed, in V79 cells, global RNA polymerase II is hyper-phosphorylated after CPT treatment and, when they exposed to prolonged CPT treatment (6 h) resulted in gradual degradation of RNA Pol II (Desai *et al.*, 2003). Taken together, these observations reinforce the idea that Top1-bounded SSBs as well as Flp-bounded SSBs complexes are probably able to stall RNAPII. The blocked polymerase could be

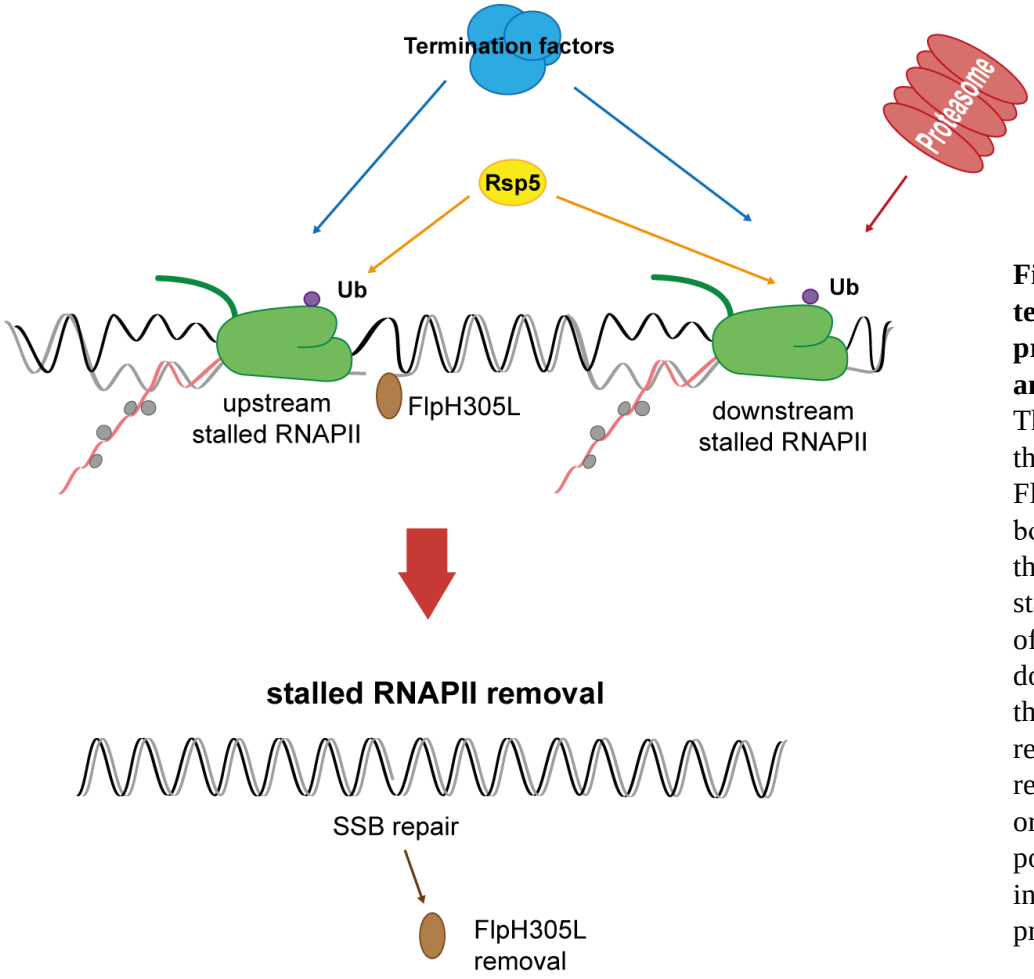


Figure D5. SSBs in the template strand produce RNAPII stall and removal.
The presence of a SSB in the template strand with FlpH305L covalently bounded to the 3'-end of the break, preferentially stalls RNAPII upstream of the damage, but also downstream. Repair of this damage probably requires RNAPII removal by termination or by degradation upon poly-ubiquitylation initiated by Rsp5 and proteasome digestion.

recognized to trigger TCR of the lesion that would mediate the removal of FlpH305L. However, if the damage persists in time, it could lead to RNAPII degradation.

Summarizing, we generated two molecular systems to measure *in vivo* the direct effect of SSBs over RNAPII at a single locus. We showed that RNAPII stalls mainly upstream of the damage site when a SSB is produced in the template strand. This RNA polymerase II stalling is accompanied by a decrease in the transcript levels. SSB in the non-template strand did not produce any noticeable effect neither in RNAPII nor in mRNA levels, suggesting that a sole SSB in the template strand at physiological conditions is able to hinder transcription and RNAPII may be transiently stalled, or degraded to repair the damage (Figure D5). Additional experiments need to be done to elucidate the fate of RNAPII and whether chromatin changes are responsible of the block.

4.4. R-loops and SSBs induce recombination between homologous chromosomes

DNA:RNA hybrids are able to increase homologous recombination between direct repeats, as it has been reported for several R-loop accumulating mutants like *hpr1Δ, sen1-1, mlp1Δ, npl3Δ, trf4Δ* or *rnh1Δ rnh201Δ* (Huertas & Aguilera 2003; Mischo *et al.*, 2011; García-Benítez *et al.*, 2017; Santos-Pereira *et al.*, 2013; Gavaldá *et al.*, 2013; Keskin *et al.*, 2016). In the plasmid systems used so far to analyse recombination, the main process taking place is the single strand annealing (SSA), in which resection of the DSB uncovers homology for the damaged DNA in the same molecule and therefore it does not require strand invasion (Pardo *et al.*, 2009). Despite hyper-recombination being one of the first phenotypes linked to R-loops, little is known about the direct role of these structures in generating DSBs and whether they could be repaired by an homologous recombination mechanism different to SSA. Taking advantage of the system that we have developed to generate R-loops in the *LYS2* gene, we designed a diploid genetic system to study the effect of DNA:RNA hybrids at inducing DNA damage and how this damage could be repaired by an HR mechanism different of SSA (Figure R31): recombination between homologous chromosomes.

When we measured the recombination frequency between homologue chromosomes induced by R-loops, we detected a significant increase respect to the non-transcribed control. However, the recombination frequencies measured in our system

were very low (in the order of 10^{-6}), suggesting a weak role of R-loops at inducing the sort of DNA damage that could be repair specifically by recombination between homologous alleles (Figure R32). This is in agreement with previous analysis performed in systems that measured SSA induced by R-loops transcribing Sμ or Sγ3, that showed low recombination frequencies in wild type conditions and they were exacerbated only in mutants as *mft1Δ* or *hpr1Δ* or in the presence of AID, that induces additional damage (Ruiz *et al.*, 2011; Gan *et al.*, 2011). This suggests that R-loops probably require additional defects in transcription or in R-loop processing to become a threat for genome stability. Another possibility is that most of the damage produced by R-loops could be repair by another pathway like NHEJ or using the sister chromatid instead of the homologous allele. Recombination between homolog chromosomes causes loss of heterozygosity, a phenotype linked to cancer. Therefore, it is not surprising that the cells have developed mechanisms to minimize it. Indeed, loss of heterozygosity in mouse cells is not favoured, as NHEJ was estimated as three order of magnitude higher than HR and recombination with the sister chromatid four orders of magnitude higher than with the homolog allele (Moynahan & Jasin, 1997; Stark & Jasin 2003). But in our system we could not detect the damage repaired by NHEJ or sister chromatid recombination.

Although how R-loops cause DSBs is still unknown, research performed in the last decade strongly suggest that it is mediated by transcription-replication conflicts (TRCs) (Castellano-Pozo *et al.*, 2013; Garcia-Rubio *et al.*, 2018). R-loops enhance TRCs when transcription is orientated head-on to replication or when they are artificially stabilized by interacting proteins (Prado & Aguilera 2005; Garcia-Rubio *et al.*, 2018). When it is co-directional, the replisome favours R-loop resolution, diminishing the damage generated (Hamperl *et al.*, 2017). Since in our system the closer replication origin is *ARS215*, that is located 12.7 kb upstream of the *LYS2* gene, the TRC that it would be produced by RNAPII stalled at the Sμ region would be co-directional and not head-on (Figure R31). Therefore, R-loops produced in this situation could potentially be resolved by the replisome, decreasing the amount of damage and HR produced. To confirm this idea, it would be interesting to invert the system to favour head-on collisions or to artificially stabilize the R-loops by overexpressing Yra1.

In parallel to the study the implications of R-loops at inducing recombination between homolog chromosomes, we analysed the effect of the Flp-nick system. To do so, we generated two diploid strains containing the *LYS2* with the FRT sequence integrated

in either orientations in one chromosome and a *lys2* point mutation in the other allele (Figure R31). Although flipase cleavage generates SSBs, they become DSBs upon replication (Nielsen *et al.*, 2009). Indeed, mutant strains of the homologous recombination machinery are highly sensitive to camptothecin while induction of a single SSB with the Flp-nick system reduced growth of *rad52* and *rad51* mutants (Nielsen *et al.*, 2009). Therefore, it is not surprising that we found no differences in recombination frequencies independently of the strand in which the FlpH305L produced SSBs (Figure R33). Induction of FlpH305L produced an increase of 3 to 4 orders of magnitude in homologous recombination. Interestingly, we measured one order of magnitude decrease in allelic recombination during *LYS2* transcription respect of the non-transcribed condition. This reduction in recombination was also independent of the template or non-template orientation of the FRT site (Figure R33). We propose that FlpH305L recruitment maybe partially impeded by the chromatin configuration during transcription or by the RNAPII itself, and the flipase is probably not able to gain access to the FRT sequence. In order to check this hypothesis, it would be interesting to quantify SSBs with or without transcription. Another intriguing alternative could be that transcription would favour SSBs repair by a TCR mechanism, eliminating the damage before replication converts it into a DSB and thus reducing the homologous recombination product. However, if this is the case, we would probably expect a reduction only when SSBs occur in the template strand and not in both.

Using the systems generated to measure homologous chromosome recombination would allow us to check for different TCR mutants that could be involved in the signalling or the repair of this damage in order to elucidate the mechanism of the RNAPII stalling and the resolution of the SSB and R-loop-induced DNA damage.

Overall, this thesis provides evidence on how the imbalance of the mRNA biogenesis induces genomic instability associated to R-loop accumulation, extending our knowledge on the connexion between mRNP assembly and R-loop formation. This thesis also opens new perspectives on how R-loops and SSBs affect transcription, demonstrating that RNAPII elongation is impeded by the presence of R-loops at physiological levels but also by a sole SSB in the DNA template strand. We have developed several systems that would allow as to check at molecular and genetic level how RNAPII is affected by the formation of these obstacles and to identify the machineries involved in their resolution.

These two approaches allowed us to better understand the mechanisms that trigger genomic instability in the cell.

5. CONCLUSIONS

1 – *DIS3* overexpression causes R-loop accumulation, increases R-loop-dependent DNA damage and affects rRNA processing similarly to *dis3* mutants.

2 – An excess of She2 increases R-loop accumulation in specific genes, increasing DNA damage.

3 – Rie1 enters into the nucleus when it is in excess, causing R-loop accumulation and DNA damage.

4 – At physiological levels, R-loops are able to transiently stall RNAPII, reducing the elongation rate and the amount of transcript.

5 – Induction of a SSB in the transcriptional template strand stalls RNAPII preferentially upstream of the damage and reduces the level of transcript.

6 – R-loops induce recombination between homologous chromosomes, but they are a poor source of DNA damage.

7 – Transcription negatively impacts the recombination frequency produced by SSBs independently of whether they occur in the template or non-template strand.

6. MATERIALS AND METHODS

6.1. Growth media and conditions

6.1.1. Bacteria culture media

- *Rich medium LB:* 0.5% yeast extract, 1% bacto-tryptone, 1% NaCl. When necessary, medium was supplemented with 100 μg/ml of ampicillin.

6.1.2. Yeast culture media

- *Rich medium, YPAD:* 1% yeast extract, 2% bacto-tryptone, 2% glucose, 20 mg/L adenine.
- *Minimum medium, SD:* 0.17% yeast nitrogen base (YNB) without amino acids nor ammonium sulfate, 0.5% ammonium sulfate, 2% glucose.
- *Complete medium, SC:* SD medium supplemented with the amino acids: leucine, tryptophan, histidine, lysine, methionine, aspartate and threonine and the nitrogen bases adenine and uracil in concentrations described in (Sherman *et al.*, 1986). The absence of one or more of these requirements is specified.
- *Complete medium, SGal*: SC medium containing 2% galactose instead of 2% glucose as carbon source.
- *Complete medium, SRaf*: SC medium containing 2% raffinose instead of 2% glucose as carbon source.
- *Complete medium SC+FOA*: SC medium with half concentration of uracil (10 mg/L) supplemented with 500 mg/L 5-Fluoroorotic acid (FOA) and 0.1% L-proline instead of ammonium sulfate as nitrogen source. 5-FOA was added to autoclaved medium cooled down to 60ºC.
- *Sporulation medium, SPO*: 1% potassium acid, 0.1% yeast extract, 0.005% glucose, supplemented with a quarter concentration of the requirements described for SC medium.
- Solid mediums were prepared adding 2% agar before autoclaving.

6.1.3. Growth conditions

Bacteria strains were grown at 37ºC in all cases. Yeast strains were incubated at 30ºC, except when specified. Liquid cultures were incubated on horizontal orbital shakers at 200 rpm. Diploid yeast strains were sporulated at 26ºC in SPO medium for 3-4 days.

6.2. Antibiotics, drugs, inhibitors, enzymes and antibodies

6.2.1. Antibiotics

- *Ampicilin, Amp* (Sigma): β-lactam antibiotic that inhibits cell division in *E. coli* preventing the cell wall synthesis. It was used to select bacteria cells carrying a plasmid. Working concentration: 100 μg/ml.
- *Doxycyclin, DOX* (Sigma): Tetracyclin family antibiotic used as transcription repressor or activator of yeast genes under the bacterial *tet* promoter.
- *G418, Geneticin* (USB): Aminoglycoside antibiotic that inhibits protein synthesis by binding to the ribosome 80S subunit (Jimenez and Davies, 1980). It was used in yeast strains to select, follow and maintain the kanamycin resistance *E. coli* gene *KAN*. Working concentration: 100 μg/ml.
- *Hygromycin B, Hyg* (Roche): Aminoglycoside antibiotic that inhibits protein synthesis by introducing the misreading of the mRNA. It was used in yeast strains to select, follow and maintain the hygromycin resistance *E. coli* gene *HPH* (Gritz and Davies, 1983). Working concentration: 250 μg/ml.
- *Nourseothricin, Nat* (Werner BioAgents): Aminoglycoside antibiotic that inhibits protein synthesis by inducing miscoding. It was used to select, follow and maintain the nourseothricin resistance gene *NAT* (Krügel *et al.*, 1993). Working concentration: 100 μg/ml.

6.2.2. Drugs and inhibitors

- *1-Naphthaleneacetic acid, NAA* (Sigma): Synthetic phytohormone of the auxin family. Allow to deplete a target protein with a degradation domain (degron) using the auxin-dependent degradation (AID) pathway from plants. Working concentration: 0.5 mM.

- *5-fluorotic acid, FOA:* Toxic analog of uracil that poison *URA3* yeasts but not *ura3* mutants (Boeke *et al.*, 1984). Working concentration: 500 mg/L.
- *(S)-(+)-Camptothecin, CPT* (Sigma): It traps topoisomerase I enzyme in a covalent linkage with DNA. Working concentration: 20 μg/ml.
- *Complete protease inhibitor cocktail* (Roche): Mix of several inhibitors of serine and cysteine proteases, but not metalloproteases. It was used according to manufacturer's recommendations.
- *Diethyl pyrocarbonate, DEPC* (Sigma): RNase inhibitor. Working concentration: 1/1000 v/v.
- *Hydroxyurea, HU* (USB): Compound that blocks the synthesis of deoxynucleotides, inhibiting DNA synthesis. It inactivates the ribonucleotide reductase, forming a nitroxide free radical that binds a tyrosyl free radical in the active site of the enzyme. Working concentration: 100 mM.
- *Phenylmethanesulfonyl fluoride, PMSF* (Sigma): Inhibitor of serine (trypsin and chymotrypsin) and cysteine proteases. Working concentration: 1 mM.

6.2.3. Enzymes

- *Alkaline phosphatase* (Roche): Hidrolyzes 5'-monophosphate groups from DNA ends generated after an enzymatic cut. Dephosphorylating a cut vector prevents religation.
- *iTaq™ universal SYBR® green supermix* (Bio-Rad): Mix for quantitative PCR amplification that contains the ampliTaq Gold®DNA polymerase and the LD DNA polymerase, dNTPs with a dUTP/dTTP mixture and the ROX fluorochrome, used in an optimized buffer for the qPCR reaction.
- *Klenow* (Roche): Major fragment of the *E. coli* DNA polymerase I, with 5'-3' polymerase and 3'-5' exonuclease activities.
- *Lysozyme* (Sigma): Enzyme from chicken egg white that hydrolyzes bacterial peptidoglycans, breaking down the cell wall.
- *MyTaq™ DNA polymerase* (Promega): Fast DNA polymerase without proofreading nor high processivity used for DNA probes and checking PCRs.

- *Phusion® high-fidelity DNA polymerase* (Finnzymes): A *Pyrococcus*-lile polymerase fused with a processivity-enhancing domain.
- *Pronase* (Sigma): Non-specific mix of proteases isolated from *Streptomyces griseus*.
- *Protein A/G Dynabeads™* (Invitrogen): Magnetic beads with recombinant Protein A or G coupled to its surface. Protein A/G binds to the Fc region of IgA, IgG and IgM inmunoglobulins.
- *Proteinase K* (Roche): Serine protease from *Pichia pastoris*.
- *Q5™ high-fidelity DNA polymerase* (New England BioLabs): High-fidelity, low error rate DNA polymerase with 3'-5' exonuclease activity, fused to a processivity enhancing Sso7d domain.
- *QuantiTect Reverse Transcription Kit* (Qiagen): Kit for cDNA synthesis with oligo-dT and random primers and for genomic DNA removal. It contains a mix of recombinant heterodimeric reverse transcriptases.
- *Restriction enzymes* (New England Biolabs and Takara): Different sequence-specific DNA endonucleases.
- *RNase A* (Sigma): Endonuclease that degrades single-stranded RNA.
- *RNase H* (New England Biolabs): Endonuclease that specifically hydrolyzes the phosphodiester bonds of RNA hybridized to DNA but does not digest single or double-stranded DNA.
- *RNase III Ambion™* (Invitrogen): Double-stranded RNA specific endoribonuclease from *E. coli*.
- *T4 phage DNA ligase* (Roche): It catalyzes the covalent binding of dsDNA ends.
- *T4 Polynucleotide Kinase* (Roche): It phosphorylates the 5′ and 3' ends of RNA or DNA.
- *Zymoliase 20T* (USB): Mix of enzymes from *Arthrobacter luteus* to digest *S. cerevisiae* cell wall to produce spheroplast. Working concentration: 2 mg/ml.

6.2.4. Antibodies

Antibodies used in this thesis are listed in Table M1 and where used following the manufacturer's recommendations.

Table M1. Primary antibodies

Antibody	Source	Epitope	Reference	Use
Anti-β-Actin	Rabbit		ab8227 (Abcam)	WB (1:2000)
Anti-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)	ChIP (10µg)
Anti-HA	Rabbit		ab9110 (Abcam)	ChIP (10µg) WB (1:5000)
Anti-RNA Pol II (8WG16)	Mouse	C-terminal heptapeptide repeat (CTD) of Rpb1, the largest subunit of RNA Pol II.	664906 (BioLegend)	ChIP (3µg)
RNA pol II CTD phospho Ser2 (3E10)	Rat	C-terminal heptapeptide repeat (CTD) of Rpb1 phosphorylated in Ser2 residue.	61083 (Active motif)	ChIP (5µg)
RNA pol II CTD phospho Ser5 (3E8)	Rat	C-terminal heptapeptide repeat (CTD) of Rpb1 phosphorylated in Ser5 residue.	61085 (Active motif)	ChIP (5µg)
RNA Pol II CTD phospho Thr4 (6D7)	Rat	C-terminal heptapeptide repeat (CTD) of Rpb1 phosphorylated in Thr4 residue.	61361 (Active motif)	ChIP (10µg)
RNA Pol II CTD phospho Tyr1 (3D12)	Rabbit	C-terminal heptapeptide repeat (CTD) of Rpb1 phosphorylated in Tyr1 residue.	MABE350 (Merck)	ChIP (5µg)

Antibody	Source	Epitope	Reference	Use
S9.6	Mouse	DNA-RNA hybrids.	Hybridome cell line HB-8730	ChIP (3µg) IF (1:300)

Table M2. Secondary antibodies

Specificity	Source	Conjugation	Reference	Use
Mouse	Donkey	Cy3	715-165-150 (Jackson laboratories)	IF (1:1000)
Rabbit	Goat	Peroxidase	A6154	IF (1:5000)

6.3. Strains and plasmids

6.3.1. Bacterial strains

All experiments with *E. coli* were done using DH5α strain: *F- endA1 gyr96 hsdR17 lacU169(F80LACZΔm15) recA1 relA1 supE44 thi-1* (Hanahan, 1983), or K-12 ER2925 strain: *ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10)TetS endA1 rpsL136 dam13::Tn9 xylA-5 mtl-1 thi-1 mcrB1 hsdR2* (Woodcock *et al.*, 1989).

6.3.2. Yeast strains

Yeast strains used in this work are listed in the Table M3.

HPR1DGK is a diploid strain obtained by mating of hpr1-d2 and SChY58a.

GLY-2D strain was generated replacing the *LYS2* endogenous promoter with a cassette containing the *NatMX6nt1* selection marker and the *GAL1* promoter, obtained by PCR from the pFA6a-NATnt1-GAL1 plasmid and inserted by homologous recombination (Results 3.2.1).

The strain GLSd-1B was obtained from GLY-2D by inserting a PCR cassette containing the Sµ350 sequence on the position 2942 bp of the *LYS2* ORF using the pML104-LYS2g plasmid (express Cas9 enzyme and a gRNA). The PCR cassette was

amplified from the plasmid pRS413-SF (Results 3.2.2). Strains GLFc-2C and GLFt-3A were generated by the same procedure but integrating the FRT sequence in the non-template or the template strand of the *LYS2* gene, respectively, using a PCR cassette amplified from plasmid pTINVFRT-1 (Results 3.2.9).

Strains GLYH-3D and GLSdH-5D were created by genetic crosses of HPBAR1-R with GLY-2D and GLSd-1B respectively. Strains GLYRH-2 and GLSdRH-1 were obtained deleting the *RNH1* and *RNH201* genes in GLY-2D and GLSd-1B respectively, by homologous recombination with a *rnh1Δ::KanMX6* and a *rnh201Δ::HygMX6* PCR cassettes in a two-step transformation.

Strains GLYN-HA1 and GLSdN-HA1 were obtained from GLY-2D and GLSd-1B, inserting a carboxyterminal 3xHA tag and the KanMx4 cassette in the *NRD1* gene by homologous recombination using a PCR product from the plasmid pFA6a-3HA-KanMX6 with 50 bp of homology in both ends as previously described (Janke *et al.*, 2002).

Strain YLYS2-6, containing a single deletion in the position 5685 of the *LYS2* gene, was obtained from Ypb250 by error prone repair of the damage induced by Cas9 enzyme expressed and directed to the target sequence using the plasmid pML104-3'mut with a specific gRNA cloned, as previously reported (Lemos *et al.*, 2018). The strains DGLSd-A6, DGLFc-D6 and DGLFt-B6 are diploid strains generated by mating of YLYS2-6 and GLSd-1B, GLFc-2C or GLFt-3A, respectively.

Table M3. Yeast strain used in this thesis.

Strain	Genotype	Source
BY4741	*MATa his3Δ leu2Δ0 met15Δ ura3Δ0*	Euroscarf
F15	*Mat α thr1 arg4*	G. Fink
F4	*Mat a thr4*	G. Fink
DGLSd-A6	*ade2-1/ade2-1 can1-100/can1-100 his3-11,15/his3-11,15 leu2-3,112/leu2-3,112 trp1-1/trp1-1 ura3-1/ura3-1 LYS2p::NatMX6-GAL1p lys2::Sμ350/lys2 bar1Δ/bar1Δ RAD5/RAD5*	This study

Strain	Genotype	Source
DGLFc-D6	*ade2-1/ade2-1 can1-100/can1-100 his3-11,15/his3-11,15 leu2-3,112/leu2-3,112 trp1-1/trp1-1 ura3-1/ura3-1 LYS2p::NatMX6-GAL1p lys2::FRTcopy/lys2 bar1Δ/bar1Δ RAD5/RAD5*	This study
DGLFt-B6	*ade2-1/ade2-1 can1-100/can1-100 his3-11,15/his3-11,15 leu2-3,112/leu2-3,112 trp1-1/trp1-1 ura3-1/ura3-1 LYS2p::NatMX6-GAL1p lys2::FRTtranscribed/lys2 bar1Δ/bar1Δ RAD5/RAD5*	This study
DY8107	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 rad5-535 trp1-1 ura3-1 spt16-11*	D. Stillman
GHY94	*MATa his3Δ200 lys2-1288 leu2Δ1 trp1Δ63 ura3-52 spt5-194*	G. Hartzog
GLFc-2C	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 LYS2p::NatMX6-GAL1p lys2::FRTcopy bar1Δ RAD5*	This study
GLFt-3A	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 LYS2p::NatMX6-GAL1p lys2::FRTtranscribed bar1Δ RAD5*	This study
GLSd-1B	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 LYS2p::NatMX6-GAL1p lys2::Sμ350 bar1Δ RAD5*	This study
GLSdH-5D	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 hpr1Δ::HIS3 LYS2p::NatMX6-GAL1p lys2::Sμ350 bar1Δ RAD5*	This study
GLSdN-HA1	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 LYS2p::NatMX6-GAL1p lys2::Sμ350 NRD1::3HA bar1Δ RAD5*	This study
GLSdRH-1	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 LYS2p::NatMX6-GAL1p lys2::Sμ350 rnh1::KanMX6 rnh201::HygMX6 bar1Δ RAD5*	This study
GLY-2D	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 LYS2p::NatMX6-GAL1p bar1Δ RAD5*	This study

Strain	Genotype	Source
GLYH-3D	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 hpr1Δ::HIS3 LYS2p::NatMX6-GAL1p bar1Δ RAD5*	This study
GLYN-HA1	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 LYS2p::NatMX6-GAL1p NRD1::3HA bar1Δ RAD5*	This study
GLYRH-2	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 LYS2p::NatMX6-GAL1p rnh1::KanMX6 rnh201::HygMX6 bar1Δ RAD5*	This study
HPBAR1-R	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 hpr1D::HIS3 bar1Δ RAD5*	M. San Martín
hpr1-d1	*MATa ade2-1 his3-11,15 trp1-1 leu2-3,112 can1-100 ura3-1::ADH1-AtTIR1-9Myc (URA3) hpr1::hpr1-aid (HygMX)*	M. San Martín
hpr1-d2	*MATα ade2-1 his3-11,15 trp1-1 leu2-3,112 can1-100 ura3-1::ADH1-AtTIR1-9Myc (URA3) hpr1::hpr1-aid (HygMX)*	M. San Martín
HPR1DGK	*MATa/MATα ade2-1/ade2-1 his3-11,15/ his3-11,15 trp1-1/trp1-1 leu2-3,112/leu2-3,112 can1-100/can1-100 ura3-1::ADH1-AtTIR1-9Myc (URA3)/ ura3-1 hpr1::hpr1-aid (HygMX)/hpr1Δ::KanMX*	This study
OY97	*MATa his3Δ200 lys2-1288 leu2Δ1 trp1Δ63 ura3-52*	G. Hartzog
RNH2-R	*MATa leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15 rnh1Δ::KanMX rnh201Δ::KanMX RAD5 bar1Δ*	M. San Martín
SChY58a	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 rad5-535 trp1-1 ura3-1 hpr1Δ::KanMX*	S. Chávez
U678-1C	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 rad5-535 trp1-1 ura3-1 hpr1Δ::HIS3*	R. Rothstein
W303-1A	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 rad5-535 trp1-1 ura3-1*	R. Rothstein
W303-1B	*MATα ade2-1 can1-100 his3-11,15 leu2-3,112 rad5-535 trp1-1 ura3-1*	R. Rothstein

Strain	Genotype	Source
WGLZN-3B	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 rad5-535 trp1-1 ura3-1 GL-LacZ::NatMX*	J. LaFuente-Barquero
WMC1-1A	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 rad5-535 trp1-1 ura3-1 mex67-5*	S. Jimeno *et al.* 2002
WPR52-1B	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 rad5-535 trp1-1 ura3-1 rad52Δ*	
WSPT-2D	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 rad5-535 trp1-1 ura3-1 spt4Δ::URA3*	A. Rondón
Ybp249	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 bar1Δ RAD5*	B. Pardo
Ybp250	*MATα ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 bar1Δ RAD5*	B. Pardo
YGL043w	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 dst1Δ::kanMX4*	Euroscarf
YGR063c	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 spt4Δ::kanMX4*	Euroscarf
YGR250C	*MATa his3Δ leu2Δ0 met15Δ ura3Δ0 rie1Δ::kanMX4*	Winzeler *et al.* 1999
YKL130C	*MATa his3Δ leu2Δ0 met15Δ ura3Δ0 she2Δ::kanMX4*	Winzeler *et al.* 1999
YLYS2-6	*MATα ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 lys2 bar1Δ RAD5*	This study
YOL021C	*MATa his3Δ leu2Δ0 met15Δ ura3Δ0 dis3-1*	Boone *et al.* 2011

6.3.3. Plasmids

Plasmids used in this thesis are listed in table M4.

Plasmid YEp351-YRA1Δi was obtained by cloning the BamHI fragment of pFS2146, containing *YRA1Δi* plus 500 bp upstream and downstream, into the BamHI restriction site of YEp351. YEpDIS3 was generated by cloning a PCR cassette of the *DIS3* ORF, including 450 bp upstream and 120 bp downstream of it and SacI-SphI restriction sites, into the SacI-SphI site of YEp351.

pYES-HRP1, pYES-NAB2, pYES-NHP6B, pYES-NPL3, pYES-RIE1, pYES-SHE2, pYES-SNF5 and pYES-SWT1 were all generated by cloning a PCR product from W303-1A of each ORF and specific unique restriction sites contained in the primers. These sites were used to clone them into pYES2 plasmid previously digested with the desired enzymes.

pYES-SHE2:YFP and pYES-RIE1:YFP were generated by amplifying each ORF without the stop codon and followed by four Alanine codons first and the YFP ORF (from pWJ1344) next by two-step PCR with the adaptor primers SHE2-PH rv or RIE1-PH rv respectively, and YFP fw . The PCR products were cloned into pYES2 KpnI-EcoRI sites.

pML104-LYS2g plasmid was generated by cloning the annealed gRNA LYS2 A and gRNA LYS2 B primers into the BglI-SwaI sites of pML104. Plasmid pML104-3'mut was generated by the same procedure but using gRNA LYS2-3' A and gRNA LYS2-3' B primers to target a region 0.75 kb downstream to the first Cas9 target site.

pCM190:FLP vector was obtained by cloning the PCR product containing FlpH305L (from pBIS-GALkFLP) and a BamHI restriction site added to the 5' end, into the BamHI-NotI sites of pCM190.

Table M4. Plasmids used in this thesis.

Plasmid	Description	Source
C17 MW90	YEp351 plasmid from the MW90 library containing an 8.3 Kb insert of Chr. XV from *S. cerevisiae*.	MW90 library
C23 MW90	YEp351 plasmid from the MW90 library containing an 7.2 Kb insert of Chr. VII from *S. cerevisiae*.	MW90 library
MW90 library	Yeast genomic DNA library constructed in YEp351.	Waldherr et al., 1993
p413GALRNH1	*GALp::RNH1* from pGALRH1 cloned into SalI-SpeI sites of pRS413 (*HIS3* marker).	Gómez-González B

Plasmid	Description	Source
pBIS-GALkFLP	pRS414 plasmid with the FlpH305L step-arrest mutant gene under *GAL10* promoter.	Tsalik & Gartenberg, 1998
pCM189-RNH1	*RNH1* under *tetO7-CYC1* control in the pCM189 centromeric plasmid.	Santos Pereira JM
pCM190	Yeast episomal expression vector driven by the *tetO7-CYC1* promoter.	Herrero E 1997
pCM190-FLP	pCM190 plasmid with the FlpH305L step-arrest mutant gene under *tetO7* promoter.	This study
pFA6a-3HA-KanMX6	pFA6-KanMX6 plasmid with 3xHA tag and *CYC1* terminator.	Bahler *et al*., 1998
pFA6-NATnt2+GALp	pFA6-NATnt2 plasmid with *GAL1* promoter and *CYC1* terminator from pYES2 cloned into SpeI-SapI sites.	LaFuente-Barquero J
pFS2146	*YRA1Di*+HA 500bp upstream and downstream cloned in BamHI YEplac22.	Zenklusen *et al*., 2001
pGALRH1	YCp pRS416 containing the *GALp::RNH1* fusion.	RJ Crouch
pML104	pRSII426 that expresses Cas9 and contains a guide RNA expression cassette.	Laughery *et al.*, 2015
pML104-3'mut	pML104 plasmid expressing a 3' *LYS2* gRNA.	This study
pML104-LYS2g	pML104 plasmid expressing a *LYS2* gRNA.	This study
pRS313	YCp vector based on the *HIS3* marker.	Sikorski & Hieter, 1989
pRS314-L	pRS314 containing two direct repeats of *LEU2* gene sharing 600 bp homology.	Prado & Aguilera 1995
pRS314-LYDNS	pRS314-L containing a YIp5 sequence with a 1.92 kb SphI-NsiI deletion, removing the *URA3* gene inserted at BglII site, located between the repeats.	Prado *et al.*, 1997

Plasmid	Description	Source
pRS413-GAL::YRA1Δi	Ycp pRS413 with the *GALp::YRA1Δi* fusion.	Gavaldá *et al.*, 2016
pRS413-SF	*Sμ350* from murine cloned into pRS413 plasmid.	Ruiz *et al.*, 2011
pTINVFRT-1	FRT sequence cloned as replacement into HO site in pRS316-TINV.	Ortega-Moreno P
pWJ1213	YCp centromeric plasmid containing the Rad52::YFP fusion and HIS3 marker.	Feng *et al.*, 2007
pWJ1344	YCp centromeric plasmid containing the Rad52::YFP fusion and *LEU2* marker.	Lisby *et al.*, 2001
pYES2	Multicopy expression vector with *GAL1* promoter with 2m origin and *URA3* marker.	ThermoFisher
pYES-HRP1	*HRP1* ORF from *S. cerevisiae* cloned with EcoRI in pYES2.	This study
pYES-NAB2	*NAB2* ORF from *S. cerevisiae* cloned in pYES2.	This study
pYES-NHP6B	*NHP6B* ORF from *S. cerevisiae* cloned with XhoI in pYES2.	This study
pYES-NPL3	*NPL3* ORF from *S. cerevisiae* cloned with XhoI in pYES2.	This study
pYES-RIE1	*RIE1* ORF from *S. cerevisiae* cloned with KpnI and EcoRI in pYES2.	This study
pYES-RIE1:YFP	pYES2 plasmid containing the *RIE1::YFP* fusion cloned into KpnI-EcoRI sites.	This study
pYES-SHE2	*SHE2* ORF from *S. cerevisiae* cloned with XhoI in pYES2.	This study
pYES-SHE2:YFP	pYES2 plasmid containing the *SHE2::YFP* fusion cloned into KpnI-EcoRI sites.	This study
pYES-SNF5	*SNF5* ORF from *S. cerevisiae* cloned with EcoRI and XhoI in pYES2.	This study

Plasmid	Description	Source
pYES-SWT1	*SWT1* ORF from *S. cerevisiae* cloned with EcoRI and XhoI in pYES2.	This study
pYES-YFP	*YFP* cloned in pYES2.	This study
YEp351	Multicopy plasmid with *LEU2* gene as marker.	Hill *et al.*, 1986
YEp351-NAB2	Fragment of *NAB2* gene cloned in YEp351.	Gallardo *et al.*, 2003
YEp351-NPL3	Fragment of *NPL3* gene cloned in YEp351.	Santos-Pereira *et al.*, 2013
YEp351-SHE2	*SHE2* gene including its promoter and terminator cloned in YEp351 at PstI site.	Rondon A
YEp351-YRA1Δi	YEp351 with *YRA1Δi* +500bp 5' and +500bp 3' BamHI inserted.	This study

6.4. Yeast methodology

Yeast methodology was carried out using standard procedures. Tetrad dissection was performed using a SINGER MSM 200 micromanipulator.

6.4.1. Yeast transformation

Yeast transformation was performed as previously described using the lithium acetate/single-stranded DNA/polyethylene glycol method (Gietz & Schiestl, 2007).

6.4.2. Cas9 gene editing

Cas9 target sites and gRNA sequences were obtained using 'CRISPR Toolset'. gRNA coding sequences were created by hybridizing the primers XXX and XXX of XXbp whose sequence was generated by the software and that were synthesised by Metabion. Plasmid pML104, that constitutively expresses the *Cas9* endonuclease and a specific gRNA cloned into de BglI-SwaI sites of the vector, was co-transformed with a DNA

cassette containing the sequence to be inserted in the DNA flanked by 50 bp homology to the target region as described (Laughery *et al.*, 2015). Positive transformants were selected in SC solid medium without uracil to maintain the plasmid and checked by sequencing. Only the cells that integrate the cassette will escape to Cas9 digestion and therefore will survive. Loss of the pML104 plasmid in the positive cells was done by selection in FOA containing medium.

6.4.3. Genotoxic damage sensitivity assay

Mid-log cultures were grown in YPAD or SC medium. 7 μl drops of 10-fold serial dilutions in sterile water were plated on solid YPAD, SC or SGal medium containing the drugs at the concentrations indicated in the figures. For UV irradiation, drops were dried before irradiated them. Plates were incubated during 2-4 days at the indicated temperature (UV-irradiated plates in the dark).

6.4.4. Recombination assays

Recombination frequencies were obtained by fluctuation test as the median value of six independent colonies as previously described (Gómez-González *et al.*, 2011). The average of at least three independent transformants is plotted. Recombinants were obtained by platting appropriate dilutions in applicable selective medium. To calculate the total number of cells, they were plated in the same media as the original transformation used.

L and LYΔNS plasmid systems. Both systems are based on the centromeric plasmid pRS314, containing the direct repeats *leu2Δ3'* and *leu2Δ5'*, with a 600 bp homology region. The L system contains a 35 bp fragment from YIp5 between the repeats (Prado & Aguilera, 1995). LYΔNS system contains the entire YIp5 with a 1.92 kb SphI-NsiI deletion, to remove the *URA3* gene (Prado *et al*., 1997). Recombinants are selected in plates without leucine.

GL-lacZ chromosomic system. This system is based on *leu2Δ3'* and *leu2Δ5'* truncations of the *LEU2* gene that share 600 bp homology. The sequence of the 3 kb long *lacZ* gene from *E. coli* was cloned between the direct repeats (Chávez & Aguilera, 1997).

The system is under *GAL1* promoter control and was cloned on *S. cerevisiae* chromosome III. Recombinants are selected in plates without leucine.

6.4.5. Detection of Rad52-YFP foci and protein localization

Spontaneous Rad52-YFP foci were visualized in cells transformed with plasmid pWJ1213 or pWJ1344 with a DM600B microscope (Leica) as previously described (Lisby *et al.*, 2001) with minor modifications. Individual transformants were grown to early-log phase, fixed for 10 min in 0.1 M K_iPO_4 pH 6.4 containing 2.5% formaldehyde, washed twice in 0.1 M K_iPO_4 pH 6.6 and resuspended in 0.1 M K_iPO_4 pH 7.4. More than 200 S/G2 cells were analysed for each transformant. Average values obtained from at least 3 independent transformants are plotted for each genotype.

For protein localization, YFP-tagged proteins were overexpressed for 3 h from a plasmid in mid-log phase cultures and fixed as described.

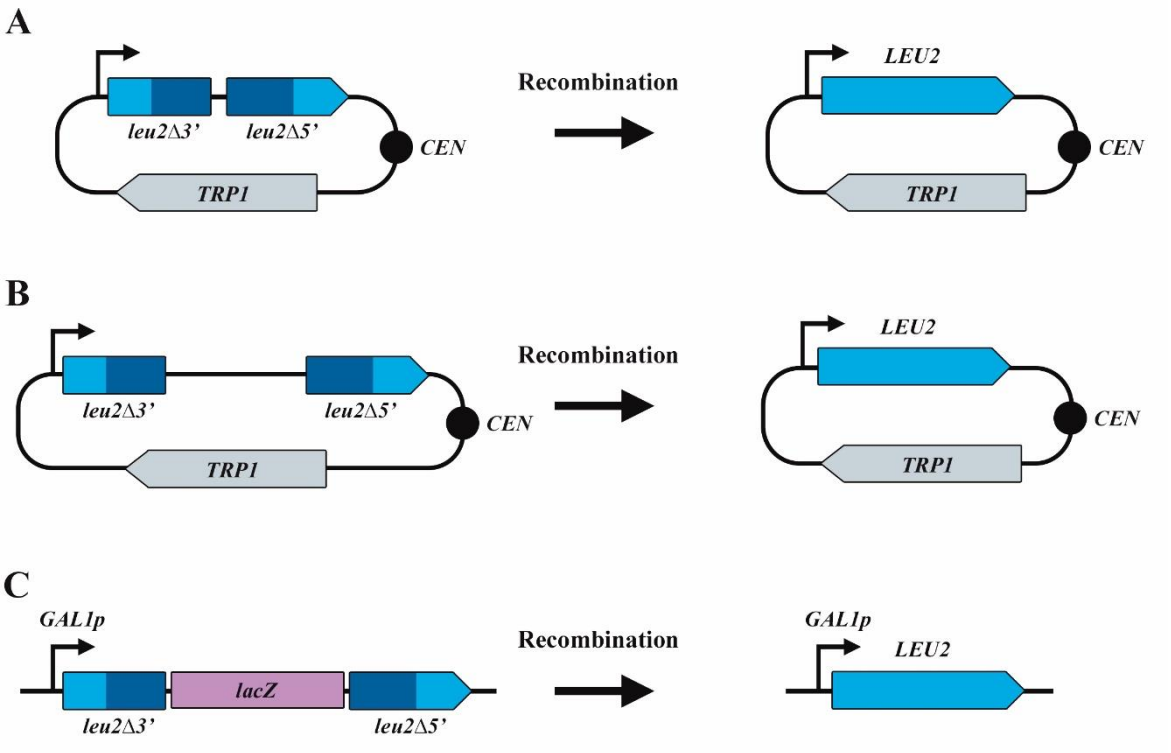


Figure M1. Recombination systems used in this thesis.
Schematic representations of the L **(A),** LYΔNS **(B)** plasmid systems or GL-lacZ chromosomic system **(C)** with the outcome of the recombination event.

6.4.6. mRNA fluorescence in situ hybridization (FISH)

Total mRNA was detected using an oligo dT 50 Cy3 fluorescent probe as described (Trcek *et al.*, 2012) with some minor changes. Briefly, early-log phase cultures were fixed with 4% formaldehyde. Cells were washed tree times with ice cold 1x buffer B (1.2 M sorbitol, 100 mM K_iPO_4 pH 7.5). Cell wall was removed by incubating in spheroplasting buffer (1x buffer B with 0.002 v/v β-mercaptoethanol, 20 mM Ribonucleoside vanadyl complex (RVC, New England Biolabs)) with 2mg/ml 20T Zymolyase. Cells were collected and washed with 1x buffer B and adhered to slides with 0.1% poly-L-Lysine incubating for 30 min at 4ºC. Slides were washed with ice cold buffer B and stored in 70% ethanol at -20ºC until used. For the hybridization, samples were rehydrated with 2x SSC and washed with hybridization solution (2x SSC, 20% formamide) for 15 min at room temperature and then incubated overnight with the hybridization mix (2x SSC, 20% formamide, 1 mg BSA ultrapure, 10 mM RVC, 10 μg herring sperm DNA, 10 μg *E. coli* tRNA and 200 ng oligo dT 50 Cy3) at 30ºC in a sealed humid chamber protected from the light. Slides were washed twice in hybridization solution at 37ºC, once with 0.1% Triton X-100 in 2x SSC at room temperature and twice with 1x SSC. Samples were stained with 0.5 μg/ml DAPI, mounted with VectaShield (Vector laboratories) and visualized with a DM600B microscope (Leica).

6.4.7. Cell cycle synchronization and FACS analysis

For cell cycle synchronization, overnight cultures were diluted to an optical density of 0.2 and grown until mid-log at 30ºC in rich (YPAD) or synthetic medium. Cells were synchronized in G1 adding 0.125 μg/ml of α-factor for *bar1*Δ mutants. After 2 h, cells were induced adding 2% galactose to the medium.

For fluorescence-activated cell sorting (FACS), cells were processed as described (Moriel-Carretero *et al.*, 2011). Briefly, 1 ml of the culture was centrifuged and washed with 1x PBS, resuspended in 1 ml of 70% ethanol and stored at 4ºC. To process the samples, cells were washed with 1x PBS, resuspended in 100 μl 1x PBS with 1mg/ml of RNase A and incubated overnight at 37ºC. Then, cells were washed with 1x PBS and resuspended in 1 ml of 5 μl/ml propidium iodide in 1x PBS, incubated in the dark for at least 30 min, sonicated 5 s at 10% amplitude and scored in a FACScalibur (Becton Dickinson, CA).

6.4.8. Chromosome spreads immunofluorescence

We followed the protocol previously described (Chan *et al.*, 2014) with some minor modifications. Briefly, 5 ml mid-log cultures were collected and washed in ice-cold solution #1 (1.2 M sorbitol, 0.1 M potassium phosphate and 0.5x magnesium chloride pH 7). Cell wall was digested in solution #1 with 10 mM DTT and 185 ng/ml of Zymoliase 20T during 20 minutes at 37ºC and then stopped by addition of 1.5 ml of solution #2 (0,1 M MES, 1M sorbitol, 1mM EDTA, 0.5 mM magnesium chloride at pH 6.4). Spheroplasts were centrifuged carefully 8 min at low speed, resuspended in solution #2, lysed with 1% Lipsol and fixed on slides with fixative solution (4% paraformaldehyde, 3.4% sucrose). Nuclei spreading was done with a glass rod and the slides were dried overnight at room temperature.

The slides were washed in 1x PBS and then blocked with blocking buffer (5% BSA, 0.2% milk in 1x PBS) during 10 minutes at room temperature in humid chambers. For the immunostaining, slides were incubated with blocking buffer containing S9.6 primary antibody (1mg/ml) in a humid chamber during 1 hour at 23ºC and washed with 1X PBS for 30 minutes. For the secondary antibody, slides were incubated 1 hour at 23ºC in the dark with Cy3 conjugated goat anti-mouse antibody (Jackson laboratories, #175-165-150) diluted 1:1000 in blocking buffer and washed with 1x PBS. Finally, slides were mounted with 50 µl of Vectashield (Vector laboratories, CA) containing 1x DAPI and sealed. More than 100 nuclei were visualized and counted to obtain the percentage of nuclei with detectable RNA:DNA hybrids.

6.5. DNA analysis

6.5.1. Southern blot

Yeast genomic DNA was digested, separate in agarose gel in native and alkaline conditions as described (Cortés-Ledesma & Aguilera, 2006) and transferred to Hybond-XL nitrocellulose membranes (GE Healthcare), with were hybridized with ^{32}P-labelled DNA probes, listed in Table M4.

6.5.2. Polymerase chain reaction (PCR)

6.5.2.1. Non-quantitative PCR

DNA amplification with temperature-stable polymerases for cloning, probe generation or strain verification were performed following standard protocols with the polymerases described in Materials and Methods 6.2.3.

6.5.2.2. Real-time quantitative PCR (qPCR)

Real-time qPCRs were performed using the iTaq^{TM} Universal SYBR® Green Supermix (Biorad). Reactions were set with 6 µl H_2O, 2 µl primer mix (0,1 mM each), 2 µl template and 10 µl SYBR® Green Supermix. Runs were always performed using the following program: 1 cycle (10 min 95ºC), 40 cycles (15 s 95ºC and 1 min 65ºC) with a final dissociation stage (15 s 95ºC, 1 min 65ºC, 15 s 95ºC and 15 s 60ºC). Samples were run in 7500 Fast Real-time PCR system (Applied Biosystem). Results were analyzed with 7500 System Software V2.0.6. A calibration curve with five 10-fold serial dilutions of a standard DNA sample was calculated for absolute quantification.

6.6. RNA analysis: Northern blot

Yeast cultures were grown in appropriate medium as indicate for each experiment. RNA was extracted from mid-log cultures using acid phenol method (Köhrer & Domdey, 1991). Northern blot was performed following standard procedures. Total RNAs were separated by agarose or acrylamide gel electrophoresis and transferred to Hybond-N nitrocellulose membranes (GE Healthcare). ^{32}P-labelled DNA probes were used. Radioactive signals were adquired using a FLA-5100 Imager Fluorescence Analyzer (Fujifilm) and quantified with the MultiGauge 2.0 analysis software (Science Lab). Signals were normalized to the SCR1 transcripts that are very stable and are transcribed by RNAPIII. Signal was plotted as arbitrarily units (a.u.).

6.7. Chromatin immunoprecipitation (ChIP)

Asynchronous or G1-synchronized mid-log cultures grown in synthetic medium at 30ºC were used. Samples were processed as described (Hecht *et al.*, 1999) with some

modifications. 50 ml cultures were cross-linked in 1% formaldehyde shacking gently for 15 min at room temperature. Reaction was stopped adding glycine to a final concentration of 125 mM and incubating for 5 min. Cells were washed twice with cold 1x PBS and stored at -80ºC. For cell extract preparation, the pellet was resuspended in 500 µl of lysis buffer (50 mM HEPES-KOH pH 7.5, 150 mM NaCl, 1 mM EDTA pH 8, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS) supplemented with protease inhibitors (1x Complete protease inhibitor cocktail (Roche) and 1 mM PMSF). Next, 1 volume of glass beads were added and cells were broken in an orbital shaker (Vibrax VXR basic, IKA) for 45 min. Samples were separated from the beads and centrifuged for 15 min to eliminate soluble proteins. Precipitate was resuspended in 300 µl of lysis buffer supplemented with protease inhibitors and sonicated using Bioruptor (Diagenode), alternating 1 min high intensity and 20 s rest pulses for 15 min. Samples were centrifuged for 15 min to eliminate cell debris. 300 µl of lysis buffer supplemented with protease inhibitors was added to each chromatin sample. 50 µl was used as a control of total DNA (input) and 100-300 µl was processed for immunoprecipitation.

The immunoprecipitation was performed overnight at 4ºC using Dynabeads Protein A or G (Invitrogen) previously incubated with the antibody at the specified concentration for 3-4 h at 4ºC rotating at low speed. Samples were washed with lysis buffer with 275 mM NaCl, lysis buffer with 500 mM NaCl, buffer III (10 mM Tris-HCl pH 8, 1 mM EDTA pH 8, 250 mM LiCl, 0.5% IGEPAL, 0.5% SDS, 0.5% sodium deoxycholate) and 1x TE. Chromatin was eluted in 100 µl elution buffer (50 mM Tris-HCl pH 7.4, 10 mM EDTA pH 8, 1% SDS) at 65ºC for 10 min. Samples were treated with 6 µl of 50 mg/ml pronase for 1.5 h at 42ºC to remove proteins and descrosslinked for 5 h at 65ºC. Macherey-Nagel NucleoSpin® Gel and PCR Clean-up purification kit was used to clean the DNA that was eluted in 100 µl of bidistilled water, employing 2 µl per qPCR reaction.

6.8. DNA:RNA hybrid immunoprecipitation (DRIP)

DNA:RNA immunoprecipitation was performed as previously described (Garcia-Rubio *et al.*, 2018). Briefly, cells were resuspended in spheroplasting buffer (1 M sorbitol, 2 mM Tris-HCl pH 8, 100 mM EDTA pH 8, 0.1% v/v β-mercapto-ethanol, 0.2% zymoliase 20T). DNA was treated with RNase A (1 h at 37ºC) and proteinase K (1 h at 50ºC) in G2

buffer (0.8 mM Guanidine HCl, 30 mM Tris-HCl pH 8, 30 mM EDTA pH 8, 5% Tween-20, 0.5% Triton X-100) and carefully extracted with chloroform:isoamylalcohol (24:1) followed by isopropanol precipitation. DNA was spooled on a glass rod, washed with 70% EtOH, gently resuspended in 1x TE and enzymatically digested with *Hind*III, *Eco*RI, *BsrG*I, *Xba*I and *Ssp*I. Samples were split and treated with *E. coli* RNase H (NEB) or mock treated.

DNA:RNA hybrid immunoprecipitation was performed by overnight incubation with Protein A Dynabeads (Invitrogen) coated with S9.6 antibody at 4ºC using binding buffer (100 mM $NaPO_4$ pH 7, 1.4 M NaCl, 0.5% Triton X-100). DNA was eluted with elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA pH 8, 0.5% SDS), treated with proteinase K and purified with Macherey-Nagel NucleoSpin® Gel and PCR Clean-up purification kit. Samples were eluted in 100 µl of bidistilled water, using 2 µl per qPCR reaction. S9.6 signal was determined by dividing the immunoprecipitated signal to the input of each sample.

6.9. Primers and probes

Primers used in this thesis for non-quantitative, quantitative PCR probe generation and cloning are described in Table M5. Real-time qPCR primers were designed using Primer Express 3.0 software (Applied Biosystems).

6.10. Protein extraction and immunodetection: Western blot

6.10.1. Protein extraction

10 ml of mid-log yeast culture was harvested and keep in ice. Proteins were extracted by adding 200 µl of ice-cold 10% TCA and 200 µl of glass beads and then vortexing 3 min four times with a pause between vortex keeping the samples in ice. Supernatant was recovered and the beads were washed twice with 200 µl cold 10% TCA. Samples were centrifuged 10 min at 3000 rpm and supernatant discarded. The pellets were resuspended using 100 µl of 2x Loading Buffer (62.5 mM Tris-HCl pH 6.8, 25% glycerol, 2% SDS, 0.01% Bromophenol Blue, 5% β-mercaptoethanol), 50 µl of distilled water and 50 µl of 1 M Tris (not-adjusted pH). Before gel loading, samples were boiled 5 min and centrifuged 10 min at 3000 rpm.

6.10.2. Sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated by electrophoresis in 29:1 acrylamide:bis-acrylamide 8% gels at 100V in running buffer (25 mM Tris base pH 8.3, 194 mM glycine, 0.1% SDS). Page Ruler plus (Fermentas, CA) was used as a marker.

6.10.3. Western blot analysis

Proteins were transferred to a nitrocellulose membrane (Hybond-ECL, GE Healthcare) using Trans-Blot system from Biorad for 1 h at 300 mA in Transfer buffer (6 g/L Tris base, 28.8 g/L glycine, 0.5% SDS and 20% methanol) at 4ºC. Membranes were stained with Ponceau S (0.1% w/v Ponceau -SIGMA- in 5% acetic acid) to check the protein loading and transference. Then, membranes were blocked with 1x TBS with 0.05% Tween-20 and 5% milk for at least 1 h at room temperature. Primary antibodies were incubated over night at 4ºC at the indicated concentrations (Table M1) in 1x TBS, 0.05% Tween-20, 3% BSA and then washed three times in 1x TBS, 0.05% Tween-20. Secondary antibodies conjugated with horseradish peroxidase were added at the indicated concentrations (Table M2) in 1x TBS, 0.05% Tween-20 with 5% milk for 1 h at room temperature and washed again. Finally, SuperSignal West Pico Plus (Thermo) was used for chemiluminescence detection.

6.11. Statistical analyses

Statistical test (Student's *t*-test) were calculated using GraphPad Prism software. In general, a *p*-value <0,05 was considered as statistically significant.

Table M5. Primers used in this thesis.

Primer	Sequence 5' to 3'	Use
DIS3-EcoRI fw	CATCATACAGGCGAATTCAACATGTCAG	Cloning
DIS3-SacI fw	GATACACCCCCAGAGCTCCAAGGTTAGA CTACTACAGC	Cloning
DIS3-SphI rv	CTAAATAGTGCATGCCACTCTACAAGAGA TATCACG	Cloning
DIS3-XhoI rv	CGTTTTTATATCTCGAGACTGAAGCATC	Cloning

Primer	Sequence 5' to 3'	Use
FCP1-XhoI fw	CATTGACCGCGCTCGAGTAAACACAAG	Cloning
FCP1-XhoI rv	AGATACGGCACTCGAGCTGCTAATC	Cloning
HRP1-EcoRI fw	GTCTGAGAAAATAGAGAATTCGTTAAATAAG	Cloning
HRP1-EcoRI rv	GTTGTTGAATTATACAAGAATTCTTTTCTCTAG	Cloning
LYS2p NATnt2-GALp fw	GGCATCGCACAGTTTTAGCGAGGAAAACTCTTCAATAGTTTTGCCAGCGGGACATGGAGGCCCAGAATAC	Cloning
LYS2p NATnt2-GALp rv	GAAAGAGTTGGATTATCCAACTTCTCTATCCAGACCTTTTCGTTAGTCATCTCCTTGACGTTAAAGTATAGAGG	Cloning
LYS2+FRT SN fw	CAACAATTAATGTGTTTGTTACCGGTGTCACAGGATTTCTGGGCTCCGAAGTTCCTATACTTTCTAGAGAATAGG	Cloning
LYS2+FRT SN rv	CTTGGCCCTGACGTGGGCAAACACTTTGAAACTGTAGTTCTTTGGAGAACGAAGTTCCTATTCGGAAGTTCC	Cloning
LYS2+FRT AS fw	GAAAAACAACAATTAATGTGTTTGTTACCGGTGTCACAGGATTTCTGGGCTCCGAAGTTCCTATTCGGAAGTTCC	Cloning
LYS2+FRT AS rv	CTTGGCCCTGACGTGGGCAAACACTTTGAAACTGTAGTTCTTTGGAGAACAAGTTCCTATACTTTCTAGAGAATAGG	Cloning
LYS2+Sm350 fw	GAAAAACAACAATTAATGTGTTTGTTACCGGTGTCACAGGATTTCTGGGCTCATTCCTGCAGCCCTGAGCTG	Cloning
LYS2+Sm350 rv	CTTGGCCCTGACGTGGGCAAACACTTTGAAACTGTAGTTCTTTGGAGAACCTGCAGCCCGGGGGAT	Cloning
NAB2-EcoRI fw	CATAAGGAAGTGGAATTCCATCAGAAATG	Cloning
NAB2-EcoRI rv	GCTTTGAATAGGTGAATTCCATCAAAAGG	Cloning
NHP6B-XhoI fw	CATAGCGCACGACTCGAGTACTAAC	Cloning
NHP6B-XhoI rv	GGTAGAGAACCTCGAGTTGTTAGATC	Cloning
NPL3-XhoI fw	CGCTAAAACCTCGAGGATAATGTCTG	Cloning
NPL3-XhoI rv	CTCAACTATCTCGAGGGCTTACCTG	Cloning

Primer	Sequence 5' to 3'	Use
NRD1 HA fw	GCTCAATTGAATTCTTTGATGAATATGCTTAACCAACAGCAGCAGCAACAACAACAAAGCCGGATCCCCGGGTTAATTAA	Cloning
NRD1 HA rv	GAGGTAGATTAGTTTTATGTACTATGAGCAAATAAAGGGTGGAGTAAAGATCTTAATGAATTCGAGCTCGTTTAAAC	Cloning
PRP2-EcoRI fw	GCGTGTATAGGAATTCATGTCAAG	Cloning
PRP2-EcoRI rv	CTGCGTTTCTAGAATTCCACATACAC	Cloning
RIE1-EcoRI rv	TTATGGACCTGAATTCATCTAGTAGTCC	Cloning
RIE1-KpnI fw	CCGCATCAGAGGTACCGAGGATG	Cloning
RIE1-PH rv	CTCCAGTGAAAAGTTCTTCTCCTTTACTCATCGCTGCGGCAGCGTAGTCCATAGAATAATCACCAC	Cloning
RNH1-MX6 fw	TAAAATTAGTTAAAGTGTCACTCCTTGCTTATCGAAGGAACTATCGATTCCTAATTATGCGGATCCCCGGGTTAATTAAG	Cloning
RNH1-MX6 rv	ATATATTTCTATTACAGGTACAACAGGTCCAGTAAGAAGCCAAGCAAAAAACAGCATTATTGGATGGCGGCGTTAGTATC	Cloning
RNH201-HYG F	ATGAGAGTGTCGAAAAACCTTGAAAACAACTACTGCACACCAAATTGATACGATTAA	Cloning
RNH201-HYG R	TGAAGTTATGACATATGTAGTATTACATGAAGATATATAGTATGTGCAAACTGGAGG	Cloning
SNF5-EcoRI fw	GTAAAGAACTACACGAATTCAACAATGAA	Cloning
SNF5-EcoRI rv	CACGATGATAATACGAATTCTTCCACGG	Cloning
SWT1-XhoI fw	CTAACATTAGTGGCTCGAGCTGCTTAC	Cloning
SWT1-XhoI rv	CATATCGCAAGCTCGAGAGTTGTGTGAG	Cloning
YFP fw	ATGAGTAAAGGAGAAGAACTTTTC	Cloning
YFP-EcoRI rv	GAGTAACTAGAGGAATTCGGAGTAAT	Cloning
gRNA LYS2 A	GATCTACATCCTTGCAGATTTGTTTTAGAGCTAG	gRNA cloning
gRNA LYS2 B	CATGCTCTAAAACAACAAATCTGCAAGGATGTA	gRNA cloning
gRNA LYS2-3' A	GATCGCCAATTCATTTTCTTTGGGGTTTTAGAGCTAGGCTAG	gRNA cloning

Primer	Sequence 5' to 3'	Use
gRNA LYS2-3' B	CTAGCTCTAAAACCCCAAAGAAAATGAATTGGC	gRNA cloning
5.8S fw	CTTTCAACAACGGATCTCTTGG	Northern probe
5.8S rv	GACGCTCAAACAGGCATGC	Northern probe
HSP104 S3 fw	GTTAGGCAACATTTCAGACCAG	Northern probe
HSP104 S3 rv	CATACTGTCCTCATTATCGTCATC	Northern probe
NAB2 sonda rv	CTGTCTGCATTGCATTCTG	Northern probe
NPL3 sonda rv	CTAACAAACAATCTGGTGTTCG	Northern probe
SHE2 sonda rv	GATTGTAGGTATCGTCAATGGC	Northern probe
LYS2 S sonda fw	GCTACATATTCGTTACAGCTACCTCAGC	Southern probe
LYS2 S sonda rv	GATGGATCGCTTAGCGCAGCAGTC	Southern probe
Sonda FRT fw	GTGGATCATCTCAAGGTGAGGTCG	Southern probe
Sonda FRT rv	GCCAAATCCATCCACTTCTCATCTG	Southern probe
FRT compr	CTATTCGGAAGTTCCTATTCTCTAGAAAG	PCR
NAT fw	AGGTCACCAACGTCAACGCA	PCR
NHP6B-IN rv	CGTGTAGCATTGTACAATTCC	PCR
SNF5-IN rv	GATGTAAGTTGTGTGGTGC	PCR
Sm350 compr	ATTCCTGCAGCCCTGAGCTG	PCR
18S_qPCR F	GGAATCGAACCCTTATTCCC	qPCR
18S_qPCR R	TCAACTTTCGATGGTAGGAT	qPCR
ASH1-E1 fw	CATTGGTGTAAGGATACAAACTATC	qPCR
ASH1-E1 rv	TTTTGATTATTAGTTAAGTTGGGTATAC	qPCR
ASH1-E3 fw	AGATCCCACAAAGGGTGAAATAAACA	qPCR
ASH1-E3 rv	ATTACAAAATAAGCAACGGTACCCTTCAAT	qPCR
GAL1 PROM up	CACTGCTCCGAACAATAAAGATTC	qPCR
GAL1 PROM lw	GGCCAGGTTACTGCCAATTTT	qPCR
GCN4 3' fw	TTGTGCCCGAATCCAGTGA	qPCR
GCN4 3' rv	TGGCGGCTTCAGTGTTTCTA	qPCR
GCN4 M	CGATGTTTCATTGGCTGATAAGG	qPCR
GCN4 M	CCAGATTGGATGGTACCAGAGAA	qPCR
HXT1 up	AGCTGGCAGAATCGACGAA	qPCR
HXT1 down	GGTCAGGTGGGCATTTGTTAA	qPCR
LYS2 1K fw	GTGTGGATTTGATGGTATGTGTGA	qPCR
LYS2 1K rv	GCAGGGTCGATAACTGAAAAGG	qPCR
LYS2 2K fw	CTGGTTAGGTCCAAGAGATAGATTGT	qPCR

Primer	Sequence 5' to 3'	Use
LYS2 2K rv	CAGTCACCGTTTGGTAGATAACGA	qPCR
LYS2 3K fw	CAGGGCCAAGGATGAAGAAG	qPCR
LYS2 3K rv	GTACCATAGGTGATACCTGCCTTT	qPCR
LYS2 4K fw	GCTCCGGAACTAGACGATAGGA	qPCR
LYS2 4K rv	CTGTCCATGCGGTGTCTTTCT	qPCR
LYS2 PROM fw	CGGTTTTTCGCGTGTGACT	qPCR
LYS2 PROM rv	CATTTGGGCGATGTTCATGTTC	qPCR
SCR1 .426 fw	GATCGCTTCGGCGGTTTAA	qPCR
SCR1 .45 up	TGGCCGAGGAACAAATCCT	qPCR and Northern probe
SCR1 .483 rv	GGCCACAATGTGCGAGTAAAT	qPCR and Northern probe
SCR1 .99 dw	CCCAAAGGGCGTGCAAT	qPCR
SNR47reg3 up	CGCGTCGGGATAACAAAGCGTAC	qPCR
SNR47reg3 dw	CCCTGTTATCCGCCTTTCTTCTTGG	qPCR
SPF1 3' F	CCCGTGGTAAACCTTTAGAAAAAC	qPCR
SPF1 3' R	ATATGAACGGCAAATTGAGACAAA	qPCR
V1 (Chr. V Interg.)	TGTTCCTTTAAGAGGTGATGGTGAT	qPCR
V2 (Chr. V Interg.)	GTGCGCAGTACTTGTGAAAACC	qPCR

7. REFERENCES

Aguilera A, Gaillard H. (2014). Transcription and recombination: when RNA meets DNA. *Cold Spring Harb Perspect Biol*. 6(8). pii: a016543.

Aguilera A, García-Muse T. (2012). R loops: from transcription byproducts to threats to genome stability. *Mol Cell.* 46(2): 115-124.

Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ. (2009). TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell.* 34(3): 387-393.

Allmang C, Mitchell P, Petfalski E, Tollervey D. (2000). Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res* 28(8): 1684-91.

Anderson JS, Parker RP. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. *EMBO J*. 17(5): 1497-1506.

Anderson JT, Wilson SM, Datar KV, Swanson MS. (1993). NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. *Mol Cell Biol*. 13(5): 2730-2741.

Arigo JT, Eyler DE, Carroll KL, Corden JL. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol Cell*. 23(6): 841-851.

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(25): 9169-9174.

Awrey DE, Weilbaecher RG, Hemming SA, Orlicky SM, Kane CM, Edwards AM. (1997). Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J Biol Chem*. 272(23): 14747-51474.

Belotserkovskii BP, Tornaletti S, D'Souza AD, Hanawalt PC. (2018). R-loop generation during transcription: Formation, processing and cellular outcomes. *DNA Repair (Amst).* 71: 69-81.

Belotserkovskii BP, Soo Shin JH, Hanawalt PC. (2017). Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Res.* 45(11): 6589-6599.

Bentin T, Cherny D, Larsen HJ, Nielsen PE. (2005). Transcription arrest caused by long nascent RNA chains. *Biochim Biophys Acta*. 1727(2): 97-105.

Björk P1, Wieslander L. (2017). Integration of mRNP formation and export. *Cell Mol Life Sci*. 74(16): 2875-2897.

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science* 277(5331),1453-1462.

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(2), 345-346.

Böhl F, Kruse C, Frank A, Ferring D, Jansen RP. (2000). She2p, a novel RNA-binding protein tethers ASH1 mRNA to the Myo4p myosin motor via She3p. *EMBO J.* 19(20): 5514-5524.

Boque-Sastre R, Soler M, Oliveira-Mateos C, Portela A, Moutinho C, Sayols S, Villanueva A, Esteller M, Guil S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc Natl Acad Sci U S A.* 112(18): 5785-5790.

Bossie MA, DeHoratius C, Barcelo G, Silver P. (1992). A mutant nuclear protein with similarity to RNA binding proteins interferes with nuclear import in yeast. *Mol Biol Cell*. 3(8):875-893.

Buchan JR, Muhlrad D, Parker R. (2008). P bodies promote stress granule assembly in *Saccharomyces cerevisiae*. *J Cell Biol*. 183(3): 441-455.

Bühler M, Haas W, Gygi SP, Moazed D. (2007). RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell*. 129(4): 707-721.

Bunch H, Lawney BP, Lin YF, Asaithamby A, Murshid A, Wang YE, Chen BP, Calderwood SK. (2015). Transcriptional elongation requires DNA break-induced signalling. *Nat Commun.* 6: 10191.

Candelli T, Challal D, Briand JB, Boulay J, Porrua O, Colin J, Libri D. (2018). High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast. *EMBO J.* 37(4). pii: e97490.

Castellano-Pozo M, Santos-Pereira JM, Rondón AG, Barroso S, Andújar E, Pérez-Alegre M, García-Muse T, Aguilera A. (2013). R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol Cell*. 52(4): 583-590.

Cerritelli SM, Crouch RJ. (2009). Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* 276(6): 1494-1505.

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 Genet.* 10(4): e1004288.

Chartrand P, Meng XH, Singer RH, Long RM. (1999). Structural elements required for the localization of ASH1 mRNA and of a green fluorescent protein reporter particle in vivo. *Curr Biol*. 9(6): 333-336.

Chávez S, García-Rubio M, Prado F, Aguilera A. (2001). Hpr1 is preferentially required for transcription of either long or G+C-rich DNA sequences in *Saccharomyces cerevisiae*. *Mol Cell Biol.* 21(20): 7054-7064.

Chávez S, Beilharz T, Rondón 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(21): 5824-5834.

Chávez S, Aguilera A. (1997). The yeast *HPR1* gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes Dev* 11(24), 3459-3470.

Cheung AC, Cramer P. (2012). A movie of RNA polymerase II transcription. *Cell*. 149(7): 1431-1437.

Chlebowski A, Lubas M, Jensen TH, Dziembowski A. (2013). RNA decay machines: the exosome. *Biochim Biophys Acta*. 1829(6-7): 552-560.

Chou DM, Adamson B, Dephoure NE, Tan X, Nottke AC, Hurov KE, Gygi SP, Colaiácovo MP, Elledge SJ. (2010). A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. Proc *Natl Acad Sci U S A*. 107(43): 18475-18480.

Colin J, Candelli T, Porrua O, Boulay J, Zhu C, Lacroute F, Steinmetz LM, Libri D. (2014). Roadblock termination by reb1p restricts cryptic and readthrough transcription. *Mol Cell*. 56(5): 667-680.

Cortés-Ledesma F, Aguilera A. (2006). Double-strand breaks arising by replication through a nick are repaired by cohesin-dependent sister-chromatid exchange. *EMBO Rep.* 7(9), 919–926.

Costantino L, Koshland D. (2018). Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. *Mol Cell.* 71(4): 487-497.

Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, Damsma GE, Dengl S, Geiger SR, Jasiak AJ, Jawhari A, Jennebach S, Kamenski T, Kettenberger H, Kuhn CD, Lehmann E, Leike K, Sydow JF, Vannini A (2008). Structure of eukaryotic RNA polymerases. *Annu Rev Biophys*. 37: 337-352.

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR, Kornberg RD. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*. 288(5466): 640-649.

Creamer TJ, Darby MM, Jamonnak N, Schaughency P, Hao H, Wheelan SJ, Corden JL. (2011). Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet*. (10): e1002329.

Crossley MP, Bocek M, Cimprich KA. (2019). R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell*. 73(3): 398-411.

Davidson L, Francis L, Cordiner RA, Eaton JD, Estell C, Macias S, Cáceres JF, West S. (2019). Rapid Depletion of DIS3, EXOSC10, or XRN2 Reveals the Immediate Impact of

Exoribonucleolysis on Nuclear RNA Metabolism and Transcriptional Control. *Cell Rep.* 26(10), 2779-2791.

Davies CJ, Trgovcich J, Hutchison CA. (1990). Homologue of TFIIS in yeast. *Nature* 345(6273): 298.

Dellino GI, Palluzzi F, Chiariello AM, Piccioni R, Bianco S, Furia L, De Conti G, Bouwman BAM, Melloni G, Guido D, Giacò L, Luzi L, Cittaro D, Faretta M, Nicodemi M, Crosetto N, Pelicci PG. (2019). Release of paused RNA polymerase II at specific loci favors DNA double-strand-break formation and promotes cancer translocations. *Nat Genet.* 51(6): 1011-1023.

Desai SD, Zhang H, Rodriguez-Bauman A, Yang JM, Wu X, Gounder MK, Rubin EH, Liu LF. (2003). Transcription-dependent degradation of topoisomerase I-DNA covalent complexes. *Mol Cell Biol*. 23(7): 2341-2350.

Domínguez-Sánchez MS, Barroso S, Gómez-González B, Luna R, Aguilera A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet*. 7(12): e1002386.

Du TG, Jellbauer S, Müller M, Schmid M, Niessing D, Jansen RP. (2008). Nuclear transit of the RNA-binding protein She2 is required for translational control of localized ASH1 mRNA. *EMBO Rep*. 9(8): 781-787.

Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev*. 18(13): 1618-1629.

Dziembowski A, Lorentzen E, Conti E, Séraphin B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat Struct Mol Biol.* 14(1): 15-22.

Eberle AB, Jordán-Pla A, Gañez-Zapater A, Hessle V, Silberberg G, von Euler A, Silverstein RA, Visa N. (2015). An Interaction between RRP6 and SU(VAR)3-9 Targets RRP6 to Heterochromatin and Contributes to Heterochromatin Maintenance in *Drosophila melanogaster*. *PLoS Genet* 11(9): e1005523.

Edelmann FT, Schlundt A, Heym RG, Jenner A, Niedner-Boblenz A, Syed MI, Paillart JC, Stehle R, Janowski R, Sattler M, Jansen RP, Niessing D. (2017). Molecular architecture and dynamics of ASH1 mRNA recognition by its mRNA-transport complex. *Nat Struct Mol Biol.* 24(2): 152-161.

Estruch F, Hodge C, Gómez-Navarro N, Peiró-Chova L, Heath CV, Cole CN (2012). Insights into mRNP biogenesis provided by new genetic interactions among export and transcription factors. *BMC Genet* 13, 80.

Fei J, Chen J. (2012). KIAA1530 protein is recruited by Cockayne syndrome complementation group protein A (CSA) to participate in transcription-coupled repair (TCR). *J Biol Chem.* 287(42): 35118-35126.

Feng Q, Düring L, de Mayolo AA, Lettier G, Lisby M, Erdeniz N, Mortensen UH, Rothstein R. (2007). Rad52 and Rad59 exhibit both overlapping and distinct functions. *DNA Repair* 6(1), 27-37.

Feroli F, Carignani G, Pavanello A, Guerreiro P, Azevedo D, Rodrigues-Pousada C, Melchioretto P, Panzeri L, Agostoni Carbone ML. (1997). Analysis of a 17.9 kb region from Saccharomyces cerevisiae chromosome VII reveals the presence of eight open reading frames, including BRF1 (TFIIIB70) and GCN5 genes. *Yeast*. 13(4): 373-377.

Gaillard H, Aguilera A. (2014). Cleavage factor I links transcription termination to DNA damage response and genome integrity maintenance in Saccharomyces cerevisiae. *PLoS Genet*. 10(3): e1004203.

Gaillard H, Wellinger RE, Aguilera A. (2007). A new connection of mRNP biogenesis and export with transcription-coupled repair. *Nucleic Acids Res*. 35(12): 3893-906.

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(26), 24225-24232.

Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X. (2011). R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev*. 25(19): 2041-2056.

García-Benítez F, Gaillard H, Aguilera A. (2017). Physical proximity of chromatin to nuclear pores prevents harmful R loop accumulation contributing to maintain genome stability. *Proc Natl Acad Sci U S A* 114(41), 10942-10947.

García-Pichardo D, Cañas JC, García-Rubio ML, Gómez-González B, Rondón AG, Aguilera A. (2017). Histone Mutants Separate R Loop Formation from Genome Instability Induction. *Mol Cell*. 66(5): 597-609.

García-Rubio M, Barroso-Ceballos SI, Aguilera A. (2018). Detection of DNA-RNA hybrids in vivo. *Methods in Molecular Biology*, 347-361.

García-Rubio M, Aguilera P, Lafuente-Barquero J, Ruiz JF, Simon MN, Geli V, Rondón AG, Aguilera A. (2018). Yra1-bound RNA-DNA hybrids cause orientation-independent transcription-replication collisions and telomere instability. *Genes Dev*. 32(13-14), 965-977.

García-Rubio M, Chávez 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(2): 123-132.

Gavaldá S, Santos-Pereira JM, García-Rubio ML, Luna R, Aguilera A. (2016). Excess of Yra1 RNA-binding factor causes transcription-dependent genome instability, replication impairment and telomere shortening. *PLoS Genetics* 12(4), e1005966.

Gavaldá S, Gallardo M, Luna R, Aguilera A. (2013). R-loop mediated transcription-associated recombination in trf4Δ mutants reveals new links between RNA surveillance and genome integrity. *PLoS One.* 8(6): e65541.

Gietz RD, Schiestl RH. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* 2(1), 31-34.

Ginno PA, Lott PL, Christensen HC, Korf I, Chédin F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell*. 45(6): 814-825.

Gómez-González B, Aguilera A. (2019). Transcription-mediated replication hindrance: a major driver of genome instability. *Genes Dev*. 33(15-16): 1008-1026.

Gómez-González B, García-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marín A, Foiani M, Aguilera A. (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J.* 30(15): 3106-19.

Gómez-González B, Ruiz JF, Aguilera A. (2011). Genetic and molecular analysis of mitotic recombination in *Saccharomyces cerevisiae*. *Methods Mol Biol* 745, 151-172.

Gómez-González B, Aguilera A. (2009). R-loops do not accumulate in transcription-defective hpr1-101 mutants: implications for the functional role of THO/TREX. *Nucleic Acids Res*. 37(13): 4315-4321.

González-Aguilera C, Tous C, Babiano R, de la Cruz J, Luna R, Aguilera A. (2011). Nab2 functions in the metabolism of RNA driven by polymerases II and III. *Molecular Biology of the Cell*, 22(15), 2729–2740.

Gritz L, Davies J. (1983). Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* 25(2-3), 179-188.

Hamperl S, Bocek MJ, Saldivar JC, Swigut T, Cimprich KA. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell*. 170(4): 774-786.

Hamperl S, Cimprich KA. (2016). Conflict Resolution in the Genome: How Transcription and Replication Make It Work. *Cell*. 167(6): 1455-1467.

Hamperl S, Cimprich KA. (2014). The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst)*. 19: 84-94.

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

Harreman M, Taschner M, Sigurdsson S, Anindya R, Reid J, Somesh B, Kong SE, Banks CA, Conaway RC, Conaway JW, Svejstrup JQ. (2009). Distinct ubiquitin ligases act sequentially for RNA polymerase II polyubiquitylation. *Proc Natl Acad Sci U S A*. 106(49): 20705-20710.

Hartzog GA, Wada T, Handa H, Winston F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in *Saccharomyces cerevisiae*. *Genes Dev*. 12(3): 357-369

Hecht A, Strahl-Bolsinger S, Grunstein M. (1999). Mapping DNA interaction sites of chromosomal proteins. Crosslinking studies in yeast. *Methods Mol Biol* 119, 469-479.

Heidemann M, Hintermair C, Voß K, Eick D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta.* 1829(1): 55-62.

Herrera-Moyano E, Mergui X, García-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(7): 735-748.

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

Hill JE, Myers AM, Koerner TJ, Tzagoloff A. (1986). Yeast/E. coli shuttle vectors with multiple unique restriction sites. *Yeast* 2(3), 163-167.

Hodroj D, Serhal K, Maiorano D. (2017). Ddx19 links mRNA nuclear export with progression of transcription and replication and suppresses genomic instability upon DNA damage in proliferating cells. *Nucleus*. 8(5): 489-495.

Houseley J, Kotovic K, El Hage A, Tollervey D. (2007). Trf4 targets ncRNAs from telomeric and rDNA spacer regions and functions in rDNA copy number control. *EMBO J*. 26(24): 4996-5006.

Huertas P, García-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(20): 7451-7465.

Huertas P, Aguilera A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell*. 12(3): 711-721.

Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. (2003). Global analysis of protein localization in budding yeast. *Nature*. 425(6959): 686-691.

Huibregtse JM, Yang JC, Beaudenon SL. (1997). The large subunit of RNA polymerase II is a substrate of the Rsp5 ubiquitin-protein ligase. *Proc Natl Acad Sci USA.* 94(8): 3656-3661.

Iyama T, Wilson DM 3rd. (2016). Elements That Regulate the DNA Damage Response of Proteins Defective in Cockayne Syndrome. *J Mol Biol*. 428(1): 62-78.

Janke C, Magiera MM, Rathfelder N, Taxis C, Reber S, Maekawa H, Moreno-Borchart A, Doenges G, Schwob E, Schiebel E, Knop M. (2002). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast*. 21(11): 947-962.

Jansen LE, Belo AI, Hulsker R, Brouwer J. (2002). Transcription elongation factor Spt4 mediates loss of phosphorylated RNA polymerase II transcription in response to DNA damage. *Nucleic Acids Res*. 30(16): 3532-3539.

Jimenez A, Davies J. (1980). Expression of a transposable antibiotic resistance element in Saccharomyces. *Nature* 287(5785), 869-871.

Jimeno-González S, Haaning LL, Malagon F, Jensen TH. (2010). The yeast 5'-3' exonuclease Rat1p functions during transcription elongation by RNA polymerase II. *Mol Cell*. 37(4): 580-587.

Jimeno S, Rondón 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(13): 3526-3535.

Kathe SD, Shen GP, Wallace SS. (2004). Single-stranded breaks in DNA but not oxidative DNA base damages block transcriptional elongation by RNA polymerase II in HeLa cell nuclear extracts. *J Biol Chem*. 279(18): 18511-18520.

Keskin H, Shen Y, Huang F, Patel M, Yang T, Ashley K, Mazin AV, Storici F. (2016). Transcript-RNA-templated DNA recombination and repair. *Nature*. 515(7527): 436-439.

Kessler MM, Henry MF, Shen E, Zhao J, Gross S, Silver PA, Moore CL. (1997). Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3′-end formation in yeast. *Genes & Development*, 11(19), 2545–2556.

Kitsera N, Stathis D, Lühnsdorf B, Müller H, Carell T, Epe B, Khobta A. (2011). 8-Oxo-7,8-dihydroguanine in DNA does not constitute a barrier to transcription, but is converted into transcription-blocking damage by OGG1. *Nucleic Acids Res*. 39(14): 5926-5934.

Köhrer K, Domdey H. (1991). Preparation of high molecular weight RNA. *Methods Enzymol* 194: 398-405.

Kouzine F, Wojtowicz D, Baranello L, Yamane A, Nelson S, Resch W, Kieffer-Kwon KR, Benham CJ, Casellas R, Przytycka TM, Levens D. (2017). Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Syst*. 4(3): 344-356.

Krügel H, Fiedler G, Smith C, Baumberg S. (1993). Sequence and transcriptional analysis of the nourseothricin acetyltransferase-encoding gene nat1 from *Streptomyces noursei*. *Gene* 127(1): 127-131.

Kumar A, Harrison PM, Cheung KH, Lan N, Echols N, Bertone P, Miller P, Gerstein MB, Snyder M. (2002). An integrated approach for finding overlooked genes in yeast. *Nat Biotechnol.* 20(1): 58-63.

LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell*. 121(5): 713-724.

Laughery MF, Hunter T, Brown A, Hoopes J, Ostbye T, Shumaker T, Wyrick JJ. (2015) New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*. *Yeast* 32(12): 711-720.

Lejeune D, Chen X, Ruggiero C, Berryhill S, Ding B, Li S. (2009). Yeast Elc1 plays an important role in global genomic repair but not in transcription coupled repair.

Lemos BR, Kaplan AC, Bae JE, Ferrazzoli AE, Kuo, Anand RP, Waterman DP, Haber JE. (2018). CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc Natl Acad Sci USA.* 115(9): E2040-E2047.

Li S, Smerdon MJ. (2004). Dissecting transcription-coupled and global genomic repair in the chromatin of yeast GAL1-10 genes. *J Biol Chem.* 279(14): 14418-14426.

Li X, Manley JL. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell*. 122(3): 365-738.

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

Liu B, Alberts BM. (1995). Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science*. 267(5201): 1131-1137.

Liu Q, Greimann JC, Lima CD. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell*. 127(6): 1223-1237.

Long RM, Gu W, Lorimer E, Singer RH, Chartrand P. (2000). She2p is a novel RNA-binding protein that recruits the Myo4p-She3p complex to ASH1 mRNA. *EMBO J*. 19(23): 6592-6601.

Lorenz MC, Heitman J. (1998). Regulators of pseudohyphal differentiation in Saccharomyces cerevisiae identified through multicopy suppressor analysis in ammonium permease mutant strains. *Genetics*. 150(4): 1443-1457.

Luke B, Panza A, Redon S, Iglesias N, Li Z, Lingner J. (2008). The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in *Saccharomyces cerevisiae*. *Mol Cell*. 32(4): 465-477.

Luna R, Rondón AG, Aguilera A. (2012). New clues to understand the role of THO and other functionally related factors in mRNP biogenesis. *Biochim Biophys Acta.* 1819(6): 514-520.

MacKellar AL, Greenleaf AL. (2011). Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II. *J Biol Chem.* 286(42): 36385-36395.

Maizels N, Gray LT. (2013). The G4 genome. *PLoS Genet*. 9(4): e1003468.

Manfrini N, Clerici M, Wery M, Colombo CV, Descrimes M, Morillon A, d'Adda di Fagagna F, Longhese MP. (2015). Resection is responsible for loss of transcription around a double-strand break in *Saccharomyces cerevisiae*. *Elife*. 4: e08942.

Mason PB, Struhl K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell.* 17(6): 831-840.

Matsuzaki K, Shinohara A, Shinohara M. (2008). Forkhead-associated domain of yeast Xrs2, a homolog of human Nbs1, promotes nonhomologous end joining through interaction with a ligase IV partner protein, Lif1. *Genetics*. 179(1): 213-225.

Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science*. 336(6089): 1723-1725.

Mayne LV, Lehmann AR. (1982). Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Res.* 42(4): 1473-1478.

Meinel DM, Burkert-Kautzsch C, Kieser A, O'Duibhir E, Siebert M, Mayer A, Cramer P, Söding J, Holstege FC, Sträßer K. (2013). Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II. *PLoS Genet*. 9(11): e1003914.

Milbury KL, Paul B, Lari A, Fowler C, Montpetit B, Stirling PC (2019). Exonuclease domain mutants of yeast DIS3 display genome instability. *Nucleus* 10(1): 21-32.

Mischo HE, Gómez-González B, Grzechnik P, Rondón AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell.* 41(1): 21-32.

Mitchell P, Petfalski E, Tollervey D. (1996). The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes Dev*. 10(4): 502-513.

Mott C, Symington LS. (2011). RAD51-independent inverted-repeat recombination by a strand-annealing mechanism. *DNA Repair (Amst).* 10(4): 408-415.

Moynahan ME, Jasin M. (1997). Loss of heterozygosity induced by a chromosomal double-strand break. *Proc Natl Acad Sci USA*. 94(17): 8988-8993.

Mullenders L. (2015). DNA damage mediated transcription arrest: Step back to go forward. DNA *Repair (Amst)*. 36: 28-35.

Murakami H, Goto DB, Toda T, Chen ES, Grewal SI, Martienssen RA, Yanagida M. (2007). Ribonuclease activity of Dis3 is required for mitotic progression and provides a possible link between heterochromatin and kinetochore function. *PLoS One*. 2(3): e317.

Neil AJ, Belotserkovskii BP, Hanawalt PC. (2012). Transcription blockage by bulky end termini at single-strand breaks in the DNA template: differential effects of 5' and 3' adducts. *Biochemistry*. 51(44): 8964-7890.

Mellon I, Spivak G, Hanawalt PC. (1987). Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell*. 51(2): 241-249.

Nielsen I, Bentsen IB, Lisby M, Hansen S, Mundbjerg K, Andersen AH, Bjergbaek L. (2009). A Flp-nick system to study repair of a single protein-bound nick in vivo. *Nat Methods*. 6(10): 753-757.

Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods* 6: 917-922.

Palangat M, Larson DR. (2012) Complexity of RNA polymerase II elongation dynamics. *Biochim Biophys Acta*. 1819(7): 667-672.

Pankotai T, Soutoglou E. (2013) Double strand breaks: hurdles for RNA polymerase II transcription? *Transcription*. 4(1): 34-38.

Pankotai T, Bonhomme C, Chen D, Soutoglou E. (2012). DNAPKcs-dependent arrest of RNA polymerase II transcription in the presence of DNA breaks. *Nat Struct Mol Biol*. 19(3): 276-282.

Pardo B, Gómez-González B, Aguilera A. (2009). DNA double-strand break repair: how to fix a broken relationship. *Cell Mol Life Sci*. 66(6): 1039-5106.

Paull TT, Carey M, Johnson RC. (1996). Yeast HMG proteins NHP6A/B potentiate promoter-specific transcriptional activation in vivo and assembly of preinitiation complexes in vitro. *Genes Dev*. 10(21): 2769-2781.

Pavri R. (2017). R Loops in the Regulation of Antibody Gene Diversification. *Genes (Basel)*. 8(6). pii: E154.

Poirier GG, de Murcia G, Jongstra-Bilen J, Niedergang C, Mandel P (1982). Poly(ADP-ribosyl)ation of polynucleosomes causes relaxation of chromatin structure. *Proc. Natl. Acad. Sci. U S A.* 79(1982): 3423-3427

Pommier Y, Redon C, Rao VA, Seiler JA, Sordet O, Takemura H, Antony S, Meng L, Liao Z, Kohlhagen G, Zhang H, Kohn KW. (2003). Repair of and checkpoint response to topoisomerase I-mediated DNA damage. *Mutat Res*. 532(1-2): 173-203.

Porrua O, Libri D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol.* 16(3): 190-202.

Prado F, Aguilera A. (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J.* 24(6): 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(10): 2826-2835.

Protić-Sabljić M, Kraemer KH. (1985). One pyrimidine dimer inactivates expression of a transfected gene in xeroderma pigmentosum cells. *Proc Natl Acad Sci USA*. 82(19): 6622-6626.

Rajagopal D, Maul RW, Ghosh A, Chakraborty T, Khamlichi AA, Sen R, Gearhart PJ. (2009). Immunoglobulin switch μ sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *J Exp Med*. 206(6): 1237-1244.

Ramiro AR, Jankovic M, Eisenreich T, Difilippantonio S, Chen-Kiang S, Muramatsu M, Honjo T, Nussenzweig A, Nussenzweig MC. (2004). AID is required for c-myc/IgH chromosome translocations in vivo. *Cell*. 118(4): 431-438.

Ratner JN, Balasubramanian B, Corden J, Warren SL, Bregman DB. (1998). Ultraviolet radiation-induced ubiquitination and proteasomal degradation of the large subunit of RNA polymerase II. Implications for transcription-coupled DNA repair. *J Biol Chem.* 273(9):5184-5189.

Richardson R, Denis CL, Zhang C, Nielsen MEO, Chiang YC, Kierkegaard M, Wang X, Lee DJ, Andersen JS, Yao G. (2012). Mass spectrometric identification of proteins that interact through specific domains of the poly(A) binding protein. *Mol Genet Genomics*. 287(9): 711-730.

Rhodes D, Lipps HJ. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res*. 43(18): 8627-8637.

Rodríguez-Navarro S, Sträßer K, Hurt E. (2002). An intron in the *YRA1* gene is required to control Yra1 protein expression and mRNA export in yeast. *EMBO Reports*, 3(5): 438–442.

Rondón AG, Aguilera A. (2019). What causes an RNA-DNA hybrid to compromise genome integrity? *DNA Repair (Amst)*:102660.

Rondón AG, Mischo HE, Kawauchi J, Proudfoot NJ. (2009). Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Mol Cell*. 36(1): 88-98.

Rondón AG, Jimeno S, García-Rubio M, Aguilera A. (2003). Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. *J Biol Chem* 278(40): 39037-39043.

Rosonina E, Kaneko S, Manley JL. (2006). Terminating the transcript: breaking up is hard to do. *Genes Dev.* 20(9): 1050-1056.

Röther S, Clausing E, Kieser A, Strässer K. (2006). Swt1, a novel yeast protein, functions in transcription. *J Biol Chem*. 281(48): 36518-3625.

Ruiz JF, Gómez-González B, Aguilera A. (2011). AID induces double-strand breaks at immunoglobulin switch regions and c-MYC causing chromosomal translocations in yeast THO mutants. *PLoS Genet* 7(2): e1002009.

Sainsbury S, Bernecky C, Cramer P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* 16(3): 129-143.

Salas-Armenteros I, Pérez-Calero C, Bayona-Feliu A, Tumini E, Luna R, Aguilera A. (2017). Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. *EMBO J.* 36(23): 3532-3547.

Santos-Pereira JM, Aguilera A. (2015). R loops: new modulators of genome dynamics and function. *Nat Rev Genet.* 16(10): 583-597.

Santos-Pereira JM, Herrero AB, Moreno S, Aguilera A. (2014). Npl3, a new link between RNA-binding proteins and the maintenance of genome integrity. *Cell Cycle* 13(10), 1524–1529.

Santos-Pereira JM, Herrero AB, García-Rubio ML, Marín A, Moreno S, Aguilera A (2013). The Npl3 hnRNP prevents R-loop-mediated transcription-replication conflicts and genome instability. *Genes Dev* 27(22): 2445-2558.

Sartori G, Mazzotta G, Stocchetto S, Pavanello A, Carignani G. (2000). Inactivation of six genes from chromosomes VII and XIV of *Saccharomyces cerevisiae* and basic phenotypic analysis of the mutant strains. *Yeast.* 16(3): 255-265.

Schmidt A, Hall MN, Koller A. (1994). Two FK506 resistance-conferring genes in Saccharomyces cerevisiae, TAT1 and TAT2, encode amino acid permeases mediating tyrosine and tryptophan uptake. *Mol Cell Biol*. 14(10): 6597-6606.

Schmidlin T, Kaeberlein M, Kudlow BA, MacKay V, Lockshon D, Kennedy BK. (2008). Single-gene deletions that restore mating competence to diploid yeast. *FEMS Yeast Res*. 8(2): 276-286.

Schneider C, Leung E, Brown J, Tollervey D. (2009). The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. *Nucleic Acids Res*. 37(4): 1127-1140.

Schwertman P, Vermeulen W, Marteijn JA. (2013). UVSSA and USP7, a new couple in transcription-coupled DNA repair. *Chromosoma*. 122(4): 275-284.

Sen D, Gilbert W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*. 334(6180): 364-366.

Shanbhag NM, Rafalska-Metcalf IU, Balane-Bolivar C, Janicki SM, Greenberg RA. (2010). ATM-dependent chromatin changes silence transcription in cis to DNA double-strand breaks. *Cell.* 141(6): 970-981.

Shandilya J, Roberts SG. (2012). The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim Biophys Acta*. 1819(5): 391-400.

Sharma K, Venema J, Tollervey D. (1999). The 5' end of the 18S rRNA can be positioned from within the mature rRNA. *RNA.* 5(5): 678-686.

Shen Z, St-Denis A, Chartrand P. (2010). Cotranscriptional recruitment of She2p by RNA pol II elongation factor Spt4–Spt5/DSIF promotes mRNA localization to the yeast bud. *Genes & Development* 24(17): 1914–1926.

Sherman F, Fink GR, Hicks JB. (1986). Methods in yeast genetics. Cold Spring Harbor, NY.

Sigurdsson S, Dirac-Svejstrup AB, Svejstrup JQ. (2010). Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol Cell.* 38(2):202-210.

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(1), 19-27.

Simonsson T, Pecinka P, Kubista M. (1998). DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res*. 26(5): 1167-1172.

Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*. 516(7531): 436-439.

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(6): 794-805.

Smith SB, Kiss DL, Turk E, Tartakoff AM, Andrulis ED. (2011). Pronounced and extensive microtubule defects in a Saccharomyces cerevisiae DIS3 mutant. *Yeast*. 28(11):755-69.

Somesh BP, Reid J, Liu WF, Søgaard TM, Erdjument-Bromage H, Tempst P, Svejstrup JQ. (2005). Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. *Cell*. 121(6): 913-923.

Spivak G. (2015). Nucleotide excision repair in humans. *DNA Repair (Amst)*. 36: 13-18.

Stark JM, Jasin M. (2003). Extensive loss of heterozygosity is suppressed during homologous repair of chromosomal breaks. *Mol Cell Biol*. 23(2): 733-743.

Steinmetz EJ, Conrad NK, Brow DA, Corden JL. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature*. 413(6853): 327-331.

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(2): 163-175.

Strasser K, Hurt E. (2000). Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J*. 19: 410-420.

Svejstrup JQ. (2002). Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol.* 3(1): 21-29.

Swaminathan S, Kile AC, MacDonald EM, Koepp DM. (2007). Yra1 is required for S phase entry and affects Dia2 binding to replication origins. *Mol Cell Biol*. 27(13): 4674-4684.

Takizawa PA, DeRisi JL, Wilhelm JE, Vale RD. (2000). Plasma membrane compartmentalization in yeast by messenger RNA transport and a septin diffusion barrier. *Science*. 290(5490): 341-344.

Takizawa PA, Vale RD. (2000). The myosin motor, Myo4p, binds Ash1 mRNA via the adapter protein, She3p. *Proc Natl Acad Sci U S A*. 97(10): 5273-5278.

Thoms M, Thomson E, Baßler J, Gnädig M, Griesel S, Hurt E. (2015). The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell*. 162(5): 1029-1038.

Tkach JM, Yimit A, Lee AY, Riffle M, Costanzo M, Jaschob D, Hendry JA, Ou J, Moffat J, Boone C, Davis TN, Nislow C, Brown GW (2012). Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nature cell biology*, 14(9), 966–976.

Tran PLT, Pohl TJ, Chen CF, Chan A, Pott S, Zakian VA. (2017). PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. *Nat Commun.* 8: 15025.

Tornaletti S, Park-Snyder S, Hanawalt PC. (2008). G4-forming sequences in the non-transcribed DNA strand pose blocks to T7 RNA polymerase and mammalian RNA polymerase II. *J Biol Chem*. 283(19): 12756-12762.

Tous C, Aguilera A. (2007). Impairment of transcription elongation by R-loops in vitro. *Biochem Biophys Res Commun.* 360(2): 428-432.

Trcek T, Chao JA, Larson DR, Park HY, Zenklusen D, Shenoy SM, Singer RH. (2012). Single-mRNA counting using fluorescent in situ hybridization in budding yeast. *Nat Protoc* 7(2), 408-419.

Tsalik EL, Gartenberg MR. (1998). Curing *Saccharomyces cerevisiae* of the 2 micron plasmid by targeted DNA damage. *Yeast* 14(9), 847-852.

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

Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol*. 15(8): 795-804.

Vasiljeva L, Kim M, Terzi N, Soares LM, Buratowski S. (2008). Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. *Mol Cell*. 29(3): 313-323.

Vinciguerra P, Stutz F. (2004). mRNA export: an assembly line from genes to nuclear pores. *Curr Opin Cell Biol*. 16(3): 285-292.

Vissers JH, van Lohuizen M, Citterio E. (2012). The emerging role of Polycomb repressors in the response to DNA damage. *J Cell Sci*. 125(Pt 17): 3939-3948.

Vítor AC, Sridhara SC, Sabino JC, Afonso AI, Grosso AR, Martin RM, de Almeida SF. (2019). Single-molecule imaging of transcription at damaged chromatin. *Sci Adv*. 5(1): eaau1249.

Wahba L, Costantino L, Tan FJ, Zimmer A, Koshland D (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev*. 30(11): 1327-1338.

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(6): 978-988.

Waldherr M, Ragnini A, Jank B, Teply R, Wiesenberger G, Schweyen RJ. (1993). A multitude of suppressors of group II intron-splicing defects in yeast. *Current Genetics* 24: 301-306.

Wang P, Byrum S, Fowler FC, Pal S, Tackett AJ, Tyler JK. (2017). Proteomic identification of histone post-translational modifications and proteins enriched at a DNA double-strand break. *Nucleic Acids Res*. 45(19): 10923-10940.

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(8): 3327-3334.

Wilson MD, Harreman M, Taschner M, Reid J, Walker J, Erdjument-Bromage H, Tempst P, Svejstrup JQ. (2013). Proteasome-mediated processing of Def1, a critical step in the cellular response to transcription stress. *Cell*. 154(5): 983-995.

Woodcock DM, Crowther PJ, Doherty J, Jefferson S, DeCruz E, Noyer-Weidner M, Smith SS, Michael MZ, Graham MW. (1989). Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. *Nucleic Acids Res*, 17(9), 3469–3478.

Woudstra EC, Gilbert C, Fellows J, Jansen L, Brouwer J, Erdjument-Bromage H, Tempst P, Svejstrup JQ. (2002). A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. *Nature* 415(6874): 929-933.

Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, Libri D, Jacquier A. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*. 121(5): 725-737.

Zenklusen D, Vinciguerra P, Strahm Y, Stutz F. (2001). The Yeast hnRNP-Like Proteins Yra1p and Yra2p participate in mRNA Export through interaction with Mex67p. *Molecular and Cellular Biology*, 21(13), 4219–4232.

Zhou W, Doetsch PW. (1993). Effects of abasic sites and DNA single-strand breaks on prokaryotic RNA polymerases. *Proc Natl Acad Sci USA*. 90(14): 6601-6605.

Zorio DA, Bentley DL. (2004). The link between mRNA processing and transcription: communication works both ways. *Exp Cell Res*. 296(1): 91-97.