Papel de factores de la transcripción y la replicación del ADN en el origen de la inestabilidad genómica

Juan F. Lafuente Barquero

Tesis Doctoral

Universidad de Sevilla

2017

UNIVERSIDAD D SEVILLA

TESIS DOCTORAL

Papel de factores de la transcripción y la replicación del ADN en el origen de la inestabilidad genómica

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), para optar al grado de Doctor en Biología por el licenciado Juan F. Lafuente Barquero.

Sevilla, 2017

El doctorando,

Juan F. Lafuente Barquero

El director de tesis,

Andrés Aguilera López

Las cosas no se pueden quedar pendientes...
así que aqui estamos.

Index of content

Index of Figures

Index of Tables

ABBREVIATIONS

4-NQO	4-Nitroquinoline 1-Oxide
5-FOA	5-Fluorotic Acid
A.U.	Arbitrary Units
ARS	Autonomously Replicating Sequence
ATP	Adenosine Triphosphate
bp	Base pair
ChIP	Chromatin Immunoprecipitation
ChIP-chip	ChIP combined with microarray
CTD	Carboxyl-Terminal Domain
DNA	Deoxyribonucleic Acid
DOX	Doxycycline
DRIP	RNA:DNA hybrid immunoprecipitation
DSB	Double Strand Break
DSBR	Double Strand Break Repair
FACS	Fluorescent-Activated Cell Sorting
FACT	FAcilitates Chromatin Transcription
GC	Guanine+Citosine
Gal	Galactose
GFP	Green Fluorescent Protein
Glu	Glucose
GO	Gene Ontology
HJ	Holliday Junction
HR	Homologous Recombination
HU	Hydroxyurea
IF	Immunofluorescence
kb	kilobases
Leu	Leucine
MMS	Methyl Methanesulfonate
MPA	Mycophenolic Acid
mRNA	messenger RNA
mRNP	messenger Ribonucleoprotein Particle
NER	Nucleotide Excision Repair
NHEJ	Non Homologous End Joining
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
qPCR	quantitative PCR
rDNA	ribosomal DNA
RER	Ribonucleotide Excision Repair
RF	Replication Fork
RNA	Ribonucleic Acid
RNAPI	RNA polymerase I
RNAPII	RNA polymerase II
RNAPIII	RNA polymerase III
RNAase	Ribonuclease
RPA	Replication Protein A
rRNA	ribosomal RNA
SC	Synthetic Complete Medium

SD	Standard Deviation
Ser	serine
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
ssDNA	Single Stranded DNA
TAM	Transcription-Associated Mutation
TAR	Transcription-Associated Recombination
TCR	Transcription Coupled Repair
TLS	Translesion Synthesis
tRNA	transfer RNA
Ura	Uracil
UV	Ultraviolet light
WT	Wild-type
YFP	Yellow Fluorescent Protein

RESUMEN

Mantener la integridad del genoma no es tarea fácil para las células. El ADN sufre constantes ataques de agentes químicos y/o físicos de origen exógeno y endógeno, pero no sólo eso, los propios procesos celulares que emplean el ADN como sustrato también pueden generar daños y roturas. Dos de los procesos más fundamentales para la supervivencia de la célula, la replicación y la transcripción, son a su vez de las fuentes más importantes de daños en el ADN, especialmente en los casos de colisiones entre las maquinarias responsables de ambos procesos.

Las colisiones entre transcripción y replicación pueden darse por diversos motivos, como la acumulación de estrés torsional en el ADN, la competencia entre polimerasas por el ADN molde en regiones del genoma con alta tasa de transcripción, o la presencia de secuencias específicas de ADN con tendencia a formar estructuras diferentes a su conformación B canónica. Dentro de esta última clasificación existe un tipo de estructura nucléica que está adquiriendo una creciente relevancia, los bucles R (más conocidos como *R-loops*). De origen primordialmente co-transcripcional, los *R-loops* están constituidos por una cadena sencilla de ADN desplazada y un híbrido de ARN:ADN formado entre el ARN y su hebra de ADN molde. Los híbridos de ARN:ADN son estructuras naturales, que presentan mayor estabilidad que la propia doble hélice de ADN, y que participan en diversos procesos celulares, como la replicación del ADN mitocondrial o el cambio de isotipo de las inmunoglobulinas. Sin embargo, la presencia de una cadena sencilla de ADN desplazada, característica de los *R-loops*, o la capacidad de estos de interferir con la replicación, pueden suponer riesgos para el mantenimiento de la estabilidad de los genomas.

La presente tesis se plantea con la meta de avanzar en el conocimiento sobre los mecanismos que dan lugar a la formación y acumulación de los *R-loops*, y la de descubrir nuevos factores implicados en el mantenimiento de su homeostasis para evitar la aparición de conflictos entre la transcripción y la replicación y la inestabilidad genética que va asociada a ellos. Adicionalmente, empleando el organismo modelo *S. cerevisiae*, investigamos el funcionamiento de la proteína humana RECQL5, que a día de hoy se considera uno de los factores con el papel más claro y directo en la coordinación de los procesos de transcripción y replicación del ADN.

Nuestros resultados apoyan la creciente noción de que la presencia de *R-loops* en las células es más común de lo que se pensaba inicialmente, incluso en fondos genéticos silvestres.

Mediante análisis de secuenciación masiva y estudios bioinformáticos hemos encontrado que los telómeros, el ADN ribosómico (rDNA), los transposones y numerosos genes transcritos por la RNAPII son regiones con enriquecimiento de híbridos de ARN:ADN en una estirpe silvestre, patrón que se mantiene con cierta constancia en mutantes *hpr1*Δ. Los mutantes en este componente del complejo THO de elongación de la transcripción muestran fenotipos de acumulación de *R-loops*, hiperrecombinación e inestabilidad genética asociada a la transcripción. Nuestro trabajo sugiere que la diferencia puede no residir en la cantidad de híbridos que se formen en el fondo mutante, si no en ciertas características que diferencien estos *R-loops* de los presentes en cepas silvestres. Sin embargo, mejoras en la metodología deben ser introducidas antes de poder arrojar conclusiones más definitivas.

Paralelamente hemos investigado cómo se originan los *R-loops*. Actualmente se considera que esencialmente son estructuras formadas durante la trascripción. No obstante, no se puede descartar la posibilidad de que un transcrito pudiera hibridar con otras regiones homólogas del genoma, generando *R-loops* en trans. Nuestros resultados no muestran ningún indicio de que la formación de *R-loops* no co-transcripcionales sea independiente de la transcripción, o que su formación tenga un impacto detectable en recombinación. Otros datos que rechazan la hipótesis de que Rad51 tenga un papel activo en la formación de *R-loops* en el mutante *hpr1*Δ.

Una búsqueda de nuevos factores implicados en la homeostasis de *R-loops* nos llevó hasta la helicasa de ADN Mph1, FANCM en humanos. Los mutantes de levadura deficientes para esta proteína o su actividad helicasa acumulaban híbridos de ARN:ADN. A pesar de esto, las células *mph1*Δ no mostraron fenotipos de hiperrecombinación, ni defectos replicativos, ni interacciones genéticas con *hpr1* o *sen1*. Serán necesarios estudios futuros para dilucidar si el papel de Mph1 en la eliminación de *R-loops* es directo o indirecto.

Finalmente, hemos demostrado que la helicasa RECQL5 humana puede expresarse en levaduras, donde interacciona con proteínas ortólogas de aquellas humanas con las que se asocia de forma natural, como RNAPII o Rad51. Sin embargo, no hemos podido relacionar la inestabilidad genómica que observamos al expresar RECQL5 en levaduras con defectos producidos en la transcripción o en la replicación. No obstante, describimos por primera vez una relación funcional entre RECQL5 y la helicasa Srs2 de la levadura, que aporta nuevas vías para comprender el funcionamiento de ambas proteínas y sus papeles en el mantenimiento de la estabilidad genómica.

Introduction

Since in 1953 two scientists working in Cambridge stormed into The Eagle pub announcing that they had discovered "the secret of life" a lot has happened (Watson & Crick, 1953). The so called molecule of life, the DNA, is widely known to hold the key for the survival of all living organisms, both thanks to its capacity to be copied and to its role as blueprint for the proteins. One could think that constituting such an important molecule, DNA would be strongly protected and maintained stable inside the cell. Much on the contrary, genomes, that comprehend all the DNA from an organism, are constantly subjected to a variety of dynamic processes, plus exogenous and endogenous agents can even damage them, generating approximately 105 lesions per day in mammalian genomes (Collins, 1999). All these factors lead to the appearance of genome instability, a process necessary for the generation of genetic variability and evolution, but also a hallmark of aging, hereditary genetic diseases and cancer-related disorders.

This thesis adds to the knowledge on how the stability of genomes is maintained, focusing our study on the interface between the two most important biological processes that use the DNA as template, transcription and replication. In this introduction we will concisely cover the most important aspects about these processes, overviewing the mechanisms that control the interplay between them and the recent advances in the understanding of the impact derived from their miss-regulation.

1. DNA metabolism and Genomic instability

The concept of genomic instability encompasses the range of genetic alterations derived from unresolved damages or modifications that the DNA can accumulate due to the action of exogenous or endogenous agents. Additionally, biological processes that take place on the DNA can also lead to the occurrence of negative effects for its integrity. Although such events may be harmful for the cell, they also contribute to the generation of genetic variability necessary for evolution. Some of these processes play specific roles since, for example, variability is essential for regulated developmental mechanisms like the immunoglobulin class-

switching (Maizels, 2005).

Nevertheless, genetic instability is more commonly associated with pathological disorders, and in humans can develop into premature ageing, inherited diseases and a number of cancer predispositions. Usually, genetic instability translates into two major problems, either the failure to properly replicate the genome, or the loss of information indispensable for cell surveillance. Both situations are directly consequence of the two main functions of the DNA molecule, that are: first, pass on the genetic information from a "mother" cell to a "daughter" cell through replication of its genome; and second, serve as template for the synthesis of the functional components of the cell through its transcription.

Therefore, DNA replication and transcription are essential cellular processes that are mediated by complex machineries that copy genetic information into complementary DNA and RNA molecules, respectively. There are cases in which crosstalk between these processes and others that take place on the DNA, like recombination, have positive effects, as for example, in the case of transcription-coupled repair (Hanawalt & Spivak, 2008). However, the use of the same template by two different machineries is not an easy task and the encounter of two processes on the same genomic region may have negative consequences. This has been long proven for the case of replication and transcription, whose conflicts constitute a remarkable natural source of genomic instability (Zeman & Cimprich, 2013; Gaillard *et al.*, 2015).

1.1. REPLICATION

DNA replication is the process necessary for genome duplication, and so, for the maintenance of life. It is important to note that it occurs exactly once during the cell cycle, and in eukaryotic cells it takes place on a protein-DNA template called chromatin, not just on DNA. This makes replication fork (RF) movement considerably slower, being only about 1,000 to 3,000 bp per minute, compared to the rapid rate of 50,000 bp per min in prokaryotes. Accordingly, larger genomes take longer to replicate and so eukaryotic cells, like yeasts, have evolved to use many origins of replication per chromosome (Bell & Labib, 2016).

In *S. cerevisiae* there are about 250 to 400 replication origins distributed among its 16 chromosomes, almost all of these origins is also an autonomously replicating sequence or ARS. Their DNA sequence is not completely shared, but many have properties in common, like an A+T-rich core consensus sequence WAAAYATAAAW (W=A or T, Y=C or T). These stretches of nucleotides are binding sites for proteins involved in replication, mainly the origin recognition complex, or ORC. This complex, constituted by 6 subunits (ORC1 to ORC6, numbered from

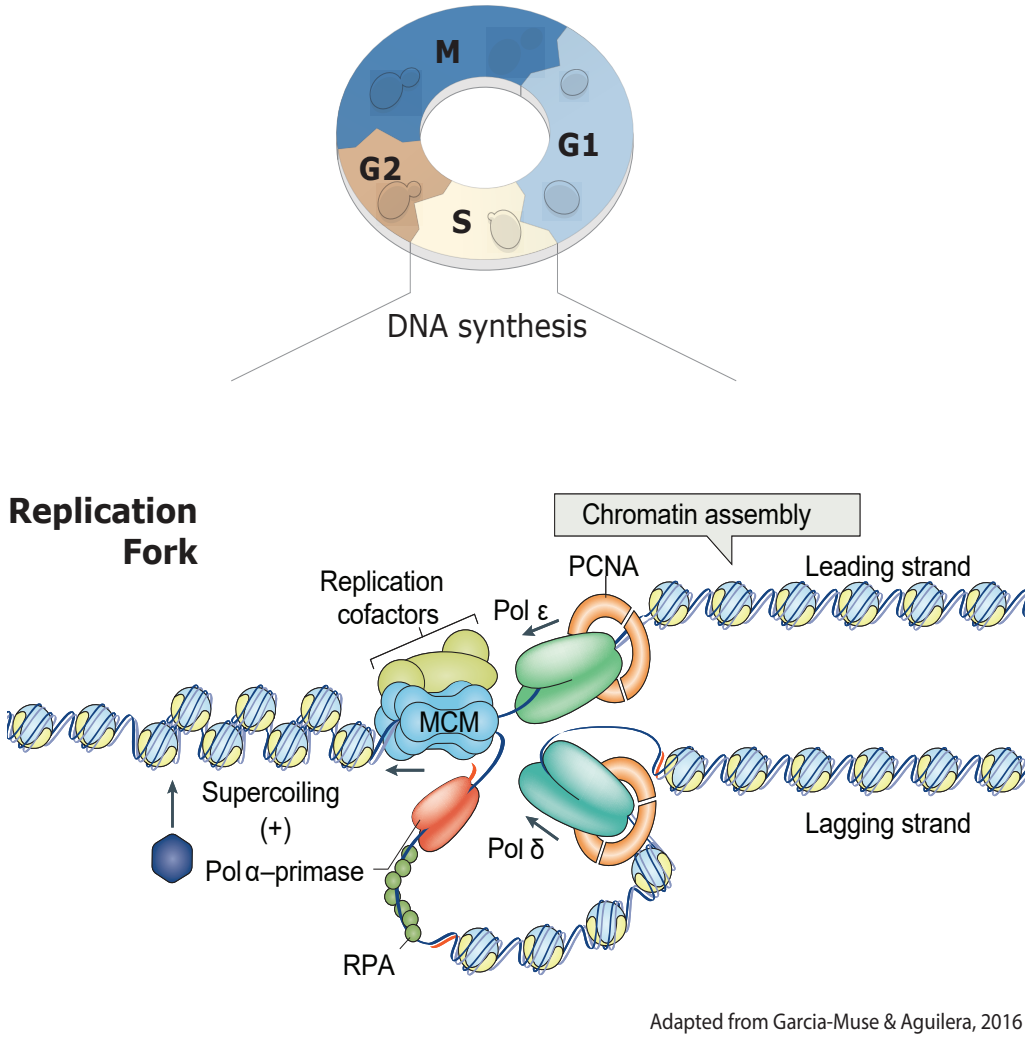


Figure I1. Cell Cycle and replication

(A) Temporal representation of S. cerevisiae 's cell cycle. Replication only takes place once during a complete turn and it is on S (synthesis) phase. (B) Replication forks are complexes constituted by the minichromosome maintenance complex (MCM9 which opens the DNA helix, and DNA polymerases Epsilon and Gamma, which extend the leading and lagging strand respectively. Pol-alpha is required for the initiation of the synthesis. RPA coats the ssDNA that is left behind in the lagging strand until its complementary strand is synthesized. Many factors are required for the correct progression of the RF, including the proliferating cell nuclear antigen (PCNA). DNA unwinding generates positive supercoiling that requires of specialized topoisomerases to be alleviated. Also, nucleosomes need to be relocated in the DNA once the RF has passed. Arrows indicate the direction of synthesis by DNA polymerases.

largest to smallest molecular weight), binds to origins of replication in an ATP-dependent manner. ORC complex is conserved to humans, and like its bacterial orthologue, DnaA, it is responsible of 3 functions at the origin: it binds to the specific DNA sequence at the origins, it participates in the DNA unwinding of the origin and it recruits other replication factors. After this, eukaryotic DNA replication process can be divided into three stages:

Licensing: In yeasts, the ORC keeps stably bound to the origin even once it has been fired. The replication initiation stage involves the assembly of the pre-replicative complex, constituted by two hexamers of Mcm proteins (minichromosome maintenance complex proteins 2 - 7), which are loaded at the replication origins exclusively during G1 phase and require the presence of the 6 subunits of the ORC, the CDC6 (cell division cycle 6) ATPase and the CDT1 (chromatin licensing and replication factor 1). After its complete assembly during the end of G1, CDC6 and CDT1 are released and the ORC and MCM are retained on the DNA ready for next step.

Origin Firing: The activation of licensed origins is triggered by the activities of cyclin-dependent kinases (CDK) and CDK-like kinases. This stage depends on the stable association of CDC45 and the DNA replication complex GINS (the name is an acronym for go-ichi-ni-san, the Japanese for 5-1-2-3, after the four subunits of the complex Sld5, Psf1, Psf2 and Psf3). Re-replication is avoided by the action of CDKs and the capture of CDT1, which prevents the reload of the MCM complex outside the G1 phase.

Progression: Replication forks (RF) are driven by holo-enzymes that are proceeded by the replicative helicases present in the CDC45-MCM-GINS complex. Leading and lagging strands are extended by Pol ε and Pol δ, respectively, which interact with additional factors that promote RF progression, like the Mrc1/CLASPIN factor or the replication pausing complex formed by Tof1-Csm3-Ctf4/TIPIN-TIM-And1 which act in the regulation and crosstalk between the replicating DNA helicases and polymerases, especially during RF stalling and restart (Errico & Costanzo, 2012). Other replication factors include the Pol α-primase complex (Pol α-pri), which contains both DNA primase and DNA polymerase subunits and

is responsible of the synthesis of each new DNA molecule initiated thanks to a RNA primer, once for the leading strand and in each Okazaki fragment for the lagging; the Rrm3 helicase and the PCNA sliding clamp along with the PCNA clamp loader, RFC.

Once DNA synthesis initiates, the new RFs start moving bidirectionally (in most cases) away from the origin. They will terminate when they meet opposing RFs from adjacent replicons.

- **Replication as source of breaks**

Although concisely explained here, such a complex process is commonly subjected to the occurrence of problems or failures, which lead to generation of replication stress. This is a source of DNA breaks and genome instability, and can be generated either by dysfunctions of the replication machinery or by blockage of the RF when it encounters an obstacle. Common barriers for replicating machineries are DNA bulky adducts, non B-DNA structures or tightly-bound DNA protein complexes. These blocks might force replication to be delayed or, in the worst cases, paused for a prolonged period of time in a phenomenon known as RF stalling. Eukaryotic cells have evolved mechanisms to restart and resume replication once the obstacles are removed, but blocked replisomes are susceptible to arrest or collapse, leading to the formation of DNA double strand breaks (DSBs) or fork regression, that would expose ssDNA stretches. Other significant source of DNA breaks is when RF encounters DNA lesions, which hamper its progression, or ssDNA gaps, which could be converted into DSBs during replication. Altogether, these kind of aggressions are liable of replication-associated genomic instability. Moreover, hyper-recombinant phenotypes of *S. cerevisiae* replication mutants, supports the notion that homologous recombination (HR) is the main repair pathway for these lesions (Hartwell & Smith, 1985; Aguilera & Klein, 1988; Yan *et al.*, 1991).

One of the common sources of replication stress are the collisions between the RF and the transcription machinery, which constitute an important source of genome instability. Since this mechanism of generation of endogenous damage will be the focus of this thesis we briefly present some details about the transcriptional process.

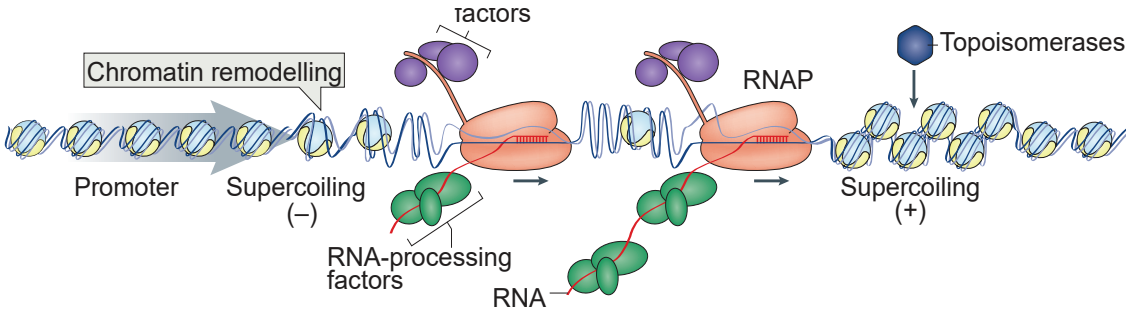
1.2. TRANSCRIPTION

Gene expression is a process that requires several steps, beginning with the transcription, which is the copying of the information from a DNA strand into a complementary sequence of RNA, and ending with the translation of that RNA into a protein in the cytoplasm. Eukaryotic transcription is evolutionarily conserved from yeast to human. It takes place inside the nucleus and the machinery required contains a core RNA polymerase that can be classified in three types depending on the nature of their products:

Table I1. Eukaryotic RNA polymerases

Polymerase	Product	Location
RNA polymerase I (RNAPI)	Larger ribosomal subunits (rRNA)	Nucleus
RNA polymerase II (RNAPII)	Messenger RNA (mRNA), most small nuclear RNAs (snRNA), small interfering RNAs (siRNA) and micro RNAs (miRNA).	Nucleus
RNA polymerase III (RNAPIII)	Transfer RNAs (tRNA), other small RNAs (snRNA & 5S rRNA), snRNA U6, signal recognition particle RNA (SRP RNA) and other stable short RNAs.	Nucleus and nucleolus

Due to RNAPII being responsible for transcription of all protein-coding genes, it has been studied in-depth and its function is well understood. In general, transcription can be divided into three stages, briefly detailed below. This process



Adapted from Garcia-Muse & Aguilera (2016)

Figure I2. Transcription.

The transcription bubble comprehends a small portion of the DNA double helix that is unwind by the RNA polymerase (RNAP) to allow the transcription of that locus. Topoisomerases are required to alleviate the negative and positive supercoiling that the RNAP generates. Also, chromatin remodeling and RNA processing factors are recruited for a successful transcription.

has some specific differences depending on the polymerase type, thus we have focused on the transcription by RNAPII:

Initiation

Transcription initiation can be also separated in two steps. In the first step, the pre-initiation complex (PIC) is assembled. Promoter elements upstream of genes are recognized by general transcription factors, like TFIIA, TFIIB, TFIID, TFIIE and TFIIH (Orphanides *et al.*, 1996; Saunders *et al.*, 2006). This phase is tightly controlled by a wide number of signaling pathways that affect sequence-specific DNA-binding transcription factors, co-activator complexes and enzymes that regulate the accessibility of DNA by modulating chromatin structure (REF). Recruitment of the RNAPII marks the end of the pre-initiation complex assembly and starts the synthesis of the first phosphodiester bonds. The second step is constituted by promoter escape and initiation. It starts when an initial transcript reaches the 10 nucleotide threshold, entering the RNA exit channel. At the end of this stage, the C-terminal tail domain (CTD) of the polymerase is phosphorylated and the interactions with the promoter elements are broken. The CTD tail of the RNA polymerase II is comprised of repeats of the heptapeptide sequence "Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7". Promoter-associated RNAPII is hypo-phosphorylated (referred as RNAPIIa) and elongating enzyme have a hyper-phosphorylated CTD (referred as RNAPIIo). All of the serine residues are susceptible of phosphorylation, and the differential patterns of CTD modification have been proposed to serve as docking platform for a variety of elongation factors and regulation proteins. These factors are required to ensure proper post-transcriptional processing of the nascent RNA as well as its export, and to help RNAPII deal with challenges due to the chromatin structure, difficult DNA sequences or lesions that the polymerase may encounter during elongation, usually causing transient pausing or enzyme arrest.

Elongation

After phosphorylation of CTD Ser5 and escaping the promoter, most of the initiation factors are substituted for new proteins required for the elongation and processivity of the polymerase (Saunders *et al.*, 2006). Double stranded DNA (dsDNA) enters from the front of the enzyme and is unzipped to provide the template

strand for RNA synthesis. DNA strands and nascent RNA oligonucleotide exit from separate channels, the two DNA strands re-hybridize at the trailing end of the transcription bubble and the RNA is quickly covered with export and splicing factors to form a ribonucleoparticle (mRNP). Full elongation capacity is achieved after a few adjustments produced during a pause in the surroundings of the promoter. These pauses are induced by the action of the DSIF complex (DRB Sensitivity Induction Factor) formed by Spt4, Spt5 and the Negative Elongation Factor (NELF). The RNAPII can stop due to a variety of reasons such as difficult template sequences, depletion of ribonucleotide pools, miss-incorporation of bases, formation of ternary structures in the RNA or presence of other proteins in the DNA strand. One key factor for the polymerase to resolve these obstacles is the presence of the TFIIS elongation factor (Dst1), which is responsible of inducing the nuclease activity of the RNAPII necessary to recover from backtracking on the template as a consequence of the stop.

It is also worth noting in this stage the role of chromatin modifiers to accomplish a successful transcription round. Since the polymerase machinery needs a protein-free DNA template, nucleosomes have to be temporarily removed either by displacement or modification. These can take place by different mechanisms that involve the action of histone chaperones, chromatin remodeling factors or histones modifying enzymes. There have been uncovered many covalent histone modifications like acetylation, methylation phosphorylation or ubiquitination (Rando & Winston, 2012). These modifications are performed by a number of complexes, like SAGA, PAF and COMPASS (Selth *et al.*, 2010). Other chromatin remodeling factors are SWI/SNF, ISWI, CHD1, INO80/SWR, RSC and FACT (formed by Spt16 and Pob3) which facilitate the transcriptional process (Winkler & Luger, 2011; Herrera-Moyano *et al.*, 2014).

Termination

The last stage of transcription, upon successful transcript formation, leads to the dissociation of the newly synthesized mRNP and release of the RNA polymerase from the template DNA. The mechanism differs between the polymerases and it is in general the most poorly understood step of transcription. For the RNAPII this step

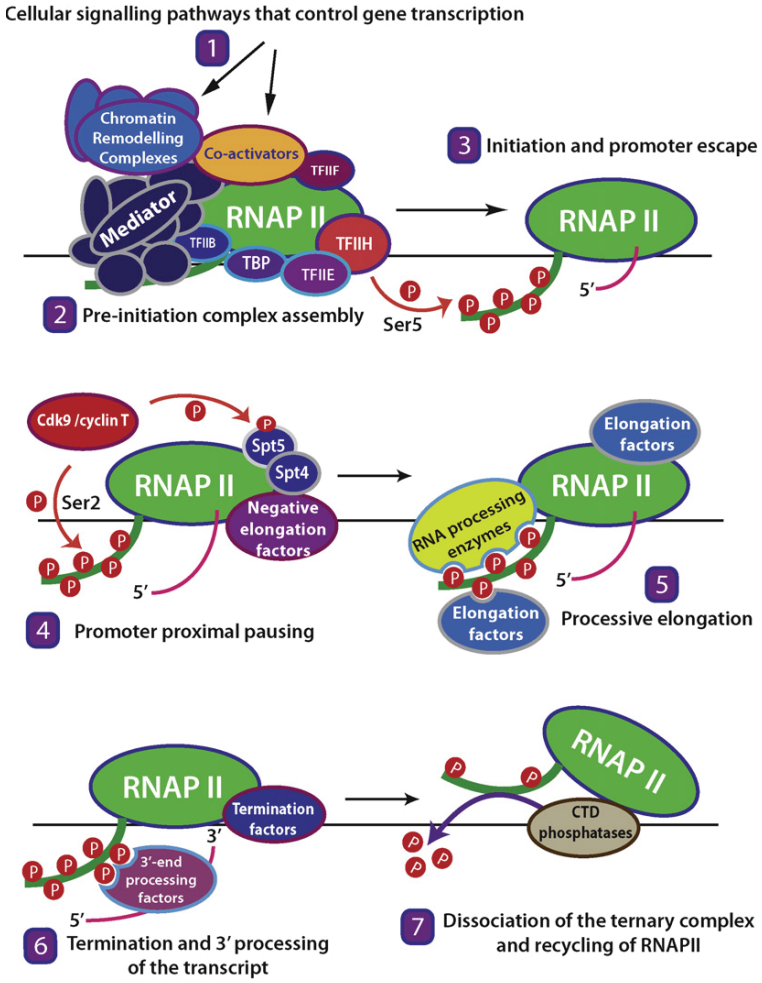


Figure I3. Transcription process

Overview of the RNAPII transcription cycle. (1) The initiation step of RNAPII transcription is tightly controlled by various cellular signaling pathways that act on the level of transcriptional co-factors and chromatin remodeling activities to ensure proper recruitment of the factors required for a complete pre-initiation complex. (2) Initiation and promoter escape (3) are characterized by the phosphorylation of the CTD, formation of an open RNAPII–DNA template complex, and synthesis of the first phosphodiester bonds of the nascent transcript. (4) The initiation–elongation transition often involves promoter-proximal pausing of early RNAPII elongation complexes by negative elongation factors and the Spt4/Spt5 complex. (5) During the processive elongation stage, the CTD is hyper-phosphorylated and forms a docking platform for the CTD-binding factors that are necessary for the processing of the elongating transcript as well as for the processivity of the RNAPII holoenzyme. (6) Termination of RNAPII transcription is triggered by termination signals that lead to 3'-end processing of the transcript and its release from the elongation complex. (7) RNAPII proceeds further downstream before being dissociated from the template. A group of CTD phosphatases de-phosphorylate the CTD and allow RNAPII to be recycled for a new round of transcription

implicates the de-phosphorylation of the CTD tail, which allows the polymerase to be recycled for transcription re-initiation. Part of the pre-mRNA processing occurs co-transcriptionally and it is coordinated with splicing. Nascent mRNA is cleaved about 20 to 30 nucleotides after the polyadenylation site (PAS) and there are two models for how termination is triggered. One, dubbed the allosteric model, implicates a conformational change in the RNAPII probably induced by the

recruited polyadenylating factors.

These cause pausing of the polymerase before its release. The alternative model (torpedo termination) is based on the fact that after pre-mRNA past the PAS, the polymerase continues producing an RNA chain that is recognized by the exonuclease Rat1 in yeast and XRN2 in humans which act as molecular triggers to release the RNAPII (Proudfoot, 2016) .

The fact that mRNA maturation factors have been found at promoter regions suggest that both initiation and termination could be connected. After the first transcription round a DNA loop is formed that promotes the restart of transcription by approaching both termination region and promoter regions (Glover-Cutter *et al.*, 2008; Singh & Hampsey, 2007; Proudfoot, 2011).

- **Transcription-associated genome instability**

Studies over the last three decades have brought to light that transcription is one important source of DNA variability . Transcription constitutes an important challenge for the stability of the DNA, and has been found to be responsible of increasing rates of mutation and recombination. These processes are conserved from prokaryotes to eukaryotes and are known as Transcription-associated mutation (TAM) and Transcription-associated recombination (TAR) (Gaillard & Aguilera, 2016).

Genome instability associated with transcription could, in principle, be explained by the transient exposure of a vulnerable ssDNA, that could constitute the non-transcribed strand displaced after the RNAPII passage. This is consistent with previous studies that demonstrate a synergistic effect between transcriptional state of a DNA region and its susceptibility to DNA damaging agents such as methylmethanesulfonate (MMS) (García-Rubio *et al.*, 2003). Another type of transcription impact on DNA derives from the topological changes that the nucleic acid must undergo to allow the movement of the transcription machinery, in which positive and negative supercoiling accumulate ahead and behind the RNAPII, respectively, making the stretches of DNA affected more susceptible to damage (Bermejo *et al.*, 2009; Sperling *et al.*, 2011).

On the other hand, the mechanisms of TAR are not completely understood, but since HR is the principal pathway for the repair of DNA breaks that take place during replication, increasing evidence suggests that TAR is consequence of transcription-replication collisions that can cause the collapsing of the RF (Aguilera, 2002; Kim & Jinks-Robertson, 2012; Gaillard *et al.*, 2013; García-Muse & Aguilera, 2016).

2. Transcription-Replication collisions and DNA damage

Given that transcription and replication are two essential processes for cell viability and proliferation, and that they occur frequently, a high incidence of encounters between both machineries is to be expected. One important difference between DNA and RNA polymerases is that when transcribing, the elongation machinery embraces the double-stranded DNA. By contrast, the replicative machinery has two DNA polymerase subunits, each using one single strand as template. Furthermore, while multiple RNA polymerases can simultaneously transcribe the same region, replisomes always move alone.

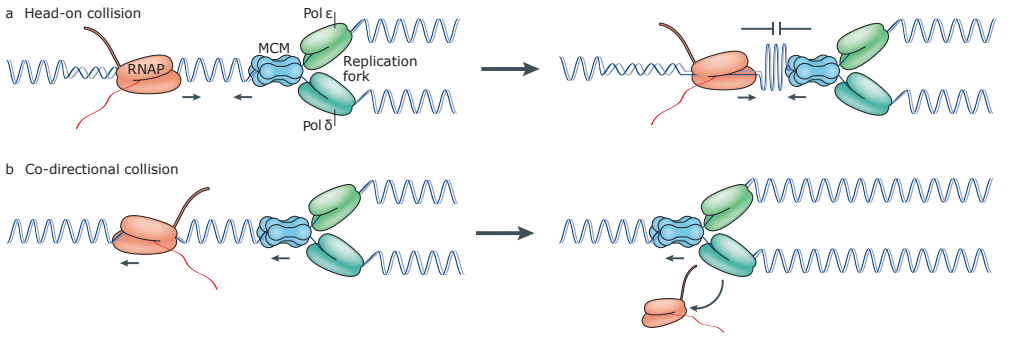
How the replication machinery progress along a double-stranded DNA molecule occupied by an RNA polymerase is an old question, due to the fact that RNA and DNA polymerases use the same template, and RNAPs are known to be one of the principle obstacles for RFs progression (Liu & Alberts, 1995; Deshpande & Newlon, 1996; Bermejo *et al.*, 2012). These collisions occur at transcribed sites driven by different polymerases (Prado & Aguilera, 2005; Gottipati *et al.*, 2008; de la Loza *et al.*, 2009; Gaillard *et al.*, 2013) and constitute a recurrent problem to deal with in all living organisms from bacteria to humans (Prado & Aguilera, 2005; Helmrich *et al.*, 2011; De Septenville *et al.*, 2012). Also, transcription-mediated chromatin changes may facilitate firing of DNA replication origins, which can have a positive effect on replication initiation but may also lead to higher probabilities of collisions.

2.1. Types of transcription-replication collisions

There are two types of transcription-replication collisions **(Figure I4)**, that depend on the direction in which the encountering polymerases advance through the DNA template. However, regardless the collision between machineries being co-directional or head-on, replisomes cannot pass through an elongating RNA polymerase. Reports suggest that the head-on type of collision is more damaging for the DNA and requires of the HR repair. Initial replisome factors likely to encounter the RNAP complex are the replicative helicases. Frontal encountering of obstacles could inactivate them producing the stalling of the RF (Srivatsan *et al.*, 2010; Prado & Aguilera, 2005). It's interesting that co-directional collisions, on the other hand, are not such dramatic events (Pomerantz & O'Donnell, 2008). A possible explanation could be that these type of encounters are more easily resolved once transcription is terminated. In any case, it has not been shown if the DNA and RNA polymerases ever actually make contact, something unlikely.

2.2. Sources of transcription-replication collisions

Supercoiling generated by transcription is one of the main sources of collisions. Transcription causes negative and positive supercoiling, ahead and



Garcia-Muse & Aguilera, 2016

Figure I4. Types of transcription-replication collisions

(A) Head-on collisions are the most dramatic for the genome stability. Opposing RNAP and a replication fork lead to pausing and blockage of the later machinery, with the possible outcome of replication fork collapsing and formation of DNA breaks. (B) Co-directional collisions when both polymerases progress in the same direction are resolved by displacement of the RNAP from the template DNA. MCM, minichromosome maintenance complex; Pol δ, DNA polymerase α; Pol ε, DNA polymerase δ.

after the RNA polymerase, respectively. DNA supercoils that accumulate between converging transcription and replication machineries impede DNA unwinding between the two advancing complexes. The resulting torsional stress in the DNA is relieved by Topoisomerases (classified as I or II depending on whether they catalyze the breakage of one or two DNA strands). Indeed, it has been shown that absence of both TOPO1 and TOPO2 (Top1 and Top2 in yeast) causes unresolved torsional stress that can promote transcription-replication conflicts, both in yeast and in humans (Brill *et al.*, 1987; Bermejo *et al.*, 2007; Tuduri *et al.*, 2009). Co-directional encounters are not subject to this situation, as the positive topological stress arising in front of the fork would neutralize the negative supercoiling behind RNAP (Pomerantz & O'Donnell, 2008).

It is not clear whether RFs slow down upon reaching highly transcribed genes in yeast. Azvolinsky *et al.* (2009) provided evidence toward forks slowing down at highly RNAPII-transcribed genes, but Sekedat *et al.* (2010) study concluded that this effect was only observed in a small number of forks. In any case, transcription rates are inherently heterogeneous, and RNAP are prone to spontaneous or regulated pausing at certain DNA sequences or can be persistently blocked by DNA lesions. Backtracked RNAPs are one type of transcriptional roadblocks, which result in the displacement of the 3' end of the nascent RNA from the active site, trapping the enzyme in a highly stable but transcriptionally inactive state (Nudler, 2012). Consistently, the absence of anti-backtracking factors in bacteria leads to RF arrest (Mirkin *et al.*, 2006) and can induce DSBs (Dutta *et al.*, 2011).

Additionally, there are DNA sequences (especially repetitive sequences or with GC enrichment) that are prone to accumulate other kind of structures different to the canonical B DNA, such as hairpins, triplex DNA or G-quadruplexes. These kind of structures represent a barrier for the advance of replication forks and have been associated with DSB hot-spots in the genome (Zhao *et al.*, 2010). The formation of these structures is suggested to be favored by the ssDNA strands that are temporally exposed during DNA replication, but it is possible that also occurs during transcription, as result of the transiently accumulated negative supercoiling behind the elongating RNA polymerase.

Another kind of transcription-associated structures that can halt replisome and generate transcription-replication conflicts are co-transcriptional generated RNA:DNA hybrids (known as R-loops when formed outside the transcription bubble), which will be reviewed later in this introduction.

2.3. Preventing transcription-replication collisions

On account of all these factors, collisions are not rare events and thus, cells have evolved mechanisms to prevent their occurrence. Bacteria have developed a bias towards genes located co-directional with replication origins (Merrikh *et al.*, 2012). In eukaryotes this bias is not obvious, but they have alternative strategies to deal with head-on collisions. As for example, *S. cerevisiae* cells have established replication fork barriers (RFBs) surrounding the highly transcribed ribosomal DNA (rDNA). Although replication is initiated bidirectionally at the ARS located near the rDNA, these RFBs work polarizing the RFs to prevent head-on collisions by blocking

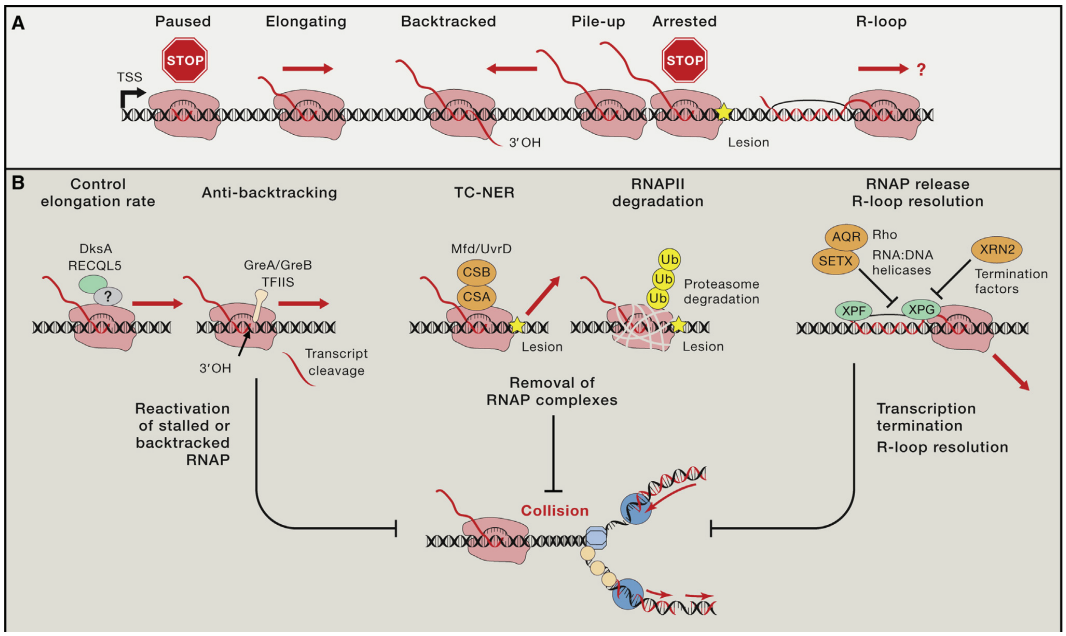


Figure I5. Co-transcriptional mechanisms to suppress transcription-replication collisions.

There are many mechanisms by which cells can resolve transcription roadblocks. Main proteins or factors involved in each pathway is shown either prokaryotic or eukaryotic. RECQL5 is a human helicase that controls transcription elongation rate to reduce conflicts with replication. The anti-backtracking factors GreA/GreB and TFIIS help to resume transcription. There are pathways that lead to RNAP proteolytic degradation when it gets stalled at DNA damage sites by TC-NER or poly-ubiquitylation signaling. Transcription termination and resolution of R-loops is mediated by the Rat1/XRN2 exonucleases and RNA:DNA helicases such as Sen1/SETX or AQR. Additionally, R-loops can by processed by the TC-NER endonucleases XPF/XPG. All of these pathways help to remove transcriptional blocks and minimize collisions with replication forks. Yellow star represents DNA lesion.

forks from progressing in the direction opposite to RNAPI transcription (Linskens & Huberman, 1988; Ivessa *et al.*, 2000). Similar RF pauses have been described also at tRNA genes in yeast and at the ribosomal gene arrays of other organisms (Labib & Hodgson, 2007). Other strategies include the redistribution by the RNAP of replication initiation factors in G1 on chromatin prior to their activation (Gros *et al.*, 2015). Evidence suggests that higher eukaryotes coordinate transcription and replication in the nucleus spatially and temporally. Reports indicate that, during S-phase, eukaryote cells compartmentalize replication so that DNA replication and transcription machineries occupy distinct nuclear territories and act at distinct times (Wei *et al.*, 1998), and more recently, a nascent RNA capture assay showed a global anti-correlation between transcription and replication timing (Meryet-Figuiere *et al.*, 2014)

Cells also count with active mechanisms to minimize transcription-replication encounters. The transcription machinery itself is an important factor on this regard as it has been inferred from the studies of defective RNA polymerase strains. Some of these mutants are able to compromise the stability of transcription complexes, being able to suppress growth defects generated by the absence of DNA helicases that help resolve collisions, such as UvrD, Rep and DING in bacteria (Baharoglu *et al.*, 2010), suggesting that more unstable transcription machineries suppose weaker obstacles for replication fork progression. Furthermore, other RNAPII mutants have been described in yeast to present elongation defects and exhibit important replication impairment, due to prolonged or tighter unions with the template DNA as determined by chromatin immunoprecipitation (ChIP) (Felipe-Abrio *et al.*, 2014). These data suggest that the capability of an RNAPII to be evicted from the DNA can vary, and this characteristic influences on replisome advance.

The same way the RNA polymerase holoenzyme represents an obstacle for the replication fork progression, other non-nucleosomal protein complexes can impede the replisome advance. In yeast, one such barrier is the protein Fob1, which is recruited to specific RFB (Replication Fork Blocking) sequences near the 35S rRNA gene. Replication progression through these RFB-Fob1 complexes requires Rrm3, and in its absence, cells present increased replication pauses at

rDNA, DNA breaks and accumulation of excised rDNA circles (Torres *et al.*, 2004). Mutants lacking Rrm3 also show increased fork pausing at tRNAs and telomeres, supporting its participation in other pause sites (Ivessa *et al.*, 2002; Ivessa *et al.*, 2003). However, although some highly transcribed genes also show RF pausing sites, Rrm3 does not accumulate in all of them, suggesting that other factors may have roles in the prevention or resolution of collisions. Consistent with this view, the *E.coli* transcription elongation factors GreA and GreB , which promote the release of backtracked and stalled RNAP seem to participate in the reduction of conflicts, as their absence in high transcription conditions leads to a complete block of replication (Tehranchi *et al.*, 2010). In yeast, the elongation factor TFIIS holds a similar role, facilitating the catalytic activity of the RNAPII in the backtracking process (Dutta *et al.*, 2015).

Another arising important factor on this regard is the human helicase RECQL5 (RecQ-like dependent DNA helicase Q5). RECQL5 is a member of the RecQ family of ATP-dependent DNA helicases, which is one key player in genome maintenance and stability. Conserved from bacteria to humans, while lower organisms, like bacteria and yeast, express only one member of the family, most mammals possess five RecQ helicases: RECQL1, BLM, WRN, RECQL4 and RECQL5. Defects in BLM, WRN and RECQL4 are associated with monogenic disease syndromes characterized by genome instability, multiple cancer predispositions and/or premature ageing phenotype. Recent studies suggest links between mutations in the two other members, RECQL1 and RECQL5, with different types and predisposition to cancer development (Fu *et al.*, 2017). RecQ helicases are involved in base excision repair (BER), DNA double strand break repair (DSBR), intra-strand cross link repair, recovery of stalled replication forks and telomere processivity and stability (Singh *et al.*, 2011; Croteau *et al.*, 2014). RECQL5 is of special interest owing to its particular versatility, forming protein complexes with a diverse set of cellular partners like the RNAPII, the PCNA complex or TOPII (Cohen *et al.*, 2013), to coordinate various processes including replication, transcription and recombination.

Up to date, RECQL5 is the protein with the best characterized active role in preventing transcription-replication conflicts. Indeed, it is the only RecQ helicase

found to interact with the RNAPII (Aygün *et al.*, 2008). Using high throughput sequencing techniques it was described that overexpression of RECQL5 led to reduced transcription rates, and its depletion induced a general transcription upregulation increasing transcription pausing, arrest and backtracking (Saponaro *et al.*, 2014). This supports the idea that RECQL5 has a negative regulatory impact on transcription elongation and that uncontrolled transcription is a source of DNA damage. RECQL5 also associates with the replicative sliding clamp PCNA (Kanagaraj *et al.*, 2006), pointing to a direct participation in replication. Furthermore, DSBs accumulate in RECQL5-depleted cells during replication in a RNAPII-dependent way, since breaks locate in actively transcribing genes and transcription inhibitors suppressed their appearance (Hu *et al.*, 2009). Recently, it has been shown that RECQL5 promotes TOP1 sumoylation, which is required for the efficient association of the topoisomerase with RNAPII and the recruitment of other mRNA-processing factors to transcriptionally active sites, as inferred by the accumulation of RNA:DNA hybrids in defective cells depleted of RECQL5 (Ramamoorthy *et al.*, 2012). These data suggest for RECQL5 an active role in the maintenance of genome stability by preventing the occurrence of transcription-replication collisions, whether by limiting transcription itself or by reducing the accumulation of co-transcriptional RNA:DNA hybrids.

Even though the role of chromatin has not been studied in relation with the regulation of transcription-replication collisions, both processes take place in the context of a highly structured template. Histone chaperones and chromatin-remodeling complexes are required not only to allow the passage of polymerases, but also to revert the DNA to its original packaged state to avoid further damages. This has been clearly shown for the FACT complex (Facilitates Chromatin Transcription), which was found to have roles originally in transcription but also in replication. Cells lacking a functional FACT complex accumulate high levels of transcription-replication collisions, and exhibit RF progression impairment which correlates with higher levels of genome instability, both in yeast and human cells (Orphanides *et al.*, 1996; Abe *et al.*, 2011; Foltman *et al.*, 2013; Herrera-Moyano *et al.*, 2014).

2.4. Resolving transcription-replication collisions

The main consequence of transcription-replication collisions is their impact on genome stability. The resulting collapsed replication forks can induce the generation of DSBs and chromosomal breakage, therefore, collisions and their damage are expected to be sensed by the DNA damage response (DDR) and other DNA repair pathways.

Although is not confirmed that DDR is able to sense transcription-replication collisions directly, the damage generated in such events is enough to activate the DNA damage checkpoint, as collisions may precede the uncoupling of the leading and lagging strands, generating long stretches of ssDNA (Lopes *et al.*, 2006; Heller & Marians, 2006). These stretches of ssDNA are sensed by the Ataxia Telangiectasia and Rad3-related (ATR, Mec1 in budding yeast)-dependent replication checkpoint, which activates to protect the integrity of the replication fork. This is unless DSBs are generated, in which scenario the Ataxia Telangiectasia Mutated (ATM, Tel1 in budding yeast)-dependent checkpoint would be the pathway triggered (Maréchal & Zou, 2013). There is evidence of ATR/Mec1 role in regulating transcription-replication collisions by dismantling the pre-initiation complex at tRNAs (Nguyen *et al.*, 2010) or by triggering the eviction of histones to facilitate the processing and release of RNAPII (Im *et al.*, 2014; Poli *et al.*, 2016). ATR/Mec1 activation starts a cascade that phosphorylates many downstream targets, including the tumor suppressor p53 in mammalian cells, which has also been recently involved in preventing transcription-replication collisions by reducing topological stress (Yeo *et al.*, 2016).

ATR-mediated replication stress response also triggers release from the nuclear pore of genes being transcribed. The phenomenon, known as 'gene gating', consists in the localization to the nuclear periphery of actively transcribed genes, presumably to facilitate the nuclear export of the transcripts, it is mediated by RNA binding proteins that coordinate the processing of the pre-mRNA and its transport through nuclear pores to the cytoplasm (Blobel, 1985; Cabal *et al.*, 2006). Interestingly, mutations in some of these factors partially suppress the replication stress phenotypes in checkpoint mutants, suggesting that blocked replisomes at

genes located near the nuclear pore cannot be released in absence of a functional checkpoint to presumably allow supercoiling relief and replication restart (Bermejo *et al.*, 2011; Bermejo *et al.*, 2012), raising the possibility that this pathway regulates fork stability through control of transcription-coupled processes.

Whether specific repair pathways are triggered at transcription-replication collisions sites after activation of the DDR by the DNA damage checkpoint is not clear yet. However, an increasing body of work suggests that transcription-replication blocks require the activity of factors involved in DNA recombination (Prado & Aguilera, 2005; Wellinger *et al.*, 2006a; Gottipati *et al.*, 2008). This is supported by recent findings that have linked the tumor suppressor BRCA1 and BRCA2,

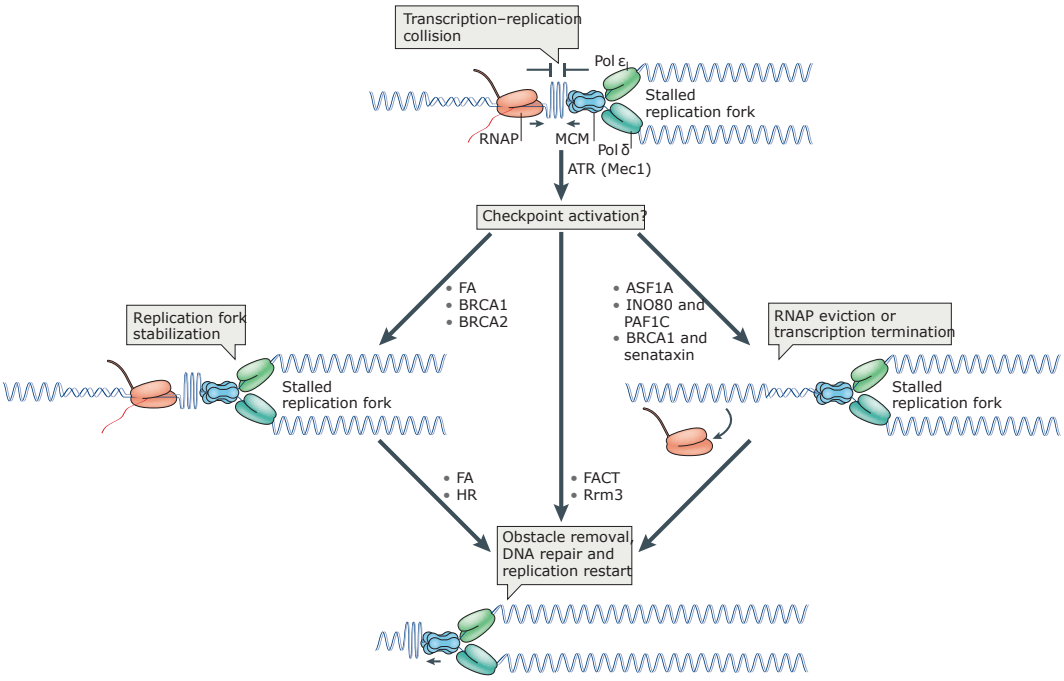


Figure I6. Response pathways for transcription-replication collisions.

A stalled replication fork can activate the ataxia telangiectasia and Rad3 related (ATR; Mec1 in yeast)-dependent checkpoint to solve collisions and avoid their consequences. Specific factors are recruited to resolve or prevent the obstacle, including DNA helicases (such as Rrm3), mRNA-processing proteins or chromatin-remodeling complexes (such as FACT). In addition, the RNA polymerase (RNAP) could be released at transcription termination sites, with the help of BRCA1 and the helicase senataxin or the chromatin-remodeling complex INO80 and transcription factor complex PAF1C, as a way to avoid the collisions. Similarly, at tRNA transcription sites, RNA Pol III is directly evicted during S phase. Stabilization and resumption of stalled forks at transcribed DNA regions can occur via DNA repair factors, such as those of the Fanconi anaemia (FA) repair pathway, including the tumour suppressors BRCA1 and BRCA2. Finally, different DNA repair pathways can act at collision sites, if these degenerate into DNA lesions. Although direct involvement of the ATR-dependent checkpoint has been reported in some examples, the degree of its involvement in other cases is yet unknown. ASF1A, anti-silencing function protein 1 homologue A; HR, homologous recombination; MCM, minichromosome maintenance complex; Pol δ, DNA polymerase δ; Pol ε, DNA polymerase ε.

which are involved in DSB repair, with the prevention or resolution of RNA:DNA hybrids (Hill *et al.*, 2014; Bhatia *et al.*, 2014), providing a mechanistic connection between recombination and replication-stress induced by R-loops. For instance, BRCA1 was shown to interact with SETX to prevent damage at certain R-loop-prone termination regions (Hatchi *et al.*, 2015), suggesting a role in R-loop-formation suppression or processing. In parallel, BRCA2 interaction with TREX-2 (Bhatia *et al.*, 2014) reinforces the hypothesis that transcription factors work together with DNA repair proteins to reduce the number and/or the impact of transcription-replication collisions. BRCA2 is also member of the Fanconi Anemia (FA) DNA-repair pathway of interstrand crosslinks, which has recently been demonstrated to be activated by transcription-replication conflicts linked to the accumulation of RNA:DNA hybrids (Schwab *et al.*, 2015; García-Rubio *et al.*, 2015). Accordingly, the Fanconi Anaemia Complementation group M protein (FANCM, Mph1 in yeast) was shown to hold the capacity of RNA:DNA branch migration *in vitro*, which has been used to suggest a role for the translocase activity of FANCM to directly resolve the fork-blocking R-Loop (Schwab *et al.*, 2015), but this has still to be demonstrated *in vivo*.

Additional mechanisms that support the existence of cooperative pathways between transcription and replication have been found. For example, it has been described in yeast that the transcription factor Hog1 has the capacity to delay replication timing by influencing early-firing origins and replication fork progression itself through the interaction with Mrc1 (Claspin in humans), which is a replisome component (Duch *et al.*, 2013) and a substrate for the replication checkpoint Mec1 (Alcasabas *et al.*, 2001). These data suggest that there is a tight regulation of mechanisms to coordinate transcription and replication in special situations, helping the organisms overcome stressful situations that require transcriptional burst, without putting in danger their genome stability.

3. R-loops

As mentioned before, RNA:DNA hybrids (more specifically R-Loops) are usually present at transcription-replication conflict sites, in some circumstances as

triggers for that collisions and in others as consequence of the blockage produced between the machineries. In the last two decades, a large quantity of data has been generated about the nature of this nucleic structure, on how they are formed and which tools have evolved the cells to regulate its presence (nicely reviewed by Santos-Pereira & Aguilera, 2015)

RNA:DNA hybrids are naturally occurring structures formed during processes like replication, in which they prime DNA synthesis, or transcription, at the active site of the RNA polymerase during ribonucleotide triphosphate incorporation. These hybrids are short and are more stable than double-stranded DNA (dsDNA), adopting a conformation midway between B form of dsDNA and the A form of double-stranded RNA (dsRNA) (Roberts & Crothers, 1992). On the other hand, so called R-loops are longer RNA:DNA hybrids formed as result of the hybridization of a nascent RNA with its complementary DNA strand, leaving the other single-stranded DNA molecule displaced. Formation and/or stabilization of R-loops *in vivo* is favored by different DNA features, such as G-content or negative DNA supercoiling, for example (Santos-Pereira & Aguilera, 2015).

Although considered rare events not so long ago, R-loops are found *in vivo* as natural intermediates in specific cellular processes, such as *E. coli* plasmid replication, mitochondrial replication or immunoglobulin (IG) class switching (Aguilera & García-Muse, 2012). However, recent studies have uncovered a wider presence than previously foreseen, and R-loops have been shown to be an important source of genome instability, a hallmark for cancer cells (Aguilera & García-Muse, 2012; Santos-Pereira & Aguilera, 2015) and to have an important impact on transcription (Skourti-Stathaki *et al.*, 2014), chromatin structure (Castellano-Pozo *et al.*, 2013) and replication (Chang & Stirling, 2017).

Recent genome-wide approaches have permitted the detection of R-loops in many loci. An i*n silico* tool was developed to identify R-loop-prone sites over the human genome and other model organisms (R-Loop DB - http://rloop.bii.a-star.edu.sg/), which identifies a high number of putative R-loop-forming sequences (250,000) in 59% of human genes (Jenjaroenpun *et al.*, 2015). Also, a variety of techniques have been applied to genome wide detection of R-loops in human cells (DRIP followed

by sequencing, DRIP-seq (Ginno *et al.*, 2012; Ginno *et al.*, 2013)) or in *S. cerevisiae* (Chromatin Immunoprecipitation followed by seq, ChIP-seq (El Hage *et al.*, 2014) or DRIP followed by microarray, DRIP-chip (Chan *et al.*, 2014) and DRIP treated with S1 nuclease followed by sequencing, DRIP-S1-seq, (Wahba *et al.*, 2016)). All revealed numerous R-loop hotspots including the rDNA region, telomeres, Ty transposons, RNA Polymerase III-transcribed genes and highly expressed RNAPII-transcribed genes.

With the discovery of this high incidence of R-loops, several cell mechanisms and pathways to regulate R-loop homeostasis are being uncovered that could be classified in two categories: those implicated in the removal of R-loops once they are formed, and those preventing their formation.

3.1. Factors that remove R-loops.

R-loops can be removed by the RNase H enzymes (RNase H1 and RNase H2), which degrade the RNA moiety of an RNA-DNA hybrid molecule. In yeast, RNase H1 is encoded by the RNH1 gene, while the RNase H2 enzyme is a trimeric complex made up of RNH201 (the catalytic subunit), RNH202, and RNH203 (Cerritelli & Crouch, 2009). Another source of RNA:DNA hybrids results from the miss-incorporation of ribonucleoside triphosphates (rNTPs) into DNA by the replicating DNA polymerase (McElhinny *et al.*, 2010). Ribonucleotide excision repair (RER) is performed primarily by the RNase H2 enzyme in combination with the endonuclease Fen1 (Sparks *et al.*, 2012). Importantly, RNase H2 is competent to cleave both mono- as well as consecutive –ribonucleotides, whereas RNase H1 requires stretches of at least 4 ribonucleotides (Cerritelli and Crouch, 2009).

Alternatively, RNA:DNA hybrids can also be removed via displacement of the RNA strand, performed by RNA:DNA helicases which unwind the RNA:DNA hybrids or limit their formation. As for example, yeast Pif1 is involved in mitochondrial DNA maintenance, telomeric DNA synthesis, rDNA replication and Okazaki fragment maturation (Boulé & Zakian, 2009). The yeast Sen1 and its human homolog SETX have been also linked to the removal of R-loops *in vivo* since mutants in these genes accumulate high levels of RNA:DNA hybrids, especially at transcription

termination sites (Mischo *et al.*, 2011; Skourti-Stathaki *et al.*, 2011). More recently, the human aquiarius (AQR) protein, which is an RNA helicase of the same family as Sen1/SETX, has also been proposed to have a role in R-loop removal, but there is still more research needed to confirm it (Sollier *et al.*, 2014).

3.2. Factors that prevent R-loop formation.

Several studies in bacteria and eukaryotes have also shown an effect of supercoiling regulation by topoisomerases (TOP) in R-loop formation. Local unwinding produced by the negative supercoiling that follows an elongating RNA polymerase is a suitable substrate for the formation of co-transcriptional RNA:DNA hybrids. This has been proven in bacteria since *topA rnhA* mutants are not viable, and overexpression of RNase H1 in *topA* backgrounds suppresses its growth defect (Drolet *et al.*, 1995), and supported in yeast, where deletion of both Top1 and Top2 increases R-loops accumulation at the rDNA locus, generating RNAPI stalling and defects in pre-rRNA synthesis (El Hage *et al.*, 2010), or in humans, where TOP1-deficient cells show DNA breaks at active genes and replication defects that are suppressed by RNase H1overexpression (Tuduri *et al.*, 2009).

The other key element for the formation of the R-loops is the nascent mRNA, as deficiencies in the assembly of the messenger ribonucleoparticle (mRNP) can strongly stimulate R-loop formation. The THO complex was the first transcription-coupled RNA processing complex that was demonstrated to limit the formation of RNA-DNA hybrids in a co-transcriptional manner (Huertas and Aguilera, 2003) by binding to the nascent RNA and ensuring that it gets properly assembled into an mRNP that is efficiently exported. Many other transcription factors have been uncovered since then, like the splicing factor SRSF1 gene in chicken and human cells (Li *et al.*, 2005), or the yeast ataxin 2 RNA-binding protein, PAB1-binding protein 1 (Pbp1), which prevents R-loop formation by interacting with non-coding RNAs (ncRNAs) generated at the intergenic spacers of the rDNA locus (Salvi *et al.*, 2014). Additionally, control of the mRNAs half-life is also important to reduce the probability of R-loop formation, as it was shown in yeast cells lacking the Trf4 polyadenylation (poly(A)) polymerase, component in the TRAMP complex (Trf4-

Air2-Mtr4) (Gavaldá *et al.*, 2013) or in cells depleted of the exosome components EXOSC3 and EXOSC10 (Rrp40 and Rrp6 in yeast, respectively) (Pefanis *et al.*, 2015).

Due to the number of factors implicated, with roles in different steps from transcription to RNA processing, export or degradation, it is likely that there are some kind of distribution in the way these proteins prevent R-loop formation. In particular, since their loss does not have the same impact in the stability of the genome and gene expression in all of them, suggesting that R-loop forming as consequence of different mutations might differ.

On this regard, there is a growing interest in the role of helicases. Helicases are ubiquitous proteins implicated in the metabolism of the nucleic acids due to their capacity to unwind double strands or displace other proteins from them. Therefore, helicases play critical roles all processes where a nucleic acid is involved, such as in DNA replication, recombination, repair, transcription and translation (Sarkar & Ghosh, 2016). Particularly, the DEAD/H-box family of helicases comprehend the two largest groups in the superfamily 2 (SF2) of helicases, the DEAD-box group with 44 proteins and the DEAH-box classification with 15 members. Most of them are ATP-dependent RNA helicases, thus having a broad role in RNA metabolism, with influence in the folding of RNA and the assembly of RNPs in processes including transcription, mRNA splicing, ribosome biogenesis, RNA storage, transport, decay and translation (Rocak & Linder, 2004). The nomenclature of this group derives from the primary aminoacid sequence of the Walker B Motif/Motif II, D-E-A-D (asp-glu-ala-asp) or D-E-A-H (asp-glu-ala-his) rather than any allusion to cell death. These proteins are multifunctional and not all members of the family are *bona fide* RNA helicases, and the substrate unwinding capacity varies from one to another. A revisit of the features and characteristics of all members from this super family is far beyond the goal of this introduction, especially after such deeper compiles are already published (Putnam & Jankowsky, 2013; Cohen *et al.*, 2013; Bourgeois *et al.*, 2016), but it is worth noting that mounting evidence points towards to an increasing importance in the appearance of genome instability and cancer onset for many of these helicases (Fuller-Pace, 2013; Sarkar & Ghosh, 2016; Cai *et al.*, 2017), which is especially interesting when these links are associated to R-loop homeostasis as

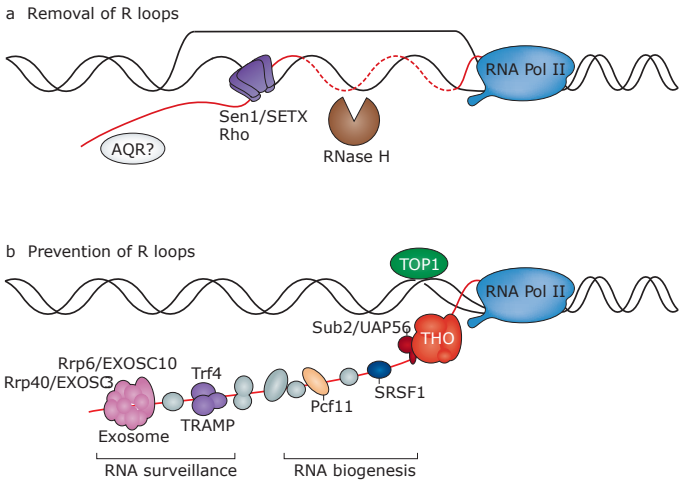


Figure I7. R-loop accumulation countermeasures.

(A) Removal of R-loops can be produced by different mechanisms. RNA:DNA hybrids structure is specifically recognized by RNAase enzymes H1 and H2, that degrade the RNA strand. Additionally, helicases such as Sen1/SETX or Rho in bacteria are able to unwind R-loops with a role in transcription termination. Other putative helicases, like the human aquarius (AQR) have been proposed to share the same capabilities. (B) Control of the R-loop formation is achieved by Topoisomerases (TOP1) that prevent supercoiling accumulation that could favor R-loop formation, and by specific RNA-binding proteins with roles in different steps of RNA metabolism from RNA biogenesis (including the THO complex, serine/arginine-rich splicing factor 1 (SRSF1) and Pcf11) to RNA surveillance (including the exoribonucleases exosome component 3 (EXOSC3) and EXOSC10 (Rrp40 and Rrp6 in yeast, respectively)). TRAMP, Trf4– Air2–Mtr4p polyadenylation complex; UAP56 (Sub2 in yeast).

in the case of DHX9 (Chakraborty & Grosse, 2011), or Dbp2, the yeast homolog of DDX5/DDX17 (Ma *et al.*, 2013).

3.3. Good R-loops (Physiological functions)

Recent evidence suggests that R-loops may have additional natural functions other than the ones already known (further reviewed in Aguilera & García-Muse, 2012). The application of genome-wide techniques allowed the discovery of RNA:DNA hybrids at promoter and terminator regions of some genes (Ginno *et al.*, 2012; Ginno *et al.*, 2013). This finding suggests a novel role for RNA:DNA hybrids in the control of RNAPII-driven transcription, likely participating in transcription activation and termination in mammals. Complementary studies support the idea that this function involve in most cases antisense RNAs or ncRNAs in association with chromatin modifications (Santos-Pereira & Aguilera, 2015). DRIP-seq analyses correlated the presence of R-loops with promoter and terminator regions enriched in CpG islands showing a strong GC skew (Ginno *et al.*, 2012; Ginno *et al.*, 2013). R-loops appeared immediately after the transcription start site, where they

protect from the action of the silencing DNA methyltransferase 3B1 (DNMT3B1). It is interesting that chromatin marks associated with transcription initiation and/or elongation, like Histone 3 lysine 4 trimethylation (H3K4me3) and H4K20me1 and H3K79me2, were also found at these CpG islands (Ginno *et al.*, 2012; Ginno *et al.*, 2013), supporting the notion that R-loops function in transcription regulation by affecting the chromatin structure in their surrounds. Additionally, R-loop-dependent regulation of mRNA transcription seems to be mediated also by ncRNAs, at least in some genes, like the human vimentin (VIM) gene, in which case the formation of an R-loop at the promoter region produced by an antisense RNA activates sense transcription by enhancing chromatin opening, as inferred from a decreased nucleosome occupancy (Boque-Sastre *et al.*, 2015).

Data indicate that the enrichment of R-loops at certain loci due to a positive GC skew also applies to terminator regions (Ginno *et al.*, 2012), suggesting a similar mechanism of protection from DNMTs to that of promoters. Regulation by ncRNAs has also been observed for transcription termination together with chromatin modifications. Such is the case of the human β-actin (ACTB) gene, for which R-loop formation at its G-rich termination pause site has been proposed to induce antisense transcription. As a consequence, the recruitment of the RNA interference (RNAi) silencing machinery leads to the formation of repressive heterochromatic sites by H3K9me2 deposition and heterochromatic protein 1γ (HP1γ) recruitment (Skourti-Stathaki *et al.*, 2014). However, there is also functional evidence that R-loops may have a more active role in transcription termination, thanks to a pausing mechanism that relies on G-rich sequences (termination pause sites) located after the poly(A) site, at which nascent mRNAs are cleaved by the exoribonucleases Rat1/XRN2 (Proudfoot, 2011). These nucleases are known to interact with Sen1/SETX, respectively (Kawauchi *et al.*, 2008; Skourti-Stathaki *et al.*, 2011), whose absence is demonstrated to increase the number of R-loops both in yeast and human cells (Mischo *et al.*, 2011). These observations support that R-loops are required for RNAPII pausing, and Sen1/SETX would help to release the RNA molecule by promoting an efficient transcription termination through the action of the nucleases.

3.4. Bad R-loops (Genome instability triggers)

R-loops are well known for their strong capacity to modulate genome structure. They are a source of DNA damage and replication stress, thus R-loops have an important impact on genome instability and genetic variation. Due to the fact that ssDNA is more mutagenic that dsDNA because it is more accessible to DNA damaging agents (Lindahl, 1993), a deregulated increase of R-loop accumulation would likely lead to a hypermutation phenotype. ssDNA is target for the action of specific mutagenic agents or enzymes, such as human activation-induced cytidine deaminase (AID), which is involved in immunoglobulin class switching recombination and somatic hypermutation in mammalian activated B cells (Chaudhuri & Alt, 2004). Indeed, when expressing AID in R-loop prone yeast THO mutants, an increase in the number of mutations in transcribed genes is observed (Gómez-González & Aguilera, 2007). Additionally, RNA:DNA hybrids could promote mutagenic DNA replication through their use as primers to initiate non-canonical replication, as it has been described in E. coli (Kogoma, 1997) or in the rDNA region of yeast lacking Top1 and Rnase H enzymes (Stuckey *et al.*, 2015). However, R-loop mediated genomic instability is manifested mainly as a form of transcription-associated recombination (TAR) that arises from DNA breaks (Aguilera, 2002).

Little is still known about how R-loops lead to ssDNA gaps or DSBs. From bacteria and yeasts to human cells it has been shown that R-loop accumulation owing to losses of different factors lead to replication stress and replication fork stalling (Wellinger *et al.*, 2006b; Boubakri *et al.*, 2010; Gan *et al.*, 2011; Gómez-González *et al.*, 2011; Santos-Pereira *et al.*, 2013; Alzu *et al.*, 2012) which are major sources of DNA breaks, recombination and genomic rearrangements (Zeman & Cimprich, 2013). Although R-loop accumulation itself may be an important source of replication stress, we cannot rule out yet the possibility that there are other structural features that may contribute to RF stalling. For example, it cannot be excluded that arrested RNA polymerases remain attached and constitute the real barrier for the replisome progression.

Different hypotheses have been proposed to explain the mechanisms

underlying the formation of DNA breaks generated by R-loop deregulation. For example, it has been reported that R-loops are processed by nucleotide-excision repair (NER) nucleases, leading to the production of DSBs (Sollier *et al.*, 2014) or ssDNA gaps that could be converted in double-strand breaks after replication. Indeed, it was shown that formation of R-loops at damaged sites of transcription block triggers a non-canonical activation of ATM DNA damage response pathway (Tresini *et al.*, 2015), supporting the formation of DSBs in presence of R-loops. More work is needed to uncover the mechanisms by which R-loops lead to the accumulation of DSBs, but the relationship between their uncontrolled accumulation and the activation of cell cycle checkpoints due to replicative stress is well proven.

Telomere maintenance is critical for cells to avoid premature senescence and ageing. Physiologically, telomerase is the protein in charge of ensuring proper telomere length, but it has been reported that in the absence of telomerase, R-loop prone mutants lacking RNase H or THO show telomere lengthening and delayed senescence in an HR-dependent manner (Balk *et al.*, 2013; Pfeiffer *et al.*, 2013; Yu *et al.*, 2014). This phenomenon is due to the presence of telomeric-repeat-containing-RNAs (TERRAS) that are transcribed at the telomeres (Luke *et al.*, 2008). Therefore, telomeres represent specific loci where cells take advantage of the capacity of R-loops to promote recombination, but this also needs some kind of regulation, since it has been shown that cancer cells with telomerase-independent recombination-mediated alternative lengthening of telomeres, RNA:DNA hybrids generated at TERRA loci accumulate, and RNase H1 overexpression reduces their rates of recombination (Arora *et al.*, 2014).

Another feature of R-loops is their capacity to influence and induce chromatin changes. Recent reports indicate that chromatin contributes to R-loop-mediated genome instability, and that their presence is linked to heterochromatin and chromatin condensation marks in yeast, *C. elegans* and human cells (Castellano-Pozo *et al.*, 2013). Lack of THO complex or SETX induces an accumulation of H3S10P in yeast an human cells, a mark associated with condensed chromosomes during mitosis and also with transcription activation (Castellano-Pozo *et al.*, 2013). Remarkably, it is known that replisomes do not easily progress through condensed chromatin (El

Achkar *et al.*, 2005; Castellano-Pozo *et al.*, 2012; Castellano-Pozo *et al.*, 2013), thus this adds to the hypothesis about how R-loops trigger replication stress. Moreover, Friedreich ataxia (FRDA) and Fragile X chromosome (FXS) arise as consequence of repeat expansions taking place in the frataxin (FXN) and fragile X mental retardation (FMR1) genes, respectively. These expansions are described to accumulate R-loops, to be associated with H3K9me2 deposition, which marks gene silencing, and to be prone to replication-dependent fragility (Groh *et al.*, 2014; Skourti-Stathaki *et al.*, 2014; Castellano-Pozo *et al.*, 2013). Another supporting relationship is stablished by the requirement of the FACT complex for correct DNA replication in regions that are being replicated, inasmuch as yeast and human FACT-depleted cells accumulate R-loops and show genomic stability phenotypes (Herrera-Moyano *et al.*, 2014). However, the formation of a RNA:DNA hybrid is associated with nucleosome-free regions and thus to an open state of the chromatin (Boque-Sastre *et al.*, 2015). It is not clear why R-loops are detected associated with condensed chromatin in some cases and with open chromatin in others, but probably they represent two different stages of their formation.

3.5. R-loops: The Ugly (Impact on human health)

Recent research has brought to light the broad presence of R-loops over the genome, and the variety of impacts that their accumulation can inflict over the cells surveillance. Therefore, many links have been uncovered between R-loop prone mutations and the development of different syndromes and human diseases associated with replication stress and genome instability. As for example, expansion of trinucleotide repeats at fragile sites are associated to the formation of R-loops, which lead to decreased expression of certain genes in neurodegenerative disorders like Friedreich ataxia or the Fragile X syndrome (Groh *et al.*, 2014). Additionally, mutations R-loop removal factors like the three subunits of RNase H2 have been linked to the neurological auto-immune disorder, Aicardi-Goutières syndrome (AGS) (Crow *et al.*, 2006); or mutations in SETX have been associated with juvenile amyotrophic lateral sclerosis type 4 (ALS4) (Chen *et al.*, 2004) and ataxia with oculomotor apraxia type 2 (AOA2) (Moreira *et al.*, 2004).

Besides, R-loops lead to genome instability and replication stress, which are hallmarks of pre-tumoral and tumoral cells (Gaillard *et al.*, 2015), so R-loops could be potential drivers of cancer. Such connection is already supported by the results that link deficiencies of BRCA1 and BRCA2 to increased R-loop accumulation and DSBs that are partially reduced by RNase H1 (Bhatia *et al.*, 2014). Additional connections have been stablished over the time, with many factors already known to lead to tumorigenesis now being associated with an increased presence of R-loops, which are deeply reviewed in (Santos-Pereira & Aguilera, 2015).

Objectives

Objectives

The main goal of this thesis is to advance in the knowledge about how R-loops are formed, what are their functions and how their presence is regulated to maintain the stability of the genome. For this purpose, we addressed the following specific objectives:

1. To analyze where R-loops are formed through the genome, in a wild-type strain and a *hpr1*Δ mutant, and study the differences.
2. To explore the possible formation of non-cotranscriptionally formed R-loops (R-loops in trans).
3. To search for new factors implicated in R-loop homeostasis maintenance.
4. To get new insights into the mechanisms of coordination between transcription and replication by the human helicase RECQL5, through its expression in the model organism *S. cerevisiae*.

Results I

1. R-loops and NGS

The study of RNA:DNA hybrids distribution across the whole genome suppose a powerful approach to help shed some light on how these structures are generated in physiological or pathological environments and why their de-regulated presence is dangerous for the cell. Up to date, a number of similar studies have already been performed with more or less solidness in their results.

Ginno *et al.* (2012) obtained a first genome-wide distribution of *in vivo* formed RNA:DNA hybrids by immunoprecipitation with the S9.6 antibody and subsequent DNA sequencing in mammal cells. The method used, called DRIP-seq (DNA-RNA hybrid immunoprecipitation followed by sequencing), allowed for the creation of the first map of RNA:DNA hybrid-prone forming regions in the human genome *in vivo*. Similar approaches have been accomplished over the years taking advantage of the model organism, *Saccharomyces cerevisiae*. Two preliminary studies (El Hage *et al.*, 2014 and Chan *et al.*, 2014) undertook the same task employing different techniques to analyse the complete genome, the first through chromatin immunoprecipitation and sequencing (ChIP-seq) and the second one by hybridization using tilling arrays (DRIP-chip). The conclusions dragged from these works showed only small differences and agreed with previous single-gene analyses, indicating that R-loops preferentially accumulated at highly transcribed genes and/or at high GC DNA regions and a major presence at the rDNA locus and retrotransposons. Recently, Wahba *et al.* (2016) improved the DRIP assay by adding a prior treatment with S1 nuclease that degrades the ssDNA displaced in the formation of the RNA:DNA hybrid. This step allows for the sonication of the samples without major disruption of the hybrids, leading to an important background reduction from the sequencing signal. Their results in wild-type strain yeasts stated that the work from Chan *et al. (2014)* overestimated positive regions and the study by el Hage *et al.* (2014) fell too short in their selection. However, a subset of the regions identified by Wahba *et al.* (2016) corroborated a number of findings from those previous DRIP-chip and/or ChIP-seq studies, including the occurrence of hybrids at the mitochondria, rDNA locus, telomeres, Ty elements, and tRNA genes. The new method allows a better

detection of the differences. It is, thus, clear that the incorporation of new steps in the DRIP processing is important to get an accurate estimate of R-loop formation through the genome at different conditions.

However, a concern with the new technique based on S1 nuclease is that a breakage in DNA may facilitate hybrids at regions where they do not spontaneously form. Further analysis with different techniques, likely not relying in the S9.6 antibody, might be necessary in the future.

1.1. RNA:DNA hybrid mapping in wild-type and *hpr1*Δ mutant yeasts

Having our laboratory identified the first factor involved in the regulation of R-loop presence, we decided to incorporate the genome-wide approach of R-loop accumulation to our studies. Experiments were performed applying a protocol similar to that performed in human cells (Ginno *et al.*, 2012) rather than those used in yeast (Chan *et al.*, 2014; El Hage *et al.*, 2014; Wahba *et al.*, 2016). The technical details of the procedures are covered in the Material and Methods' chapter. Briefly, we performed a very delicate chromatin extraction followed by digestion with a cocktail of endonucleases for the fragmentation of DNA. Afterward, samples were split and whether or not treated with RNase H before being subjected to immunoprecipitation with the S9.6 antibody. RNA:DNA hybrids were purified and amplified with a Whole Genome Amplification Kit before sending to the Genomics Laboratory for the sequencing with an Ion Torrent® platform. Inputs were used as controls.

We performed a first experiment using a S9.6 antibody purified in-house from the original hybridoma (HB-8730TM, ATCC®) to immunoprecipitate R-loops formed in wild-type and *hpr1*Δ mutant strains with a W303 background. Sequencing results were quality checked, pre-processed and aligned to the *S. cerevisiae* sacCer3 reference genome build using BOWTIE. Generated BAM files were further analysed with a number of tools including MACS2 for the peak calling of positive regions and deepTools for the generation of profiles, among others. We used the data previously published from GSE53420 (El Hage *et al.*, 2014), E-MTAB-2388

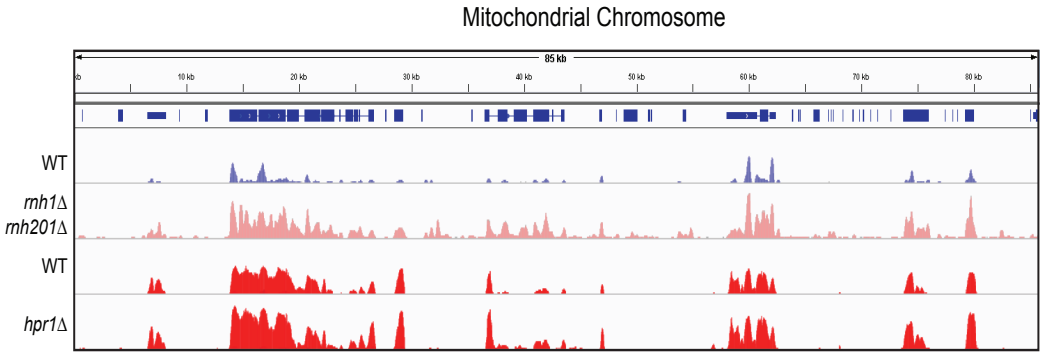


Figure R1. S9.6 signal profile over mitochondrial chromosome.
Coverage profiles over the mitochondrial chromosome. Wild-type and double rnhΔ strains data from GSE53420 (El Hage et al., 2014). Wild-type and hpr1Δ mutant (DRIP A) performed with in-house S9.6 antibodies, respectively.

(Chan *et al.*, 2014) and SRP071346 (Wahba *et al.*, 2016) to compare our results and estimate the precision of our methodology.

After generation of the coverage profiles created with the alignment of the sequencing reads, we started the evaluation of our results by comparing with available datasets. The only study that provided with data in a format that we could compare in this way was El Hage *et al.*, (2014). We loaded their read coverage files (bigwig format) directly into IGV (Integrative Genomics Viewer v2.3.48, Broad Institute) along with ours, and used the diagram generated at the mitochondrial DNA of the yeast as reference to determine the efficiency of RNA:DNA hybrid immunoprecipitations. R-loops are necessary for mtDNA replication (Baldacci *et al.*, 1984; Reyes *et al.*, 2013), therefore the pattern of IP obtained in this region could be used as control. As can be seen in **Figure R1** the pattern obtained was quite similar to the reference

Next we evaluated the overall performance of the experiment by analysing the sensitivity of hybrid-forming regions to RNase H treatment. For this, we normalized our sequencing results employing the same methodology and scripts used by El Hage *et al.*, (2014). Briefly, we measured the read depth in all intergenic regions (not annotated regions) according to the most updated version of sacCer3 table of features at SGD. The mean value of this measurement was used to normalize the number of reads in each sample, so they could be comparable. **Figure R2** collects the profiles for INPUT samples (blue lines) and S9.6 IP samples pre-treated with

RNase H (green lines) or not (red lines). Although RNH-treated samples show an overall lower coverage, this decrease was not as clear as expected. The analysis of the immunoprecipitation results taking into account the RNH treated samples for the identification of hybrid-positive peaks produced a dramatic reduction in the detection of R-loop-prone loci in comparison with the values previously published, suggesting that the RNase H treatment had not been efficient. Although the treatment with RNase H performed was the same as that used in DRIP-qPCR experiments, conditions in which we have already proven it to work correctly, it was not sufficient, likely due to artefacts introduced by the additional amplification step required for these kind of IP-seq protocols. Consequently, we continued the analysis without including the RNH-treated samples. We realize that, given a low affinity of the S9.6 antibody for dsRNA, our analysis should still give us reliable relative values of RNA:DNA hybrids distribution between the wild-type and the mutant strains. To be ascertain of this, we validated our data with those previously published in WT yeast cells.

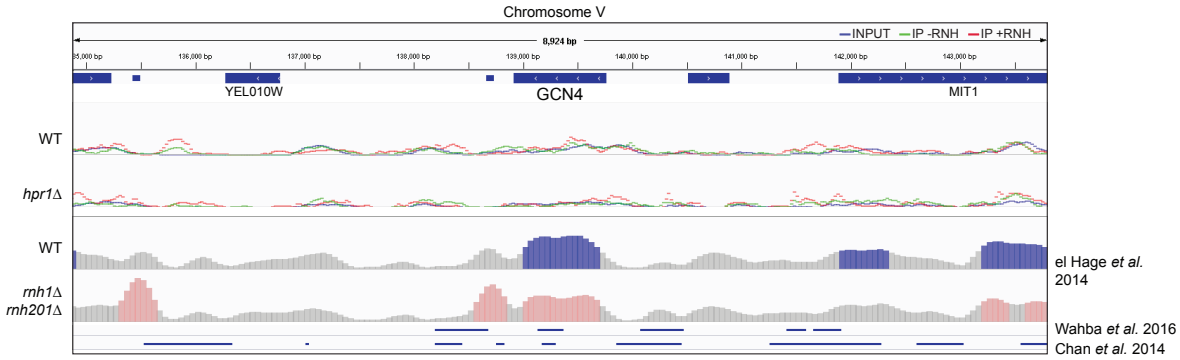


Figure R2. RNase H treatment for control.

Snapshot of DRIP-seq reads on GCN4 locus in chromosome V. Schematic representation of GCN4 locus and surrounds are depicted on top rows. First two tracks correspond to coverage data for WT and hpr1Δ S9.6 IP of RNH pre-treated sample (green line) or not (red line) and INPUT (blue line). Third and fourth track represent data as can be obtained from GEO accession for (El Hage et al., 2014); blue (WT) and pink (rnh1Δrnh201Δ mutant) regions were MACS2-predicted as hybrid positive regions. Fifth and sixth interval tracks represent hybrid accumulating regions in wild-type, identified by Wahba et al., 2016 and Chan et al., 2014, respectively

To identify hybrid-prone regions from our data, we performed statistical analyses with the software MACS2 v2.1.0. Peak calling was performed by taking our INPUT samples as control and IP samples as treatment. The software fractionates the genome in defined bins or windows, and evaluates through a Poisson distribution the significance of the signal enrichment in the treatment over the control samples. Using this method, we defined a number of regions **(Table R1)** with a significative RNA:DNA hybrid presence over the INPUT with a p-value < 0.05.

Table R1. RNA:DNA hybrid-positive features in wild-type strains

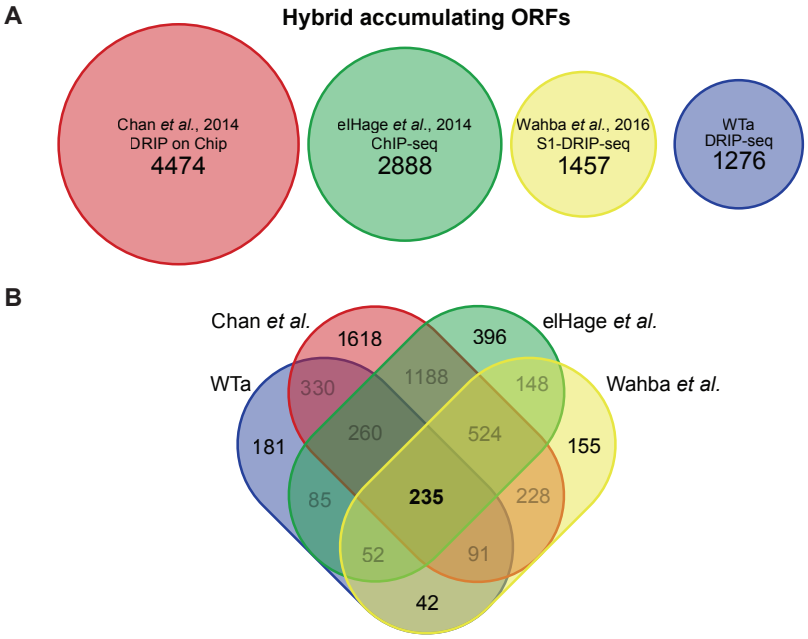
Colour gradient code range from white (0%) to red (100%). Percent represents proportion of positive hits respect the respective total number for each feature in the genome.

	Genome	WTa	el Hage *et al.* 2014	Chan *et al.* 2014	Wahba *et al.* 2016
ORF	6604	1276 (19.3%)	2888 (43.7%)	4474 (67.7%)	1475 (22.3%)
Centromere	16	0 (.%)	0 (.%)	0 (.%)	0 (.%)
Telomere	32	17 (53.1%)	31 (96.9%)	32 (100.%)	21 (65.6%)
ARS	352	43 (12.2%)	25 (7.1%)	147 (41.8%)	49 (13.9%)
Intron	377	41 (10.9%)	60 (15.9%)	135 (35.8%)	39 (10.3%)
sn/snoRNA gene	83	28 (33.7%)	12 (14.5%)	26 (31.3%)	54 (65.1%)
tRNA gene	299	91 (30.4%)	72 (24.1%)	171 (57.2%)	54 (18.1%)
rRNA gene	27	4 (14.8%)	7 (25.9%)	25 (92.6%)	4 (14.8%)
ncRNA gene	18	4 (22.2%)	7 (38.9%)	11 (61.1%)	4 (22.2%)
Transposon	91	83 (91.2%)	65 (71.4%)	91 (100.%)	1 (1.1%)
Total Hits	16455	3462 (21.%)	6455 (39.2%)	10651 (64.7%)	3472 (21.1%)

Through comparison of the number of positive regions obtained and the mapping to known features, we observed that our experiment was more similar to those of Wahba *et al.* (2016). Both experiments identified a similar number of regions, and the distribution of hits is quite similar except for transposons and sn/snoRNA genes. Due to the high presence of repetitive sequences within transposons, the change may be derived from differences in the processing of the sequencing results. **Table R1** shows that Chan *et al.* (2014) study probably fell into overestimation of positive regions. It also noticeable that there are no centromeres represented in any of the analyses using wild-type strains, and the telomeres are vastly present among positive regions.

It is widely accepted that the main source of R-loops in the cell is as co-

transcriptional by-products. Therefore, we focused our following analysis in the characteristics of the ORFs identified as hybrid-prone genes. For this, we compared the ORFs identified in all four studies and obtained a list of 235 common genes **(Figure R3B)**, from which we subtracted 13 genes present in a list of 108 genes characterized to be immunoprecipitated with high probability (Teytelman *et al.*, 2013). As can be seen in **Figure R3C**, genes identified to accumulate R-loops are long, rich in G+C and are highly expressed compared to the median of the these values for the genome. These results are consistent with the conclusion that our reference studies established for these three parameters.



C

ORFs features	Genome	Hybrid Positive	p-value
Length (bp)	1071	1755	<0.0001
G+C content (%)	39.7	40.6	<0.0001
Expression levels	0.8	1.2	<0.0001

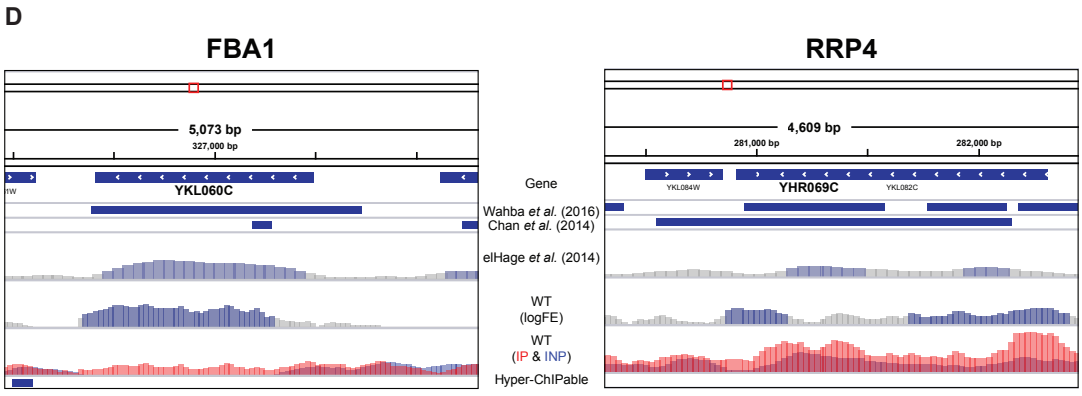


Figure R3. Analysis of RNA:DNA hybrid positive ORFs in WT

(A) Proportional representation of the number of hybrid-positive ORFs from each study. (B) Venn diagrams sowing the overlap between the RNA:DNA hybrid positive-ORFs between the wild-types in our experiment and the ones used as references. (C) Statistical analysis of length, G+C content and expression values of hybrid-positive genes (Mann-Whitney U test) (D) Snapshot of DRIP-seq reads on FBA1 and RRP4 locus in chromosome XIV. Schematic representation of genes loci and surrounds are depicted on top rows. Tracks for WT coming from reference studies by Wahba et al., 2016, Chan et al., 2014 (El Hage et al., 2014). Last two histograms from our WT (WRBb.9D) DRIP-seq. First histogram shows MACS2-calculated hybrid-accumulating positive regions and second track depicts raw sequencing coverage for S9.6 IP (red) and input (blue).

For the visualization of these results, we elaborated a new track that collects the logFE (Fold Enrichment in logarithmic scale) from IP signal over the INPUT signal. This track shows all the regions over the genome in which IP signal is at least one time higher than the signal registered in the INPUT sample. With this track we can observe only the regions in which the S9.6 is greater than the background, and marked the previously identified regions that are significant with colour over the grey background **(Figure R3D)**.

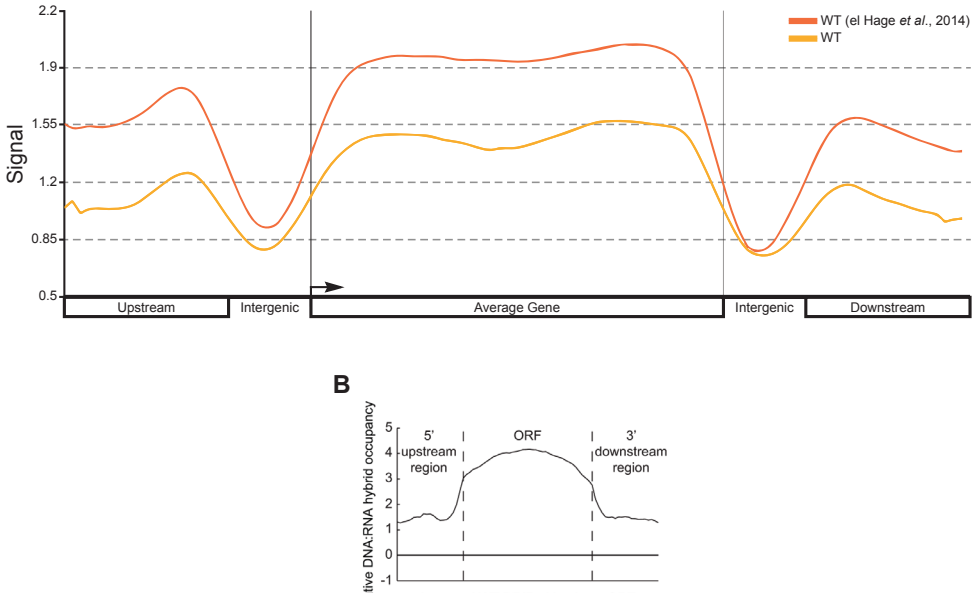


Figure R4. RNA:DNA hybrid distribution over genes in WT

(A) Average distribution over the 235 genes identified as prone to accumulate RNA:DNA hybrids and their adjacent 1kb regions are represented. Orange line corresponds to data obtained from from GSE53420 (El Hage et al., 2014), yellow line to DRIP-seq of WT strain (WRBb.9D). (B) Average gene profile of RNA:DNA hybrids at ORFs enriched for RNA:DNA hybrids in wild type, adapted from Chan et al. 2014.

Finally, we analysed the relative distribution of RNA:DNA hybrids along the length of the ORFs as explained above. All hybrid-positive gene coding regions and surrounding ±1 kb were subdivided in equal number of fragments, for each segment, the median of the signal (as integer calculated with the MACS2 peak caller) was represented. **Figure R4** shows a similar pattern of distribution for our wild-type and the one published by el Hage *et al.* (2014), in which R-loops are present along the length of the ORFs with a slightly higher accumulation towards the 3' end of the gene. The dissimilarity in signal values in this case is probably due to differences either in the efficiency of the experiment or the processing of the data before the point of this analysis. The data files available in the online databases from Chan *et al.*, (2014) and Wahba *et al.*, (2016) were not compatible with the software that we used for the representation in **Figure R4**. Therefore, comparing our result to those shown in the publications (taken directly from the papers's figures, meaning that their profiles are calculated with different set of RNA:DNA-hybrid positive genes), the distribution found by Chan *et al.* (2014) is more skewed towards an accumulation in the middle region of the gene. The data were not comparable with similar analysis from Wahba *et al.* (2016) since they focused their attention in the distribution of hybrids accumulated in the double mutant *rnh1*Δ *rnh201*Δ and did not show this analysis for their wild-type.

Since our data for the wild-type seemed to fit between the variability of the analyses already published, we also considered interesting to examine the results obtained for the *hpr1*Δ mutant strain, in order to get some insights.

General inspection of the tracks showed a close correlation between the positive regions identified in the WT (dark blue histogram) and the *hpr1*Δ strain (yellow histogram), with some extra clusters or more intense ones in the mutant, as can be observed in the example presented in **Figure R5**. To manage a more broad overview, we used the R-package *ChIPseeker* (Yu *et al.*, 2015), with this software we could elaborate a representation of the distribution of RNA:DNA hybrid positive regions over the *S. cerevisiae* chromosomes in our WT and mutant strains. As can be observed in **Figure R6**, RNA:DNA hybrids are distributed all over the genome,

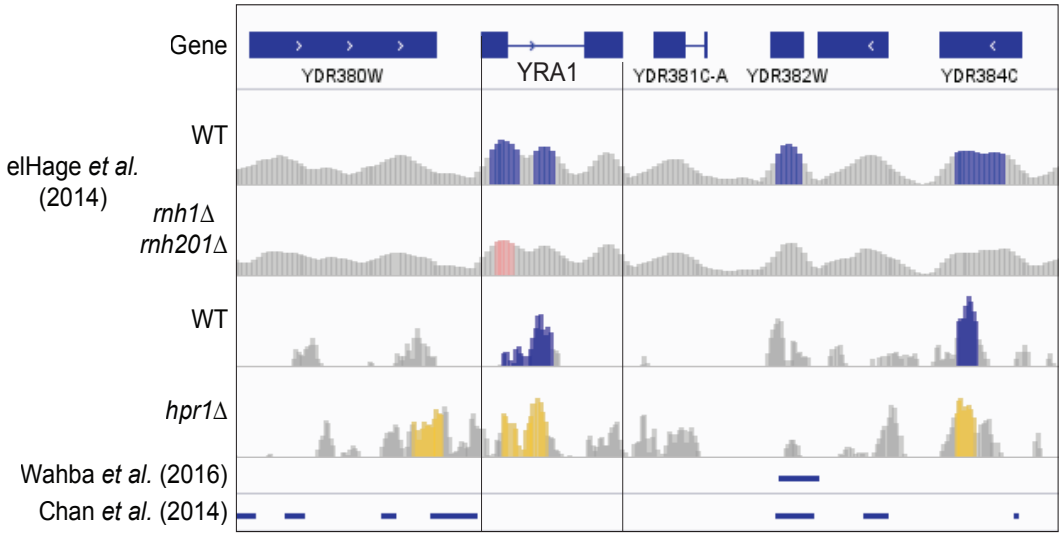


Figure R5. Distribution of hybrid positive regions at YRA1 locus.

Snapshot of DRIP-seq profile over the YRA1 locus in chromosome IV. Schematic representation of YRA1 locus and surrounds are depicted on top rows. First two tracks correspond to coverage data as can be obtained from GEO accession for El Hage et al. (2014); WT (blue) and rnh1Δrmh201Δ mutant (pink) regions were MACS2-predicted as hybrid positive regions for a fold enrichment (FE) of 2 over the background signal. Tracks WT (WRBb.9D –dark blue-) and hpr1Δ (HRBb.12 –yellow-) represent logFE enrichment of at least 1 of S9.6 signal over input (grey). Coloured regions correspond to intervals with a significative enrichment over input signal $p<0.05$. Last two tracks of S9.6 positive regions obtained from specified publications.

appearing to be more widely distributed in the *hpr1*Δ mutant although the overall signal (measured as relative count of reads) did not seem to be higher compared to the WT. This result is consistent with the only other published work in which *hpr1*Δ mutant was analysed, were it was shown that distribution signal was generally lower, but still, significantly high and had more genome coverage.

Next, we mapped the positive regions identified to the features in the *S. cerevisiae* genome **(Figure R7)** Respect to the total number of positive intervals (hits), ORFs represented approximately 40% of them in both strains (36.9% in WT and 40.3% in *hpr1*Δ), the second and third features most enriched were tRNAs (2.6% and 1.8%) and transposons (2.4% and 1.8%). Relative to the total number of features present in the genome for each category, as can be seen in **Table R2**, transposons and telomeres are genomic regions clearly prone to the formation of RNA:DNA hybrids in both strains. Another interesting difference is the dramatic reduction in the formation of hybrids over sn/snoRNAs and tRNAs in the *hpr1*Δ mutant. Again,

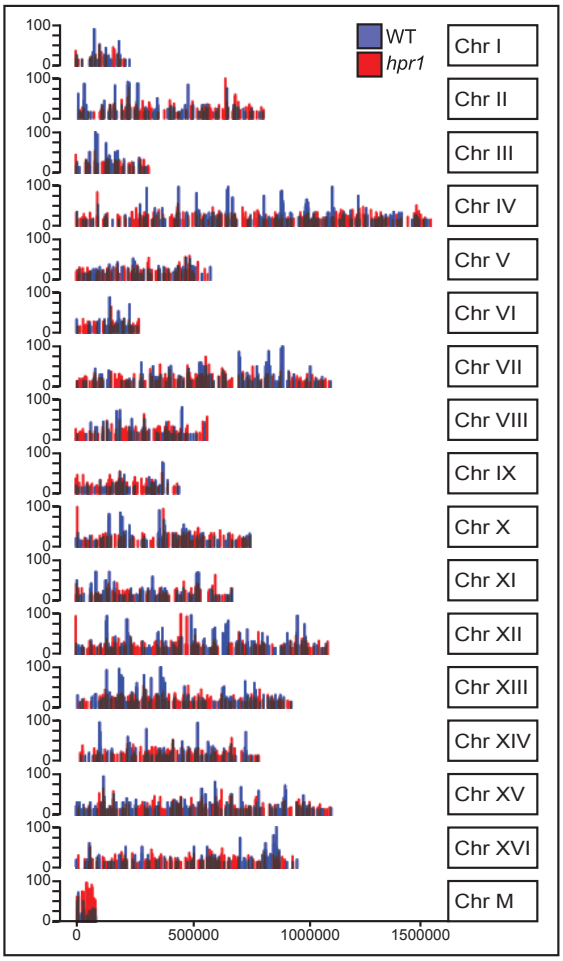


Figure R6. Distribution of RNA:DNA hybrids across S. Cerevisiae genome.
Genomic view of RNA:DNA hybrids in WT(WRBb.9D) -blue- and *hpr1*Δ (HRBb.12) –red- cells are shown. Representation of chromosomes with the value corresponding to Integer Score provided in the broadpeak file as result of peak calling using MACS2. The X-axis corresponds to coordinates in bp.

it was previously published that from the common RNA:DNA hybrid accumulating mutants (*hpr1*Δ, *rnh1*Δ *rnh201*Δ, *sen1-1*) the HPR1 deficient mutant presents the lowest signal values, specifically shown for these two types of non-coding RNAs in the study by Chan *et al.* (2014). Focusing in the analysis of ORFs, about 22% of the genes identified to accumulate hybrids did so both in wild-type and in *hpr1*Δ strains. Deletion of the *HPR1* gene leaded to the formation of hybrids in genes with different features to those accumulating spontaneous R-Loops in the wildtype. The main characteristics **(Figure R7C)** being that they were longer, presented a slight but significative reduction in their G+C content, and their expression levels according to published data (Beyer *et al.*, 2004) were significantly lower compared to the median for the wild-type, but had not statistical difference compared to the median of the genome.

Metagene analysis of all 457 hybrid-prone ORFs common to both wild-type and *hpr1*Δ strains indicated that hybrids distribution was similar independent of

the transcription factor deletion. The signal clusters spreaded along the length of transcription units with a slight peak in signal at the 5' end and a little higher density towards the 3' end of the genes **(Figure R8)**.

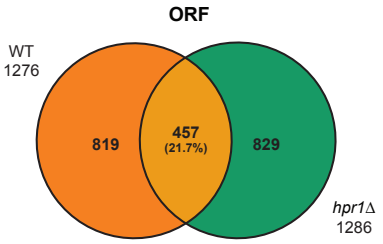
Table R2. Features mapped in WT and hpr1

The usefulness of DRIP profiling is widely demonstrated, and uncovers very interesting perspectives when the results are backcrossed with the huge quantity

A

	Genome	WT	*hpr1*Δ
ORF	6604	1276 (19.3%)	1286 (19.5%)
Centromere	16	0 (.%)	1 (6.3%)
Telomere	32	17 (53.1%)	17 (53.1%)
ARS	352	43 (12.2%)	43 (12.2%)
Intron	377	41 (10.9%)	36 (9.5%)
sn/snoRNA gene	83	28 (33.7%)	5 (6.%)
tRNA gene	299	91 (30.4%)	58 (19.4%)
rRNA gene	27	4 (14.8%)	4 (14.8%)
ncRNA gene	18	4 (22.2%)	3 (16.7%)
Transposon	91	83 (91.2%)	59 (64.8%)
Total Hits	16455	3462 (21.%)	3191 (19.4%)

B



C

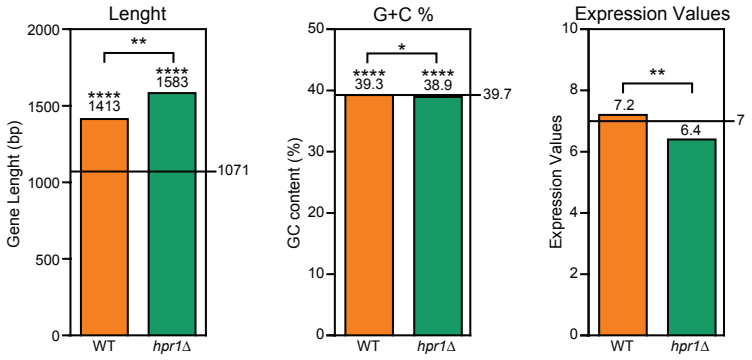


Figure R7. Analysis of RNA:DNA hybrid positive features in WT and *hpr1*Δ.

(A) Table R2. RNA:DNA hybrids positive features in WT (WRBb.9D) and hpr1Δ (HRBb.12) strains. Numbers represent the number of hits in each category. Percentages calculated respect to the total number of features in the genome for each category (B) Venn diagram showing the overlap between the WT and hpr1Δ positive genes (C) Statistical analysis of Length, G+C content and model-based expression values for genes mapped to be R-loop prone in WT and hpr1Δ strains. Median values are represented. Black line for the whole-genome median. *, P<0.05, **P<0.001, ****P<0.0001(Mann-Whitney's U-test)

readily available about transcription and replication in the yeast and mutant strains. Therefore, although due to time limitations, this will be a future project of research ahead of this thesis.

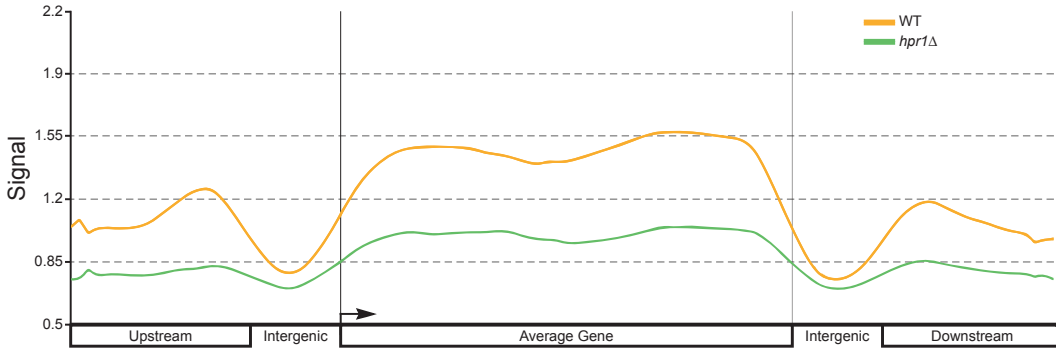


Figure R8. Signal distribution over hybrid-positive ORFs in wild-type and *hpr1*Δ.

Metagene that compiles the signal median over all the hybrid-prone ORFs common to the WT (WRBb.9b) –orange line- and hpr1Δ (HRBb.12) –green line- strains.

2. *trans* RNA as source of genomic instability

The fact that the nascent RNA is able to induce hyperrecombination during the transcriptional process in absence of a fully functional THO complex has already been well stablished (Chávez & Aguilera, 1997; Huertas & Aguilera, 2003). It is also known that the deficiency of this complex and other RNA related factors also lead to the accumulation of mRNAs in the nucleus (Strässer *et al.*, 2002; Jimeno *et al.*, 2002; Paul & Montpetit, 2016). This supports that the co-transcriptional binding of many RNA processing and transcription factors is important to prevent hybrid formation by restricting the access of nascent RNA molecules to the DNA template at the site of transcription (Aguilera & García-Muse, 2012). However, it has been proven that these RNA biogenesis factors are not sufficient to prevent transient hybrid formation at some loci in wild-type budding yeast; rather, hybrids may form naturally but are removed by factors like the two endogenous RNase H enzymes and presumably Sen1p, an RNA:DNA helicase (Mischo *et al.*, 2011; Wahba *et al.*, 2011). Some recent studies have also proposed that mRNA molecules are not only susceptible of forming a DNA:RNA hybrid during the transcription process but also independent of it and even at different sites to where they were originally being produced, that is *in trans* (Wahba *et al.*, 2013; Keskin *et al.*, 2014). Therefrom, we wanted to further investigate the capability of RNA molecules to induce hyperrecombination *in trans*. In detail, our goal was to know if transcription of a DNA fragment is able to induce recombination of another homologous fragment not being actively transcribed at that moment.

In order to achieve this, we devised a strategy based on the use of different inducible promoters to control the bacterial *LacZ* gene expression in *S. cerevisiae* cells. We initially cotransformed different yeast mutants with two plasmids **(Figure R9)**:

(a) The plasmid pCM190 bearing different sequences of the *LacZ* gene under the control of a *tet* promoter, which could be repressed by addition of doxycycline to the media. This transcriptional fusion would be the source of the mRNA in *trans*.

(b) The plasmid pRS314-GL-*LacZ,* which carried the GL-*LacZ* recombination system (González-Barrera *et al.*, 2002). This construct consisted in the sequence of the *LacZ* gene integrated between two truncated fragments of the *LEU2* gene, which share 200bp of homology. This transcriptional fusion is controlled by a *GAL1* promoter, activated by addition of galactose to the media or completely repressed in presence of glucose. Active transcription of this plasmid would produce mRNA *in cis*.

With these plasmids, depending on the simultaneous state of both promoters, we could work with a total of four possible conditions to study the impact of RNA from two different sources on the recombination frequency of our system, measured as the frequency of Leu$^+$ colonies.

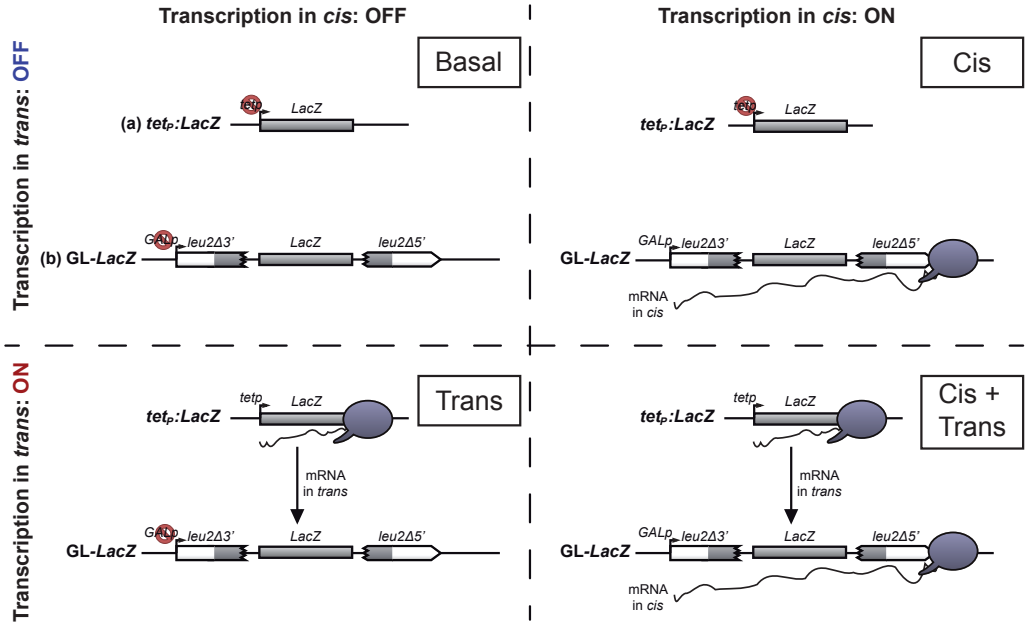


Figure R9. Schematic representation of the assay.

Schematic representation of the two mRNA sources used in the experiments and the four combinations of the promoters' activation states.

2.1. Impact of RNA in *trans* over the recombination frequency

We started studying the capability of an mRNA molecule to induce recombination at a different homologous target in a physiological environment. For this, we cotransformed the wild-type strain W303.1A with the recombination plasmid pRS314-GL-*LacZ* and the plasmid pCM190 carrying the gene *LacZ* or not. Recombination frequencies for the GL-LacZ system in each of the four conditions tested are shown in **Figure R10**. When no transcript was generated, recombination levels were as expected for a wild-type strain. Generation of an in *cis* mRNA induced a significant increase of almost 3-fold in the recombination frequency, connected to the transcription-associated recombination (TAR) (Prado & Aguilera, 2005). However, when transcription of the *LacZ* gene was active *in trans* (**Figure R10B**, red bars) no significant difference could be observed in either condition, whether transcription *in cis* was active or not. These results suggested that recombination of a sequence induced by a homologous mRNA transcribed from a different locus do not represent a detectable source of genetic instability as measured by means of our system in wild-type cells.

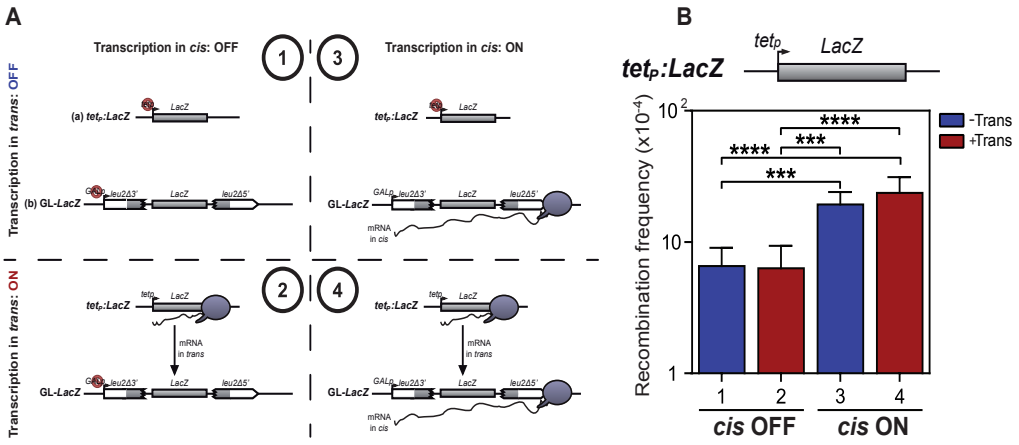


Figure R10. Analysis of the effect of RNA in trans on genetic recombination (I)

(A) Schematic representation of the two mRNA sources used in the experiments and the four combinations of the promoters activation states. (1) cis OFF/ trans OFF (2) cis OFF/ trans ON (3) cis ON/ trans OFF (4) cis ON/ trans ON (B) Recombination analysis in WT (W303) strain carrying GL-LacZ plasmid system (pRS314-GL-LacZ) plus either the empty vector pCM190 (- Trans mRNA) or the same vector (+ Trans mRNA) carrying the LacZ gene. Average and SD of at least three independent experiments consisting in the median value of six independent colonies each are shown. ***, p≤0.001; ****, p≤0.0001. (Student's t-test). A scheme of LacZ transcriptional fusion used in this experiment is shown on top.

For the formation of a hybrid between DNA and RNA synthetized in a different location, a spontaneous annealing may be required to take place. Transcription is a tightly regulated process, in which the constitution of the ribonucleoparticle (mRNP) is essential to ensure the stability and export of the nascent mRNA. Therefore, we wondered if the accumulation of unprotected RNA molecules like those generated in THO mutants (Luna *et al.*, 2012) could develop in an increase of spontaneous recombination events *in trans.*

For this, we cotransformed a W303 wild-type strain and the congenic *mft1*Δ and *hpr1*Δ mutants. Both proteins are components of the THO complex, and their absence cause phenotypes related to the formation and accumulation of DNA:RNA hybrids, but with the first mutant presenting more moderate effects than the latter. Recombination frequencies obtained are shown in **Figure R11**.

When both promoters are switched off (**Figure R11A**, blue bars), there is no *LacZ* mRNA being produced, and so, recombination levels stay at low values in the wild-type. The increase observed in the mutants can be attributed to the background instability phenotype of these mutations or to a leaky transcription

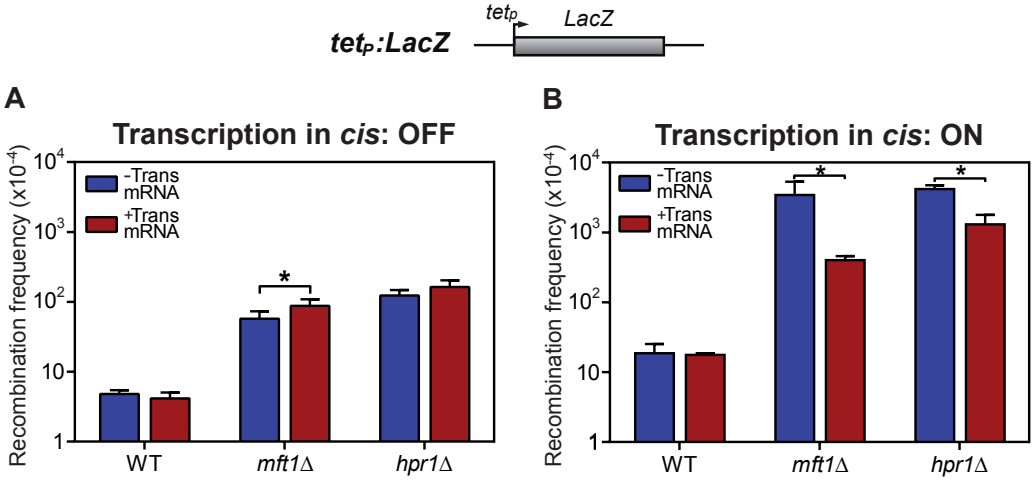


Figure R11. Analysis of the effect of RNA in trans on genetic recombination (II)

Recombination analysis in WT (W303), mft1Δ (WMK.1A) and hpr1Δ (U678.4C) strains carrying GL-LacZ plasmid system (pRS314-GL-LacZ) plus either the empty vector pCM190 (- Trans mRNA) or the same vector carrying the LacZ gene (+ Trans mRNA). Average and SD of at least three independent experiments consisting in the median value of six independent colonies each are shown. *, p≤0.05. (Student's t-test). A scheme of LacZ transcriptional fusion used in this experiment is shown on top.

from the promoter. Under this condition, the presence of an mRNA in *trans* (**Figure R11A**, red bars) had no effect on recombination, only a minor statistically significant difference was observed in the *mft1*Δ strain alone. On the other hand, when transcription was activated in *cis,* the recombination frequency increased noticeably as it was expected from these mutations (**Figure R11B**, blue bars). Unexpectedly, presence of mRNA molecules *in trans* led to a moderate but significant suppression of hyper-recombination in both mutants (**Figure R11B**, red bars). In view of these results, we reasoned that failure in the formation of a functional ribonucleoprotein (mRNP), as it does occur in the THO mutants studied, can lead to the initiation of recombination events in homologous regions where transcription is not active. However, these events seem to take place at a very low frequency and we still would need to figure out, with a molecular assay such as a DRIP, whether the HR is due to the spontaneous formation of a DNA:RNA hybrid in *trans* or not.

The suppression effect observed when both *cis* and *trans* mRNAs were produced at the same time could have different explanations. For example, the *trans* mRNA could interfere with R-loop formation by affecting the *cis* transcription, resulting in a reduction of RNA production. Alternatively, the *trans* mRNA could anneal with the displaced ssDNA in the *cis* transcribing locus, interfering in some way with the ability of the co-transcriptional R-loop to trigger HR.

For this reason, we wanted to further elucidate the mechanism behind the capacity of the mRNA in *trans* to partially suppress the hyper-recombination phenotype of these mutants. For this, we cotransformed the plasmid pGL-*LacZ* carrying the recombination system together with a plasmid, in which the *LacZ* sequence had been cloned in the opposite direction of transcription (inverted *LacZ*, *LacZi*). As can be observed in **Figure R12A**, the wild-type strain did not manifest any effect generated by the production of a *trans* mRNA molecule. Transcription of an antisense transcript *in trans* did, however, led to a small but significant reduction in the recombination frequency of both mutant strains assayed, opposing to the result obtained when the *trans* was sense to the *cis* gene target in the recombination

system (**Figure R11A**, red bars). *Cis* transcription activation led to the expected TAR-associated increase in the recombination frequency value when *trans* mRNA was absent (**Figure R12B**, blue bars). Then again, the presence of the antisense mRNA produced *in trans* when the *in cis* transcription was active produced a significant drop on the recombination levels (**Figure R12B**, red bars), but just in the THO mutant strains, as values for the wild-type remained unaltered.

Up to this point, our data indicated that genomic instability associated to the production of *trans* mRNA did not have a observable impact on a wild-type strain assayed with our system. Additionally, the THO mutants assessed did show a difference when *trans* mRNA was being produced, and both behaved similarly, as expected for two components of the same complex. Because *mft1*Δ mutant grew better than *hpr1*Δ, we decided to broaden the analysis substituting the Hpr1 deficient strain for a different one. Double RNase H mutants, *rnh1*Δ *rnh201*Δ*,* are also prone to the accumulation of mRNAs, but present no defects in the transcriptional process (Huertas & Aguilera, 2003). For this experiment we cotransformed yeast cells with the plasmid pCM189 carrying the *tetp::LacZ* fusion and with the pGL-*LacZ* plasmid for the analysis of recombination.

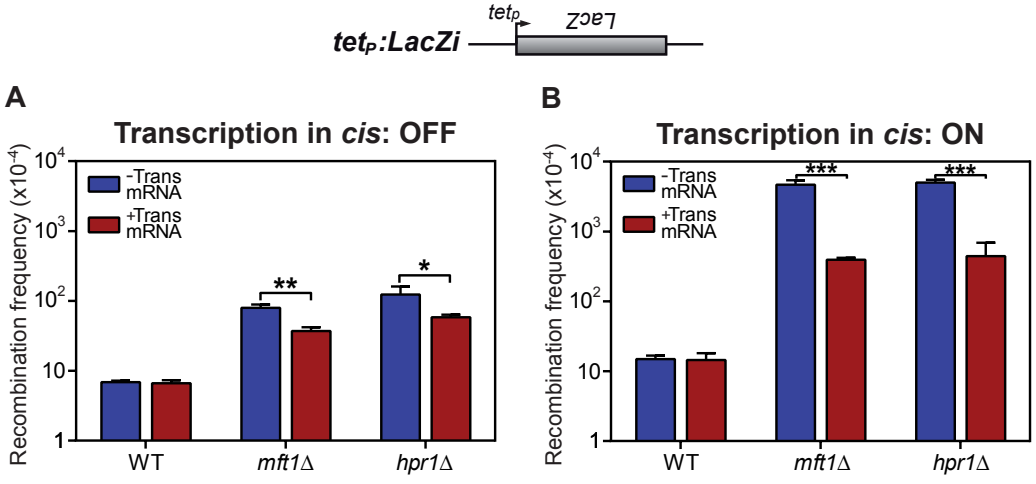


Figure R12. Analysis of the effect of RNA in trans on genetic recombination (III)

Recombination analysis in WT (W303), mft1Δ (WMK.1A) and hpr1Δ (U678.4C) strains carrying GL-LacZ plasmid system (pRS314-GL-LacZ) plus either the empty vector pCM190 (- Trans mRNA) or the same vector carrying the inverted sequence of the LacZ gene (+ Trans mRNA). Average and SD of at least three independent experiments consisting in the median value of six independent colonies each are shown. *, p≤0.05; **, p≤0.01; ***, p≤0.001. (Student's t-test). A scheme of LacZ transcriptional fusion used in this experiment is shown on top

Blue bars in **Figure R13** show that transcription *in cis* alone is responsible of a clear impact on recombination levels in all three strains assayed, with values for the double *RNH* mutant a little lower than that of the THO mutant, *mft1*Δ. For this experiment, the values obtained when simultaneously expressing the *trans* mRNA show no effect in neither condition, with the *cis* transcription repressed (**Figure R13A**, red bars) or active (**Figure R13B**, red bars). We obtained consistent results with the wild-type strain, in which we could not observe any significant effect of the *trans* mRNA expression. Even though the *mft1*Δ mutant displayed a small statistically significant increase in instability in the first experiment when transcribing the *trans* mRNA and the *cis* target was repressed **(Figure R11A)**, we could not replicate that result. This also occurred when we activated the *cis* transcription; the hyper-recombination phenotype could not be as effectively repressed as in previous assays by the *trans* mRNA molecule. Also, loss of RNH function didn't seem to increase the odds for the *trans* mRNA to generate more DNA:RNA hybrids that led to an increase in recombination frequency as measured by this assay.

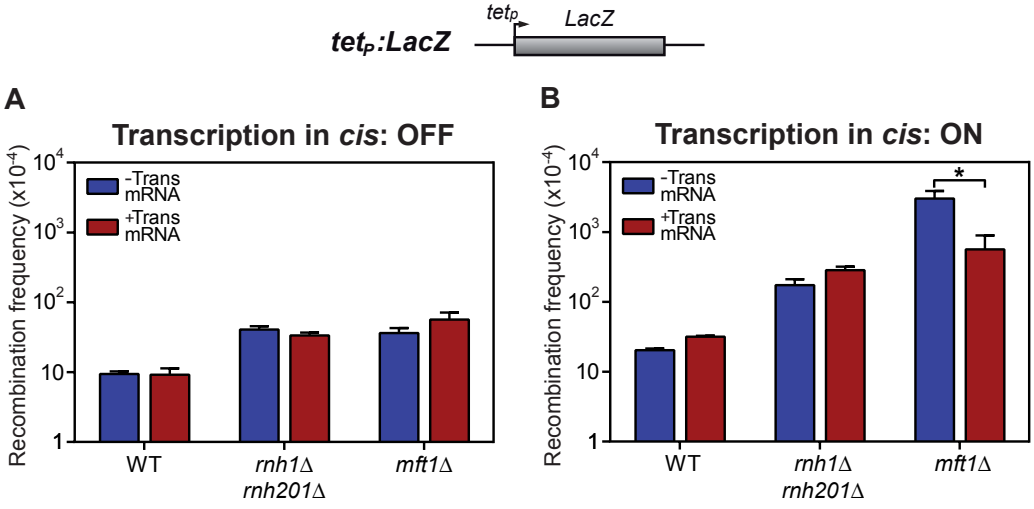


Figure R13. Analysis of the effect of RNA in trans on genetic recombination (IV)
Recombination analysis in WT (W303), rnh1Δ rnh201Δ (HRN2.1C) and mft1Δ (WMK.1A) strains carrying GL-LacZ plasmid system (pRS314-GL-LacZ) plus either the empty vector pCM190 (- Trans mRNA) or the same vector carrying the LacZ gene (+ Trans mRNA). Average and SD of at least three independent experiments consisting in the median value of six independent colonies each are shown. *, p≤0.05. (Student's t-test). A scheme of LacZ transcriptional fusion used in this experiment is shown on top.

Consistent with this approach, we presume that spontaneous contacts between the *trans* RNA and the target DNA sequence must have very low chances to occur. Moreover, the bacterial *LacZ* ORF consisting of a 3kb sequence with approximately 50% G+C content is difficult to transcribe and lead to a non-very stable mRNA even for a wild-type yeast strain (Chavez *et al.*, 2001). Therefore, we reasoned that a smaller fragment could be transcribed at higher frequency and into a more stable RNA than the whole gene, and thus favour the accumulation of higher quantities of mRNA molecules in the nucleus that could increase the chance of hybrid formation. For this reason, we decided to assay a new condition by generating a new transcriptional fusion with the last 400bp of the bacterial gene transcribed from a *tet* promoter. This new plasmid was cotransformed with the pGL-*LacZ* system into the *mft1*Δ and *rnh1*Δ *rnh201*Δ mutants. The recombination frequency values can be observed in**Figure R14**.

As can be seen, the transcription of a shorter fragment of the *LacZ* gene did not generate a significant change in the effects previously observed by a full *trans* mRNA on genetic recombination. When *cis* transcription was repressed, no

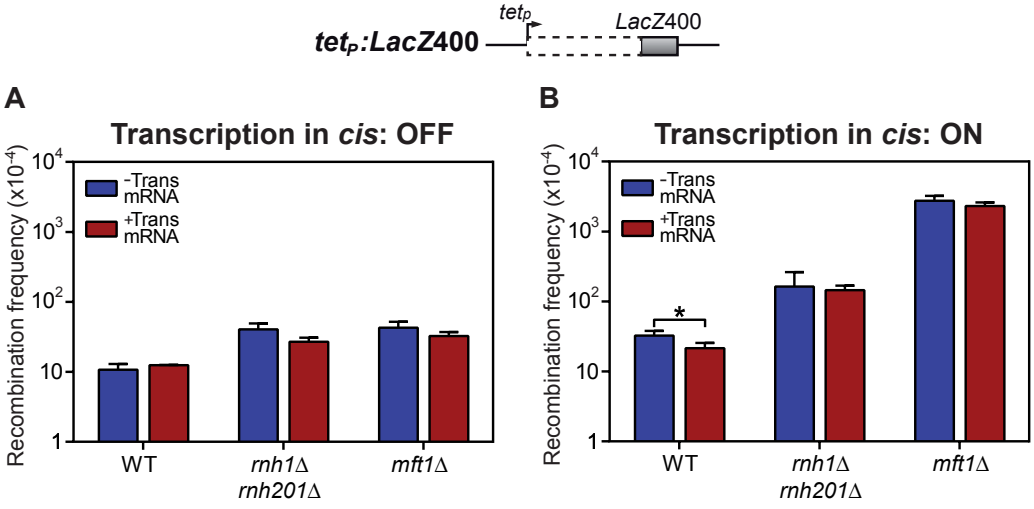


Figure R14. Analysis of the effect of RNA in trans on genetic recombination (V)

Recombination analysis in WT (W303), rnh1Δ rnh201Δ (HRN2.1C) and mft1Δ (WMK.1A) strains carrying GL-LacZ plasmid system (pRS314-GL-LacZ) plus either the empty vector pCM190 (-Trans mRNA) or the same vector carrying the last 400 bp from the 3' end of the LacZ gene (+ Trans mRNA). Average and SD of at least three independent experiments consisting in the median value of six independent colonies each are shown. *, p≤0.05. (Student's t-test). A scheme of LacZ transcriptional fusion used in this experiment is shown on top.

significative impact was detected, although the *trans* mRNA presence induced a small recombination increase on the *mft1*Δ mutant. Under active *cis* transcription, just the THO mutant (*mft1*Δ) shows a statistical significative suppression of the hyper-recombination phenotype when the *LacZ* was being transcribed, similarly to the values obtained when expressing the full length *trans* mRNA. This result suggests that however the *trans* mRNA is decreasing the impact of TAR on *mft1*Δ cells, it is not dependent on the length of the *trans* transcript.

2.2. The homologous recombination machinery is not needed for R-loop formation

So far we had shown that non co-transcriptionally produced mRNA was not able to induce any impact on the genomic stability on a wild-type strain. Additionally, only the *mft1*Δ mutant, defective in the formation of the THO complex, presented a slight increase in the recombination frequency when the *trans* mRNA was expressed and the *cis* transcription process was switched off. Unexpectedly, when the transcription of both *cis* and *trans* mRNAs was simultaneous, hyper-recombination in the THO mutants was largely reduced. These experiments were conducted based on the assumption that formation of a DNA:RNA hybrid is not actively facilitated by a specific factor, and that these structures are also spontaneously occurring even in wild-type cells (Aguilera & García-Muse, 2012). Inded, it is widely accepted that the increase in genetic instability is mostly derived from the uncontrolled accumulation of hybrids between the nascent mRNA and the template DNA due to the absence of factors responsible for their prevention or removal. However, the study of Wahba *et al.* (2013) supports the idea that there are also proteins responsible of the annealing process required for the formation of RNA:DNA hybrids. Otherwise hybrids would be too difficult to occur as a consequence of spontaneous interaction between the homologous molecules. Thus it was suggested that the homologous recombination protein *rad51* was needed for hybridization. We decided to test with our tools their hypothesis.

For this, we generated a new strain (*hpr1*Δ *rad51*Δ) to assay whether or not the

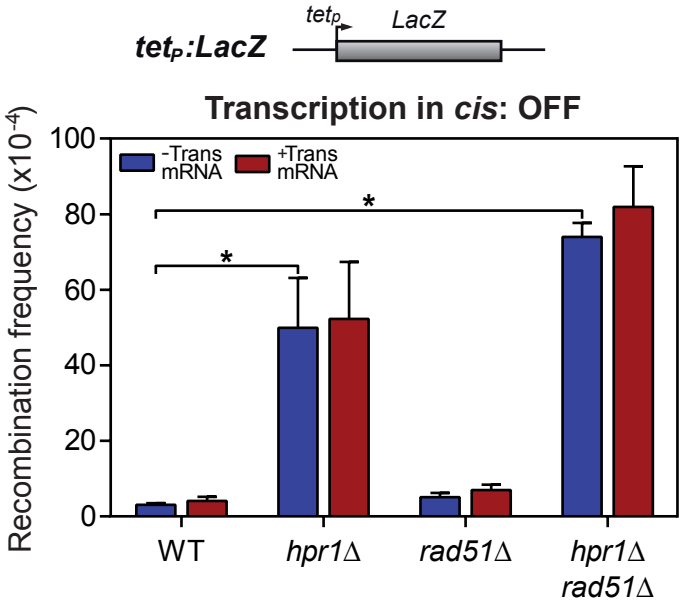


Figure R15. Analysis of the effect of RNA in trans on genetic recombination (VI)
Recombination analysis in WT (W303), hpr1Δ (U678.1C), rad51Δ (WSR51.4A) and hpr1Δ rad51Δ (HPR51.15A) strains carrying GL-LacZ plasmid system (pRS314-GL-LacZ) plus either the empty vector pCM190 (- Trans mRNA) or the same vector carrying the LacZ gene (+ Trans mRNA). Average and SD of at least three independent experiments consisting in the median value of six independent colonies each are shown. *, p≤0.05. (Student's t-test). A scheme of LacZ transcriptional fusion used in this experiment is shown on top.

recombination proteins are needed to regulate the formation of RNA:DNA hybrids in *trans*, provided that this would occur at a detectable level. We cotransformed the double mutant and its corresponding congenic single mutant and wild-type strains with the plasmids pGL-*LacZ* and that carrying the transcriptional fusion *tetp:LacZ*. For this assay we only studied the effect of the *trans* mRNA in the repression state for the *cis* transcription. Results are shown in **Figure R15**. Consistent with previous results, the wild-type recombination frequency values did not show any significative difference between the conditions of expression or absence of the mRNA *in trans,* as neither of the other strains, despite showing small increases. This result supports the idea that the formation of R-loops *in trans* cannot be detected and are consistent with the notion that if Rad51 would be required, the fact that *rad51*Δ has no effect confirms that *trans* RNA:DNA hybrids do not occur spontaneously.

Another way to assay the DNA:RNA hybrids impact in genetic instability is by the analysis of Rad52-YFP foci. This study can be complemented with the overexpression of RNH1, that would suppress any foci-accumulation phenotype

dependent on the presence of DNA:RNA hybrids; and by expression of the Activation-Induced cytidine Deaminase, AID, which would exacerbate the DNA damage by using the displaced ssDNA of the R-loops as target. Consequently, we cotransformed the strains from the previous experiment with a plasmid carrying the Rad52-YFP fusion, and either the plasmid pCM189 or the same plasmid carrying the *RNH1* or the *AID* gene, or both at the same time. The accumulation of Rad52-YFP foci in a *rad51*Δ background has been proposed before to be likely attributed to the persistence of the foci due to the downstream role of Rad51 in the homologous recombination process initiated by the incorporation of Rad52 to the resected DNA breaks (Alvaro *et al.*, 2007). The results obtained, shown in**Figure R16**, support the conclusion that the genetic instability phenotype caused by the *rad51*Δ mutation is independent of DNA:RNA hybrids formation, as there are no significative changes in the Rad52 foci levels when expressing either the RNH or the AID proteins, contrary to what is shown for the positive control, the *hpr1*Δ strain. Moreover, the percentage of replicating cells presenting Rad52 foci in the *hpr1*Δ *rad51*Δ double mutant was pretty similar to that of the Rad51 deficient cells, suggesting that R-loops formed in the absence of Hprl do not require the action of Rad51 to be generated, in which case, *hpr1*Δ instability would have been suppressed by the *rad51* deletion rather

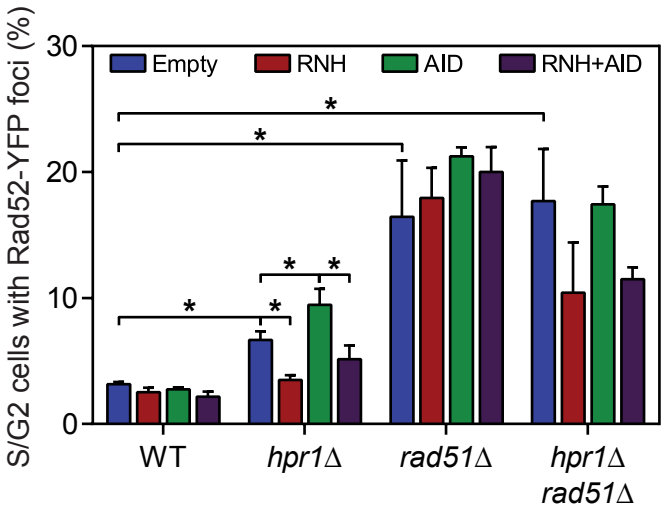


Figure R16. Analysis of Rad52 foci in hpr1Δ rad51Δ double mutants.

Spontaneous Rad52-YFP foci formation in WT (W303), hpr1Δ (U678.1C), rad51Δ (WSR51.4A) and hpr1Δ rad51Δ (HPR51.15A) strains carrying the empty vectors pCM184 and pCM189, or a combination of both carrying the RNH1 or AID genes as indicated in the legend. Average and SD of at least three independent experiments are shown. *, p≤0.05. (Student's t-test).

than having caused an increase. However, the over-expression of the RNH reduced the levels of Rad52 foci in the double mutant, both alone or in the combination with the AID, although when AID was expressed separately did not produce any effect. Thus, these results support the notion that there was not an increment in the target DNA associated to elevated levels of R-loops.

In conclusion, our results do not support the proposition that *rad51*Δ is necessary for the formation of the majority of R-loops in the cell, neither in *trans* nor in *cis*, since the double *rad51*Δ *hpr1*Δ mutant showed increasing levels of recombination and Rad52 foci, with the last phenotype partially suppressed by the overexpression of *RNH1*.

2.3. Analysis of R-Loop in *trans* mediated recombination in a chromosomal system

The necessity to cotransform two plasmids at the same time for this study was an important limitation to the further advance in our research. For this reason, we decided to integrate the GL-*LacZ* recombination system in the chromosomal *LEU2* locus (Chromosome III). This would allow us to transform the cells with a variety of different plasmids to test the effect of different mRNAs in *trans,* either different at the sequence level or in the control of the transcription and, moreover, other plasmids carrying genes or constructs of interest.

After the generation of the wild-type strain that carried the GL-*LacZ* system at the chromosome, we generated a *hpr1*Δ mutant by genetic crosses. We cotransformed the WT and *hpr1*Δ strains with the plasmid carrying the transcriptional fusion *tetp:LacZ* and either the empty plasmid pCM189 as control or the same one carrying either *RNH1* or *AID*.

The results presented **(Figure R17)** only show the assay in which the transcription in *cis* was repressed, since we only wanted to evaluate the effect induced by an mRNA in *trans.* In such a situation, R-loop formation appeared not to be favoured, since there was no major difference between both conditions either in the wild-type or in the mutant. *RNH1* overexpressing strains showed no effect either, consistent with the conclusion that R-loops able to induce recombination do

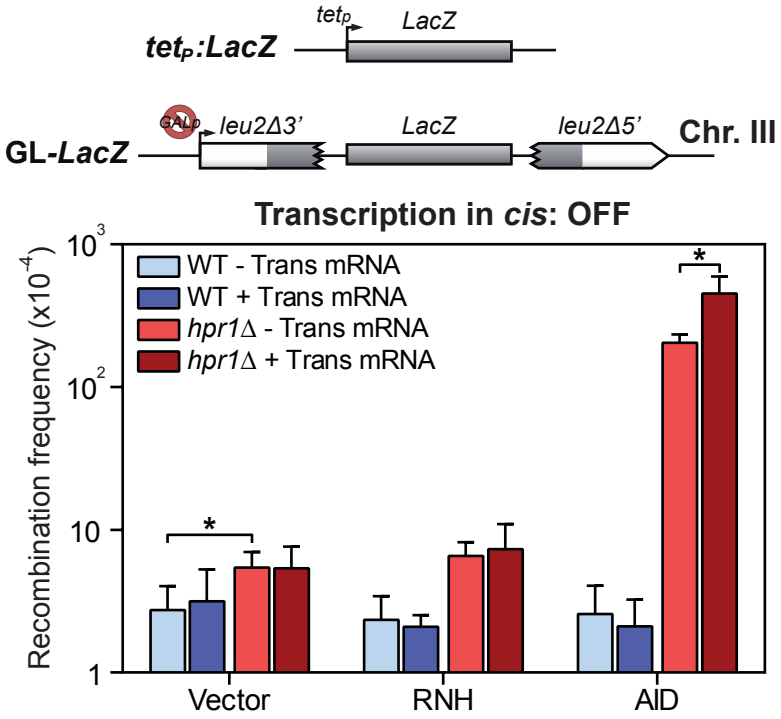


Figure R17. Analysis of the effect of RNA in trans on genetic recombination (VII)

Recombination analysis in WT (WGLZN), hpr1Δ (HGLZN) strains carrying the GL-LacZ recombination system integrated in chromosome III. Strains were transformed with empty vector pCM190 (- Trans mRNA) or the same vector carrying the LacZ gene (+ Trans mRNA) plus either plasmid pCM184 (Vector) or the same vector carrying RNH1 or AID genes (RNH and AID, respectively). Average and SD of at least three independent experiments are shown consisting in the median value of six independent colonies each. *, p≤0.05. (Student's t-test). A scheme of the recombination system is shown on top.

not form *in trans*. Interestingly, however, the AID expression in the *hpr1*Δ mutant apparently affected the hyper-recombination phenotype, as can be interpreted by the significant increase in the value, even without transcription of the *trans* mRNA. It is likely that a basal leaky transcription of the *GAL1* promoter from the recombination system harbouring the *cis* transcription unit was able to induce the formation of a small number of DNA:RNA hybrids in an *hpr1* mutant background, causing the expected effect of AID. Further analysis of the effect of AID would have to be undertaken to understand the molecular nature of this increase.

Results II

3. Search for novel factors involved in DNA:RNA hybrid metabolism

DNA:RNA hybrid excess or its deregulated presence are a well-demonstrated source of genomic instability (Huertas & Aguilera, 2003; Castellano-Pozo *et al.*, 2013; Herrera-Moyano *et al.*, 2014; García-Rubio *et al.*, 2015). However, they are natural occurring structures with a variety of functions, both in prokaryotic and eukaryotic cells (Aguilera & García-Muse, 2012). These facts lead to the idea that there must be many factors and proteins involved in the useful production and on time degradation of these nucleic hybrids. Many recent studies have proposed or shown the importance on these processes of transcription factors, replication regulators and other factors. (Gómez-González *et al.*, 2011; Herrera-Moyano *et al.*, 2014; Mischo *et al.*, 2011). However, the physiological meaning and mechanisms of R-Loop processing are still not understood. In order to gain further insight into this problem, we decided to search for proteins that had not been previously related with R-loop metabolism.

3.1. Selection of candidates

Rad52-YFP foci are formed at recombination centres in the nucleus of the cell (Lisby *et al.*, 2004). Their abundance is directly related to the manifestation of instability problems. Thus, they constitute a sensitive tool to determine if mutants have difficulties in DNA repair or during the DNA replication process. Taking advantage of this visible and measurable phenotype, we designed our screening aiming at finding mutations that had increased Rad52-YFP foci formation that could be suppressed by the overexpression of RNase H. This procedure would let us screen for novel factors potentially implicated in R-loop homeostasis.

We chose to focus our efforts in a selected number of proteins, members of the DEAD box superfamily of helicases. The list of candidates **(Table R3)** was elaborated by picking up mutants according to previous data generated in our laboratory or recently published by other researches related.

Table R3. Dead Box Helicases analyzed

Source	Systematic	Name	Human homolog
BRCA2 Co-IP			
	YBR142W	MAK5	DDX24
	YHR065C	RRP3	DDX47
	YDL084W	SUB2	DDX39A, DDX39B
	YDL160C	DHH1	DDX6
	YDR021W	FAL1	DDX21, DDX50, EIF4A3
	YDR243C	PRP28	DDX23
	YGL171W	ROK1	DDX52
	YLR276C	DBP9	DDX56
	YNL112W	DBP2	DDX17, DDX5
	YOR046C	DBP5	DDX19A,
	YOR204W	DED1	DDX3X, DDX1,
	YPL119C	DBP1	DDX3X, DDX1,
	YBR237W	PRP5	DDX23
	YJL138C	TIF2	EIF4A3
	YKR059W	TIF1	EIF4A3
	YGL078C	DBP3	DDX21, DDX50
HBD Co-IP			
	YNL112W	DBP2	DDX17, DDX5
	YOR204W	DED1	DDX3Y
	YPL119C	DBP1	DDX3Y
Stirling, 2012			
	YLR274W	MCM5	MCM5
	YPL235W	RVB2	RUVBL2
	YBL023C	MCM2	MCM2
Chan, 2014			
	YDL084W	SUB2	DDX39, UAP56
	YNR038W	DBP6	DDX51
	YKR024C	DBP7	DDX31
	YLR424W	SPP382	TFIP11
	YLR430W	SEN1	SETX
	YPL082C	MOT1	BTAF1
Other			
	YIL084C	SDS3	BRMS1
	YIR002C	MPH1	FANCM
	YCL061C	MRC1	Claspin
	YJL092W	SRS2	RETL1
	YMR190C	SGS1	BLM
	YHR031C	RRM3	PIF1

The specified factors were selected by the criteria detailed below:

a. The BRCA2 candidates were selected among a list of positive interactors identified in a co-immunoprecipitation assay performed with protein extracts from HeLa cells (V. Bhatia, unpublished).
b. Three DEAD-Box helicases were identified among other factors when immunoprecipitating DNA:RNA hybrids from HeLa's chromatin extracts expressing the HBD-GFP construct (Bhatia *et al.*, 2014).
c. Three positive candidates related to de DEAD-Box helicase family, associated with DNA replication processes, were selected among the mRNA cleavage and polyadenylation factors involved in the maintenance of chromosome stability identified by Stirling *et al.*, (2012).
d. Four DEAD box helicases selected among a list of mutations described to confer chromosome instability phenotypes that increase R-loops as detected by means of immune-fluorescence using the S9.6 antibody in yeast chromosome spreads (Chan *et al.*, 2014).
e. Six factors including five helicases, involved in DNA repair or replication (SRS2, SGS1, RRM3, MRC1, MPH1 and SDS3).

3.2. Screening

We transformed *Saccharomyces cerevisiae* cells with a plasmid bearing the Rad52-YFP construct under its endogenous promoter. The 25 mutants were divided in two groups, one for the strains from the KO deletion collection of non-essential genes (EUROSCARF, (Winzeler, 1999)) and the second batch with all the essential genes handpicked from the yeast Tet Hughes collection of essential genes (yTHC hereafter) (Mnaimneh *et al.*, 2004), in which the endogenous promoters from selected genes had been substituted by a tetracycline controlled transactivator. The genetic background for all the mutants is derived from BY4741, and the cells were transformed using the lithium acetate method (Ito *et al.*, 1983) adapted according to (Schiestl & Gietz, 1989).

Three different transformants of each mutant were grown in different days to exponential phase in 3 ml of minimum selective media at 30°C. Cells were fixed and

visualized in a Leica Microscope (see M&M 4.9).

From the initial round of selection **(Figure R18)** we could identify 6 mutants from the non-essential helicases and 7 from the essential ones that showed a spontaneous increase in the accumulation of Rad52 foci with respect to their wild-types. Our values correlate quite well with previously published data for some of the strains, as for example, our positive control *hpr1*Δ shows an accumulation around 20% of S/G2 cells with Rad52-YFP foci equally observed by (Alvaro *et al.*, 2007) giving consistency to our results.

Once identified those helicases whose mutations affected the stability of the genome, we assayed whether such alterations could be derived from the accumulation of DNA:RNA hybrids. For this, selected strains were transformed again with a plasmid carrying the RNaseH gene (*RNH1*) under the control of a tetracycline promoter in the case of the KO mutants, and a GAPDH promoter for the yTHC strains.

Only those mutants that showed a consistent reduction in the percentage of cells with Rad52 foci **(Figure R19)** were considered positive candidates for the screening, and are described in the following table **(Table R4)**

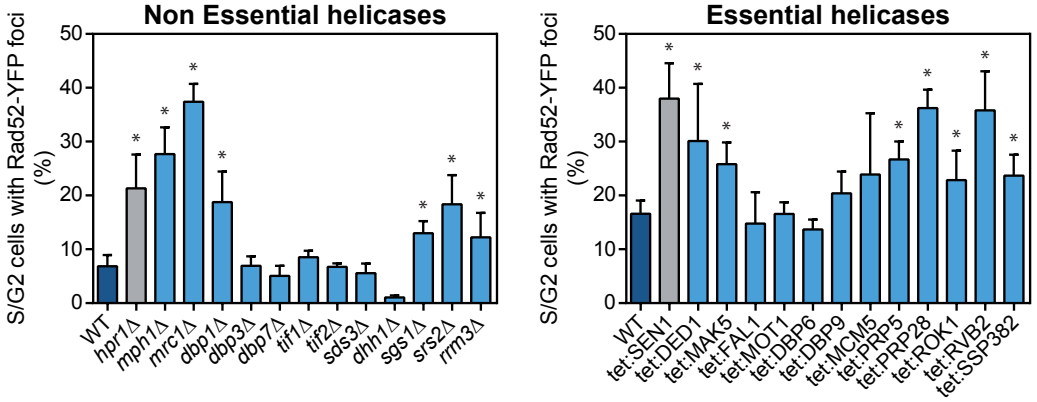


Figure R18. Genetic instability in different DNA and RNA helicase mutants.

Spontaneous Rad52-YFP foci formation in WT and mutant candidates of the selected helicases. The WT used for the non-essential gene series is BY4741 strain and for the essential gene series is R1158. Average and SD of at least three independent experiments are shown. *, p≤0.05 (Student's t-test).

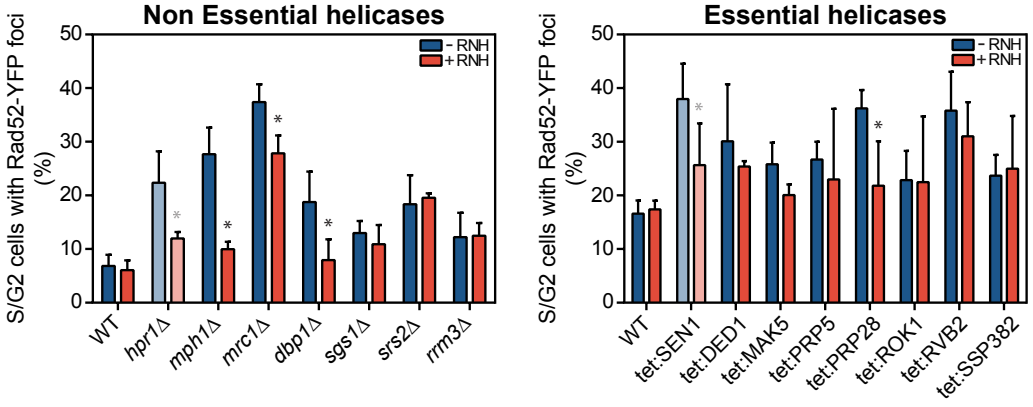


Figure R19. Supression of Rad52-YFP foci by RNH1 overexpression in different helicase mutants

Spontaneous Rad52-YFP formation in WT and selected candidates carrying the empty plasmid pCM189 (-RNH) or pCM189:RNH1 (+RNH) in the deletion collection strains and the empty plasmid pRS425 (-RNH) or pRS425:RNH1 (+RNH) in the yTHC strains. Average and SD of at least three independent experiments are shown. *, p≤0.05 (Student's t-test) for each mutant comparing -/+ RNH conditions.

From these proteins, we discarded for further investigations the RNA binding protein Prp28. It was the only essential gene in the list and was not a helicase itself. The variety of different experimental conditions used to be able to obtain a reliable result suggested the need for further re-evaluation before considering it as a definitive candidate.

Table R4. Selected candidates

Gene	Name	Function
Mph1	YIR002C	DNA helicase. Involved in translession and RF restart. Similar to FANCM
Mrc1	YCL061C	S-phase checkpoint protein required for DNA replication. Couples DNA helicase with polymerase.
Dbp1	YPL119C	Putative ATP-dependent RNA helicase. Required for translation of mRNA. Paralog of DED1.
Prp28	YDR243C	RNA binding protein; no helicase activity. Required for processing of pre-mRNA

3.3. Effect of AID on the genetic instability of candidate mutants

Next, we assayed whether the instability phenotype observed was directly related to the accumulation of R-loops. To gain further insight on this regard, we expressed the human cytidine deaminase AID (Activation Induced Deaminse). AID uses the single stranded DNA displaced in the R-loop as target generating a greater instability defect in yeast strains prone to the accumulation of DNA:RNA hybrids (Gómez-González & Aguilera, 2007; Mischo *et al.*, 2011). We expressed the AID gene under the *tet* promoter from the pCM189::AID plasmid in yeast cells expressing the RAD52-YFP fusion and measured the accumulation of Rad52 foci as indicated above.

Interestingly, although overexpression of *RNH1* consistently suppressed the increase in Rad52 foci, expression of AID did not significantly increase DNA damage above spontaneous levels **(Figure R20)**

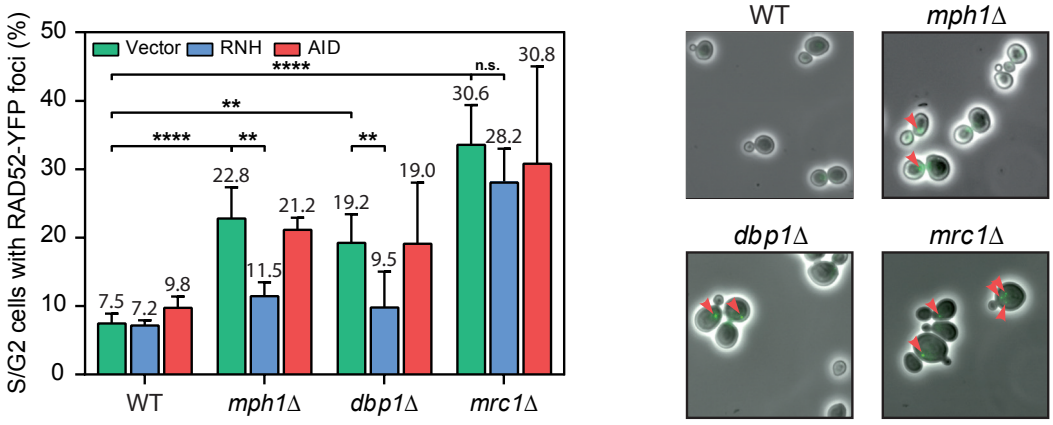


Figure R20. Effect of AID and Rnase H1 expression on genetic instability in selected mutants

Spontaneous Rad52-YFP foci accumulation in WT (BY4741), mph1Δ (YIR002C), dbp1Δ (YPL119C) and mrc1Δ (YCL061C) strains carrying the empty vector pCM189 or expressing RNase H1 from the plasmid pCM189:RNH1 (RNH) or AID from the plasmid pCM189:AID (AID). Average and SD of at least three independent experiments are shown. Representative microscope images are shown on the right. *, p≤0.05; **, p≤0.01; ***, p≤0.001; ****, p≤0.0001. (Student's t-test).

3.4. R-Loop accumulation as detected by immunofluorescence

Suppression of Rad52 foci by *RNH1* overexpression and their exacerbation by AID over-expression serve to indirectly assess the presence of DNA:RNA hybrids. To ascertain the accumulation of these nucleic structures in our mutant candidates we performed immunofluorescence assays using the S9.6 antibody on chromosome spreads from the selected candidate strains. The experiment was done including as control the known R-loop-accumulating strains such as *rnh1*Δ *rnh201*Δ and *hpr1*Δ.

Our results **(Figure R21)** show that the double RNase H mutant strain is the one that shows the highest signal accumulation value of all mutants, consistent with previous observations (Chan *et al.*, 2014). Interestingly, our 3 selected candidates also show a significative increase over the wild-type levels, reassuring a novel connection of these proteins with DNA:RNA hybrids metabolism.

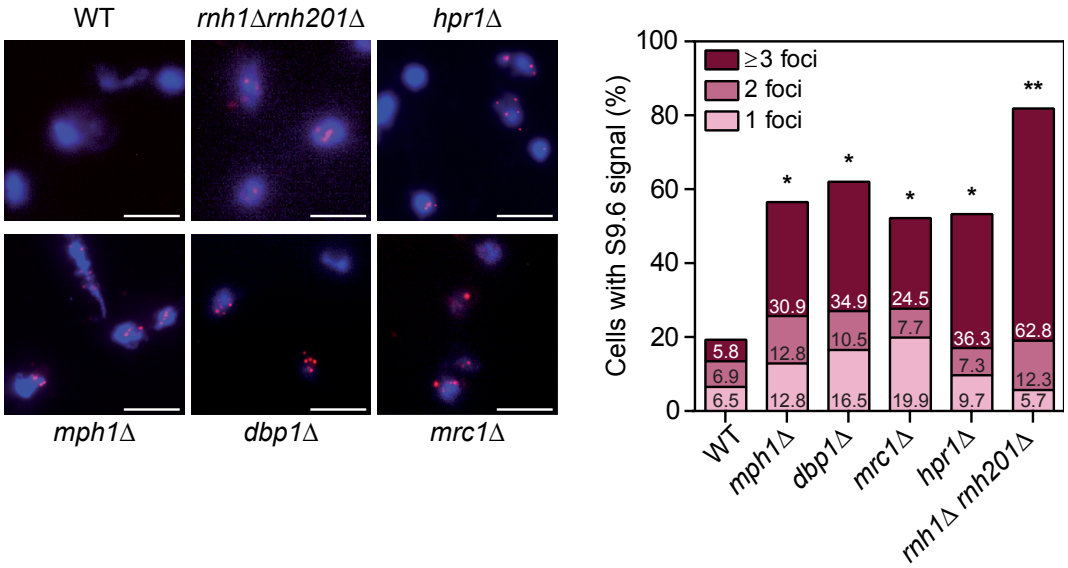


Figure R21. Immunofluorescence of DNA:RNA hybrid accumulation in candidate strains.

Percentage of nuclei and chromosomal masses containing S9.6 signal as detected by indirect immunofluorescence on yeast chromosome spreads in WT (W303.1A), mph1Δ (WMPH1-2B), hpr1Δ (HPBAR1-R) and rnh1Δ rnh201Δ (RNH-R) strains. Representative chromosome spreads are shown: blue stain is DNA (DAPI), red foci are DNA:RNA hybrids detected with S9.6 antibody and white bars indicate 10μm scale. Data represent mean from at least three independent experiments. Statistical analysis calculated with total number of signals for each mutant respect to WT values. *, p≤0.05 (Student's t-test). Experiments performed in collaboration with J. Kaplan

3.5. Effect on recombination and transcription dependency of the mutations selected

Next, we inspected the possibility that the elevated genomic instability, previously observed as high levels of the Rad52 foci, was resolved through the homologous recombination pathway or not. To assess this, we transformed the 3 deletion mutants with the plasmid-borne recombination system LYΔNS, which allowed us to study events of recombination between direct repeats leading to deletions that are detected as Leu+ recombinant colonies. Our findings revealed that despite the increase in the general genetic instability observed previously, neither the absence of Mph1 nor Dbp1 favour the processing of such breaks through homologous recombination, as can be seen in the **Figure R22**. Instead, *mrc1*Δ mutant did show a significative increase over the wild-type value (1.5 fold, $p \leq 0.05$), consistent with previous reports (Robert *et al.*, 2006).

At this point, the *mrc1*Δ mutant was discarded from further studies. The reason behind this decision was the fact that we could not see a significant suppression of

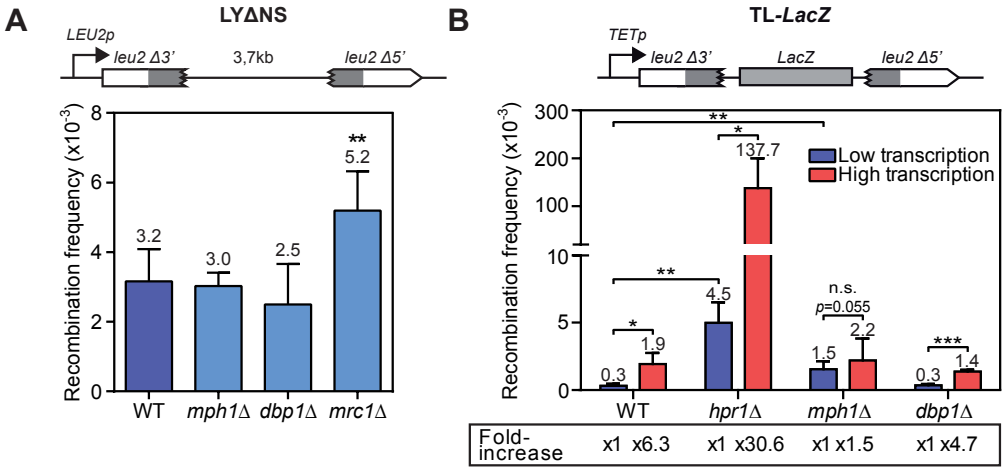


Figure R22. Recombination analysis of selected helicase mutants.

(A) Recombination analysis in WT (BY4741), mph1Δ (YIR002C), dbp1Δ (YPL119C) and mrc1Δ (YCL061C) strains carrying the LYΔNS plasmid system. Average and SD of at least three fluctuation tests consisting in the median value of six independent colonies are shown. (B) Recombination analysis in WT (BY4741), mph1Δ (YIR002C), dbp1Δ (YPL119C) and hpr1Δ (YDR138W) strains carrying the TL-LacZ plasmid system. Cells were grown in the presence of doxycycline (10mg/ml) (Low transcription) or absence of the drug (High transcription). Average and SD of at least three independent experiments consisting in the median value of six independent colonies each are shown. *, p≤0.05; **, p≤0.01; ***, p≤0.001. (Student's t-test). A scheme of the recombination system is shown on top.

the instability phenotype by RNH1. Although the *mrc1*Δ shows a difference in S9.6 signal accumulation respect to the wild-type, by the general role of this factor in replication and as a replication checkpoint, consistent with the hyperrecombination phenotype, suggests that its role, if any, in R-Loop metabolism might be indirect. Therefore, we left it for a future project.

To date, the primary known source of R-loops in yeast is associated with the transcriptional process. Consequently, since *mph1*Δ and *dbp1*Δ did not show any hyperrecombination with the LYΔNS system, we wondered whether we could see an effect using recombination assays in which we could control transcription. To test this, we determined the effect of *mph1*Δ and *dbp1*Δ mutations on recombination in the L-*lacZ* and GL-*lacZ* systems carrying 0.6- kb *leu2* direct repeats flanking the *lacZ* ORF under conditions of low (*GAL1* promoter in 2% glucose), medium (*LEU2* promoter) and high levels of transcription (*GAL1* promoter in 2% galactose). As

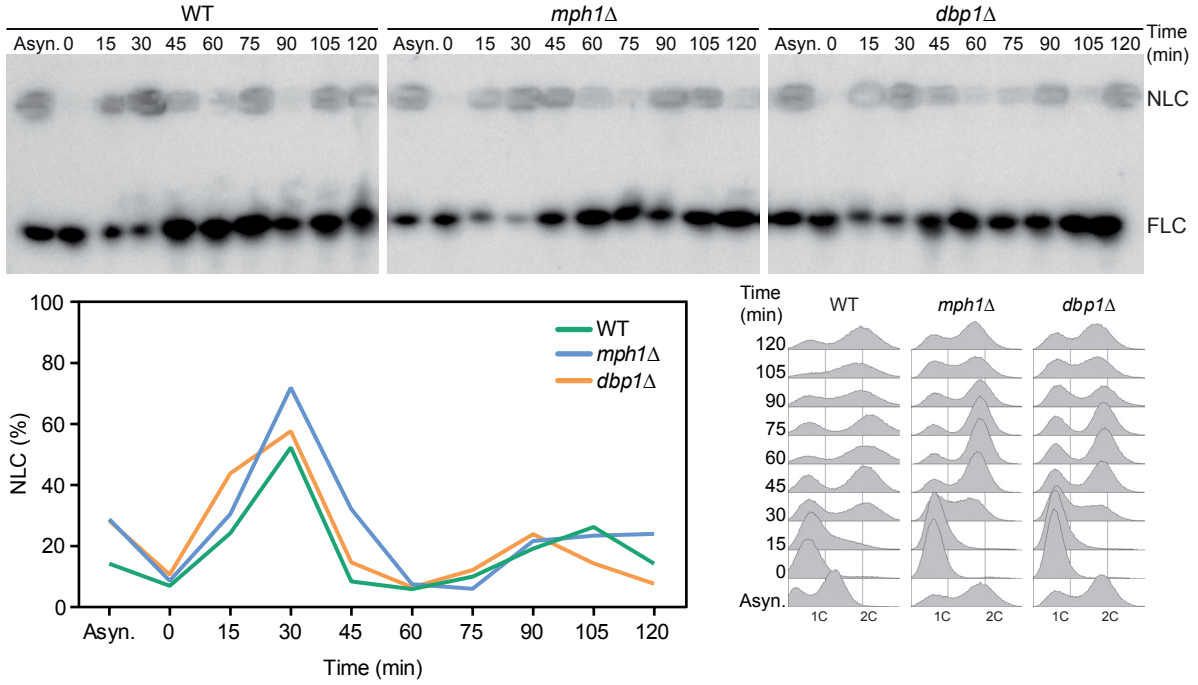


Figure R23. Analysis of the interference with replication process of the selected helicase mutants.

Pulse field gel electrophoresis (PFGE) of DNA from WT (W303), mph1Δ (WMPB1) and dbp1Δ (WDBPB1) cells synchronized in G1 using alpha-factor. Bands reveal chromosome VII as visualized by hybridization with a probe of the ADE5,7 locus. Non-linear chromosomes (NLC) include replication intermediates and remain in the well, full-length linear chromosomes (FLC) are pre- and post-replicated chromosomes and enter the gel. Graph depicts the quantification of NLCs with respect to the total signal of each line. FACS profiles from samples of the same experiment are shown. Experiment performed in collaboration with J. Kaplan.

can be seen in **Figure R22**, the level of transcription increased recombination rates even in the wild-type, as expected. However, in *hpr1*Δ cells, our positive control, this effect is quite higher with a 30-fold increase in recombination under transcription over the non-transcribing state. For our selected mutants, it seems that there is not a significant correspondence between the Rad52-foci accumulation and hyperrecombination. The increases in recombination after transcription observed are not higher than those observed in the wild-type strain.

3.6. Characterization of the impact of the *mph1*Δ and *dbp1*Δ mutation on replication

Once we have discarded a phenotype of hyperrecombination, we tested the hypothesis that *mph1*Δ or *dbp1*Δ strains could present some form of impairment in the replication fork progression. For this, we performed a Pulse Field Gel Electrophoresis (PFGE) assay. This technique allows us to determine the percentage of chromosomes that are under active replication as the DNA fraction that is unable to enter the gel, remaining stacked in the wells during the electrophoresis.

Cells were synchronized in G1 with α-factor for two hours and then released into fresh medium, samples were taken at indicated times for the PFGE and to be analysed by FACS for studying the cell cycle progression through the S phase. The experiment showed that our mutants exhibit a very similar kinetic to that of the wild-type **(Figure R23)**, with only *mph1*Δ showing a little delay in S-G2 phase termination at 45 minutes after G1 release, where about 30% of the DNA molecules still remained in the well compared to the 10% of the wild-type. However, this delay is quickly overcome as, at 60 minutes after G1 release, all three strains show again similar values. Thus *mph1*Δ cells only present a subtle defect in the replication process, if any, that doesn't probably account for their cell cycle progression retardation.

At this point, we decided not to continue working with *dpb1*Δ mutant. Dbp1 deficiency was not causing any visible defects either in the transcriptional process or in cell cycle progression. Moreover, since it is a protein which principal role has been so far associated with translation outside the nucleus, it was not obvious how

mutants of this helicase present a genetic instability phenotype that is suppressed with RNH1 overexpression as detected with the Rad52-YFP foci via a direct effect. Instead, FANCM, which is the human homolog of Mph1 (Whitby, 2010), is a component of the Fanconi Anaemia (FA) pathway of DNA repair, involved in the resolution of Inter Strand Crosslinks (ICLs) that block replication fork progression. Interestingly, some key components of this pathway, like FANCD2 or FANCA had just been reported to be involved in DNA:RNA hybrid resolution (García-Rubio *et al.*, 2015; Schwab *et al.*, 2015), adding a remarkable argument for us to continue our work with Mph1.

3.7. Genetic interaction of *mph1*Δ with other DNA:RNA accumulating mutations.

Although we already had some clues pointing to the existence of a role in R-loop metabolism for the *MPH1* gene, we wondered if the addition of other

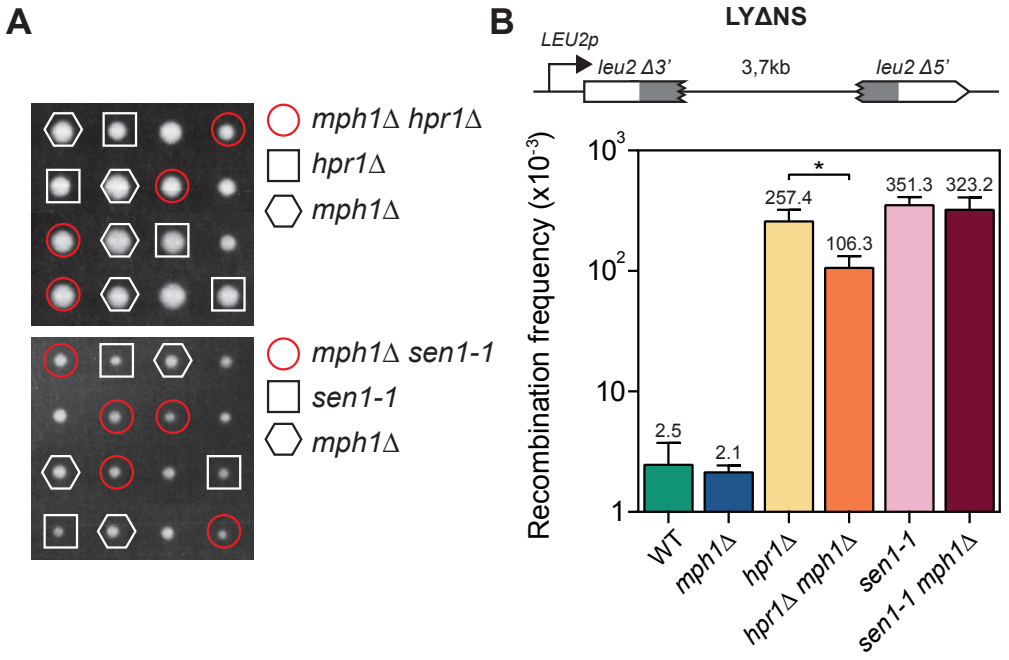


Figure R24. Analysis of the genetic interaction between DNA:RNA hybrid accumulating mutants and mph1Δ.

(A) Tetrad analysis from crosses between mph1Δ (WMPH.2B) and the well characterized R-loop accumulating mutants hpr1Δ (HPBAR-R1) and sen1-1 (SEN1-R). Photos were taken after 4 days of growth after micromanipulation at 30ºC and 26ºC for the hpr1Δ and the sen1-1 crosses respectively. (B) Recombination analysis of WT (W303.1A), mph1Δ (WMPH1.2A), hpr1Δ (HPBAR-R1), hpr1Δmph1Δ (WMPHP.5A), sen1-1 (SEN1-R), sen1-1mph1Δ (WMPSEN.1C) strains. Mean and SD for at least three fluctuation tests consisting in the median value of six independent colonies each are shown. *, p≤00.05 (Student's t-test). A scheme of the recombination system is shown on top

mutations already described to cause an important DNA:RNA hybrid accumulation could lead to the intensification of the *mph1*Δ phenotypes as a signal of genetic interactions. For this reason, we crossed *mph1*Δ with the well-characterized *hpr1*Δ mutant, member of the THO complex (Huertas & Aguilera, 2003; Gómez-González *et al.*, 2011) and the point mutant *sen1-1* of the essential DNA:RNA helicase homologous to the human SENATAXIN (Mischo *et al.*, 2011). As can be seen in **Figure R24A**, the tetrad analysis shows that both double mutants *hpr1*Δ *mph1*Δ and *sen1-1 mph1*Δ were viable and did not show any dramatic growth defect. The colony size of double mutants was pretty similar to that of the respective single mutants.

We further tested these strains for other interactions, in this case we assayed the recombination frequency of the double mutants. It was known that *hpr1* and *sen1* mutants present an elevated genomic instability that translates in high values of recombination. We transformed the double mutants and their respective parental strains with the LYΔNS recombination system. As can be seen in the **Figure R24B**, *mph1*Δ consistently showed similar values to that of the wildtype strain, and parental strains *hpr1*Δ and *sen1-1* behaved as expected. Double mutant *sen1-1 mph1*Δ did not show any different effect on the recombination frequency to that of the *sen1-1* single mutant. On the other hand, *hpr1*Δ *mph1*Δ cells show a significant lower recombination than that of the single *hpr1*Δ strain. More experiments are needed to confirm this result, with other recombination assays. However, this result may be explained by the fact that Mph1 itself is required for recombination (see Discussion).

3.8. Mph1 foci are induced in *hpr1*Δ cells

The study of repair foci formation is a well-established tool for the in vivo characterization of mutants in DNA damage response and HR, as they mark sites of DNA damage and ongoing repair (Lisby & Rothstein, 2015). Mph1 foci had already been used to determine its function and interactions in telomere maintenance and regulation of crossovers (Silva *et al.*, 2016). Our current hypothesis is that R-loops present a major impediment to replication fork progression, thus increasing the frequency of RF stalls. Since Mph1 may be required at stalled replication forks, similar to Fanconi Anemia factos in human cells, and it is well-established that the

main source of phenotypes in *hpr1*Δ mutant is the accumulation of DNA:RNA hybrids in the nucleus (Huertas & Aguilera, 2003) due to the failure in creating a functional ribonucleoparticle (mRNP) with the nascent mRNA (Peña *et al.*, 2012). We decided to test whether a high presence of R-loops could tether the formation of Mph1 foci.

For this, we transformed a wild-type strain and a congenic *hpr1*Δ mutant both carrying the Mph1-YFP fusion, with plasmids pRS315 (-RNH) or pRS315-GAL:RNH1 (+RNH). Cells were grown under induction conditions until mid-log phase and analyzed under the microscope. As can be observed in **Figure R25**, consistent with the idea that Mph1 had a role at DNA:RNA hybrids, we detected a significan increase in the frequency of spontaneous Mph1-YFP foci formation in *hpr1*Δ mutant cells, and this increment is supressed by the overexpression of RNase H1. This result supports our hypothesis about the novel role of Mph1 in R-loop metabolism. Additionally, on the basis that a percentage of these foci are derived from the presence of DNA:RNA hybrids, our results suggest that Mph1-YFP foci could be used as an additional tool to search for new factors involved in R-loop homeostasis.

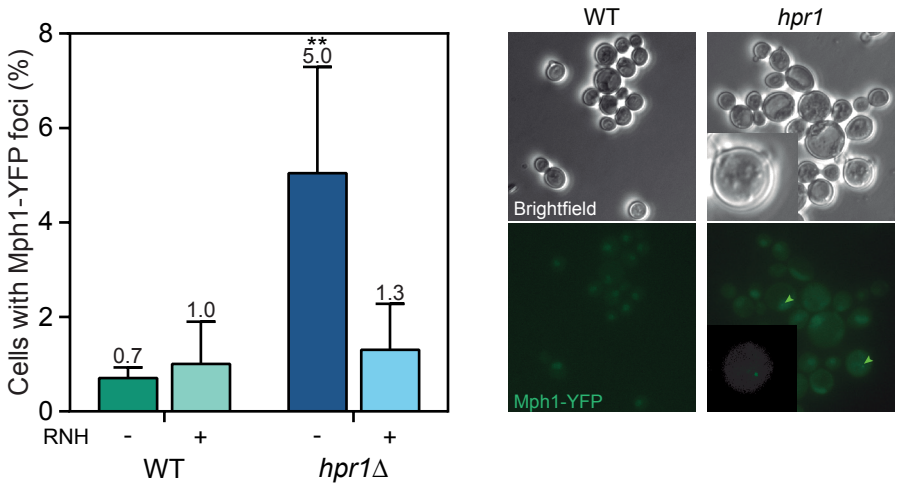


Figure R25. Accumulation of Mph1 foci in hpr1Δ cells.

Spontaneous Mph1-YFP foci were quantified in exponentially growing wild-type (ML66-11A) and hpr1Δ::KANMX (YBG722) cells transformed with pRS315 or pRS315-GALRNH1 at 30ºC. Mean and SD of four experiments performed with independent transformants are plotted. 100-400 cells were analysed in each case. ** p≤0.01 (two-tailed unpaired t-test).

3.9. DNA:RNA hybrid detection by immunoprecipitation in *mph1*Δ mutants

To demonstrate that R-loops accumulate at a molecular level in the *mph1*Δ mutant we performed a direct immunoprecipitation assay using the S9.6 monoclonal antibody at different genes. DNA:RNA hybrids clearly accumulated in the mutant in all DNA regions tested **(Figure R26A)**. We also analysed this increase in two point mutants of the *MPH1* gene mutated at key residues of the helicase domain. The mutant E210Q, which has a replacement of the glutamic acid in the DEAH motif, and the mutant Q603D with a similar modification in the key residue of the helicase motif. These mutants have the helicase function impaired, but behave like *mph1*Δ for DNA damage sensitivity (Chen *et al.*, 2009). As can be observed in **Figure R26B**, these mutants also show a significant increase in the accumulation of DNA:RNA hybrids in the locus analysed. Thus, the helicase function, already characterized to be essential for the Mph1 role in recombination (Schürer *et al.*, 2004; Prakash *et al.*, 2009; Banerjee *et al.*, 2008) is necessary to prevent R-loop accumulation.

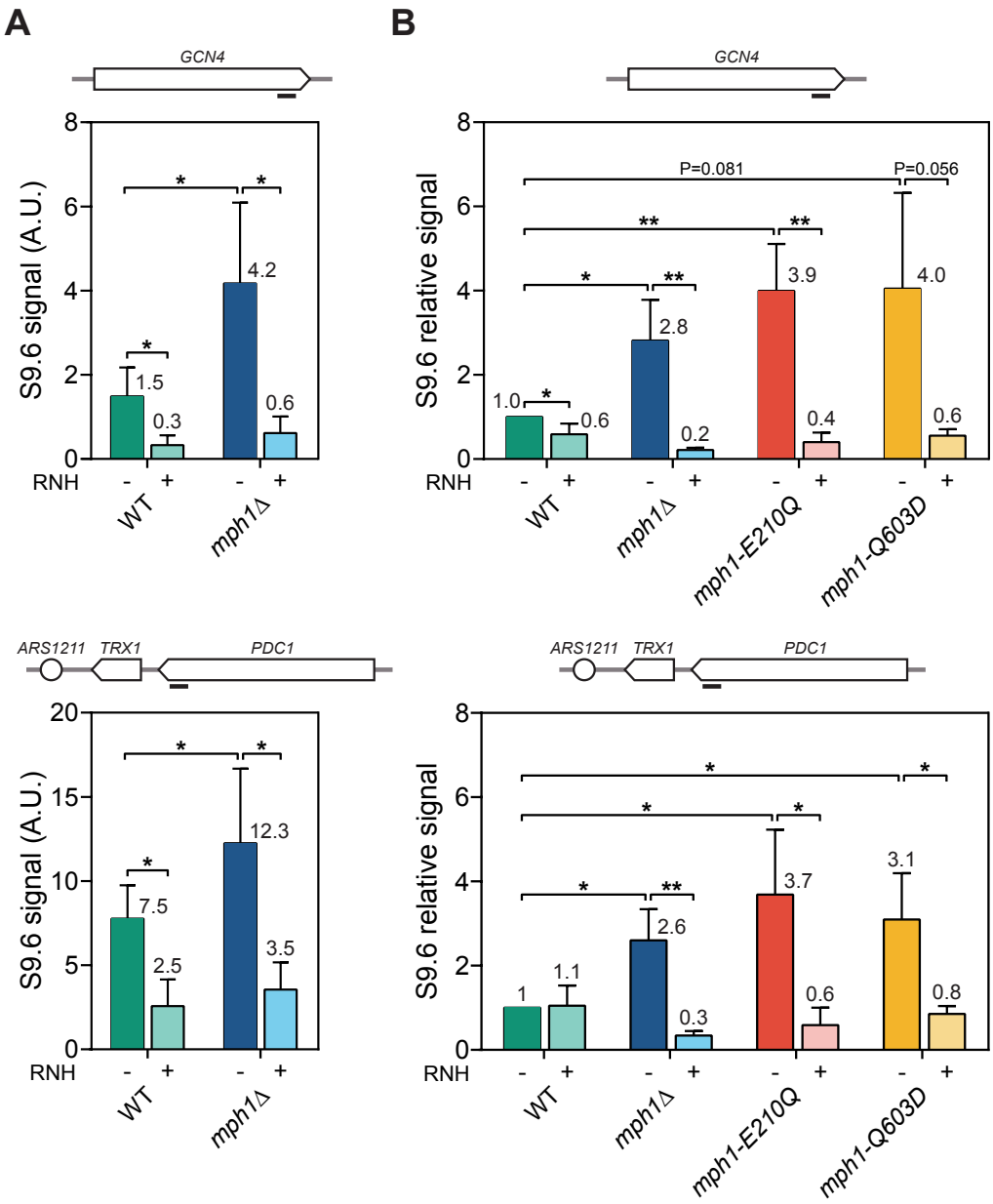


Figure R26. DNA:RNA hybrid accumulation in mph1Δ deletion and helicase-dead mutants.

S9.6 immunoprecipitation of DNA:RNA hybrids (DRIP) in (A) WT (W303), mph1Δ (WMPH1.1A) or (B) WT (W303), mph1Δ (WMPH1.1A), mph1-E210Q (T617), mph1-Q603D (T597-1) strains at the indicated regions (black line). Samples were treated with RNase H before immunoprecipitation. Mean and SD for at least three independent experiments are shown. *, p≤0.05; **, p≤0.01 (Student's t-test).

4. Function of the human helicase RECQL5 in the model organism *Saccharomyces cerevisiae*

The human helicase RECQL5 is a protein member of the well-conserved RecQ family of helicases. Demonstrated genome caretakers (futher reviewed in (Chu & Hickson, 2009; Monnat, 2010; Bernstein *et al.*, 2010; Singh *et al.*, 2011), there are a total of five members present in most mammals, while there is usually just one in lower organisms. RECQL5 is the only one that has been shown to directly interact with the RNA polymerase II (RNAPII) (Aygün *et al.*, 2008; Islam *et al.*, 2010). This relationship makes this helicase unique among the family, since their roles in genome stability maintenance through the regulation of diverse DNA repair processes had been well stablished. The fact that RECQL5 is present at the transcription interface raises many questions about the regulation of the interactions between this process and others that use the DNA as template, such as the replication or the repair processes.

Owing to the absence of assays to address these questions in human cell cultures and the potential of model organisms for the study of the DNA metabolic processes, we decided to assess the capacity of the yeast *Saccharomyces cerevisiae* to serve as a tool for the functional study of the human helicase RECQL5.

4.1. Expression of RECQL5 in yeast

Human RECQL5 transcript presents three alternatively spliced forms: RECQL5α, RECQL5β and RECQL5γ (Shimamoto *et al.*, 2000). The second splice variant, hereinafter referred as RECQL5, comprises a total of 991 amino acids and it is the most abundant and better characterized. It is expressed ubiquitously and independently of cell-cycle phase and cell-type (Kitao *et al.*, 1998; Kawabe *et al.*, 2000). In the scheme in **Figure R27A**, it is shown the domain organization of the functional wild-type protein, which comprises (from left to right) the DEAD-Box helicase (yellow box), the RecQ C-terminal domain (RQC, blue box), which contains the structural features specific of the RecQ helicases and the KIX domain (dark green box), present in several transcriptional regulators (Parker *et al.*,

1996) and responsible for the interaction with both the hypo- (IIa) and the hyper-phosphorylated (IIo) forms of the RNAPII (Islam *et al.*, 2010). This domain has been mutated, as depicted in the lower scheme in **Figure R27A**, (RECQL5 ID, Interaction Domain mutant) with a 27-bp deletion in the sequence, that supposedly abolished its interaction with the polymerase, as previously shown (Aygün *et al.*, 2009). Other important domains are the BRC-repeat variant (BRCv, light green box), a structural motif also found in the tumor suppressor BRCA2 (Islam *et al.*, 2012), which localized to the Rad51 binding region and required for the anti-recombinase activity of RECQL5. The nuclear localization signal (NLS, brown box) and the SRI domain (Set2 Rpb1-interacting domain, orange box) present in the histone methyltransferase Set2 (Kizer *et al.*, 2005), known to interact with the phosphorylated C-terminal domain of the RNAPII (pCTD).

We transformed the wild-type strain W303.1A with multicopy yeast expression plasmids carrying both versions of RECQL5 under the control of a *GAL1* promoter. **Figure R27B** and C illustrates that overexpression of both versions of RECQL5 in *S. cerevisiae* dramatically affected growth as shown in plates of synthetic media with galactose as the sole carbon source. Due to this severe effect, that did not let us pursue the experiments, we undertook two approaches to reduce the levels of RECQL5 in the cells. First, we added different amounts of glucose to the synthetic media containing 2% galactose, in order to reduce the activation level of the *GAL1* promoter, in addition, we supplemented the cells with an alternative carbon source. We selected the media with 0.05% glucose and 2% galactose, in which cells showed reduced colony size that evidenced the expression of RECQL5, but were still able to grow.

On the other hand, we cloned both versions of the human gene into a different vector backbone, in this case, either pCM184 (*TRP1*) or pCM189 (*URA3*). These are centromeric vectors that contain a *tet* chimeric promoter that is repressed by addition of doxycycline to the medium **(Figure R27D)**. The new transcriptional fusion *tetp::RECQL5* generated did not seem to have any visible effect in yeast viability. The correct expression at low levels of the RECQL5 protein was demonstrated by western blot **(Figure R27E)**.

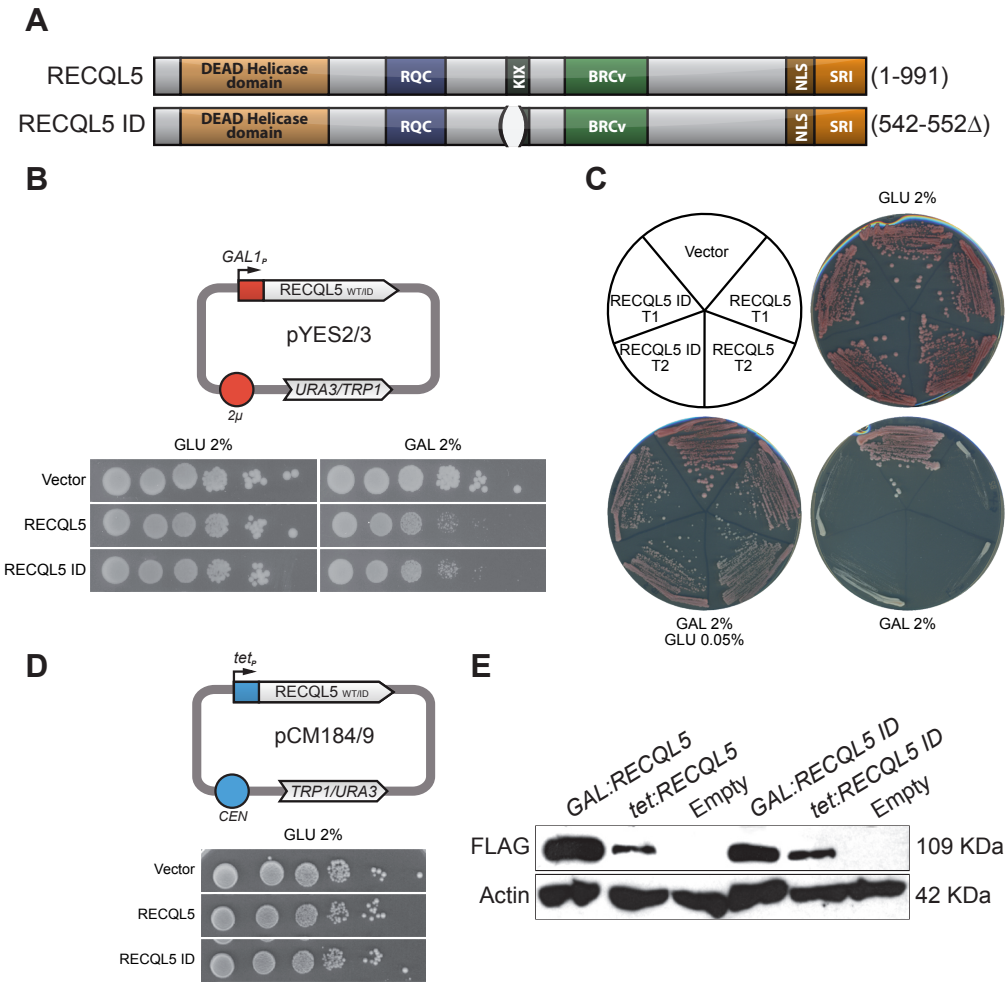


Figure R27. Expression of the human helicase RECQL5 in S. cerevisiae.

(A) Domain organization of the RECQL5 protein. Top scheme corresponds to the wild-type version (RECQL5). Bottom scheme shows a mutant version with a deletion of 9 aa in the KIX domain (RECQL5 ID). (B) Top, scheme showing the multicopy (*2μ*) and galactose inducible (*GAL1p*) vectors used, pYES2 (*URA3*) and pYES3 (*TRP1*). 10-fold serial dilutions of the W303.1A strain transformed with the pYES2 vector, or carrying RECQL5 or RECQL5 ID on minimum synthetic selective media with different carbon sources are shown. (C) Growth of wild-type yeast strain W303.1A transformed with plasmids either pYES3 (Vector), pYES3:RECQL5 (RECQL5) or pYES3:RECQL5 ID (RECQL5 ID) and streaked on synthetic media with different concentrations of sugars. Two different transformants (T1 and T2) are shown for the plasmids carrying genes. Photographs were taken after 3 days of growth at 30°C. (D) Scheme showing the centromeric (*CEN)* and doxycycline repressible (tetp) vectors used, pCM184 (*TRP1*) and pCM189 (*URA3*). 10-fold serial dilutions of the W303.1A strain transformed with the pCM184 vector or carrying RECQL5 or RECQL5 ID on selective media are shown. Photographs were taken after 3 days of growth at 30°C. (E) Western Blot analysis using the FLAG antibody to detect the tagged RECQL5 protein levels expressed from the or the tet promoter. The strain W303.1A was transformed with the indicated plasmids. Actin immunodetection is shown as loading control.

4.2. RECQL5 recruitment to yeast chromatin

We have seen that the heterologous expression of the human helicase was clearly interfering with some cellular processes in the yeast, generating viability defects at higher doses. Our first objective, however, was to assay whether RECQL5 behaved in our model organism similarly as it had been proposed in human cells, by regulating transcription in order to avoid conflicts with replication (Aygün *et al.*, 2009).

For this, we performed co-immunoprecipitation (Co-IP) assays using an antibody against the FLAG epitope fused to our RECQL5 proteins and an antibody that recognised the Rpb3 subunit of the RNAPII. We transformed the W303 yeasts cells with pYES3 plasmids, either empty or carrying the RECQL5 or the mutant ID genes, and grew them for 3 hours in galactose until exponential phase. Through western blot of the corresponding immunoprecipitated protein extracts we demonstrated that RECQL5 was able to interact with the *S. cerevisiae* RNAPII **(Figure R28A)**. Surprisingly, the mutant RECQL5 ID, which was supposed to have this capability abolished (Aygün *et al.*, 2008), was also co-IP with RNAPII the same way as the wild-type helicase.

We additionally performed ChIP (Chromatin Immunoprecipitation) to ensure that this interaction correlated with an active RNAPII, and not just with unspecific interaction. We transformed the W303.1A strain with the plasmid pCM184 either empty or carrying the wild-type RECQL5 or the mutant RECQL5 ID under the control of a *tet* promoter. Purified chromatin from these exponentially growing cells, either in glucose (repressed transcription) or galactose (active transcription), was subjected to IP using the antibodies described before. Results from this experiment **(Figure R28B)** also evidenced that both versions of RECQL5, wild-type and mutant, interacted with yeast's chromatin and in a transcription-dependent manner. Our finding that the interaction of RECQL5 did not significantly affect the recruitment of RNAPII to the transcribing gene was opposed in to that observed in human cells by Saponaro *et al.* (2014), published later the completion of these experiments. This outcome was indicative that, although the interaction domains with the RNA polymerase could be conserved from yeast to humans, RECQL5

was not acting in the same exact way as when interacting with the human RNAPII in a physiological environment. In any case, our results prompted us to continue with the work, since the interaction between RECQL5 and the yeast RNAPII and its chromatin was confirmed, we could now study if this was the cause of the viability defects observed or there were other roles implicated that could help us learn more about the still unclear role of RECQL5.

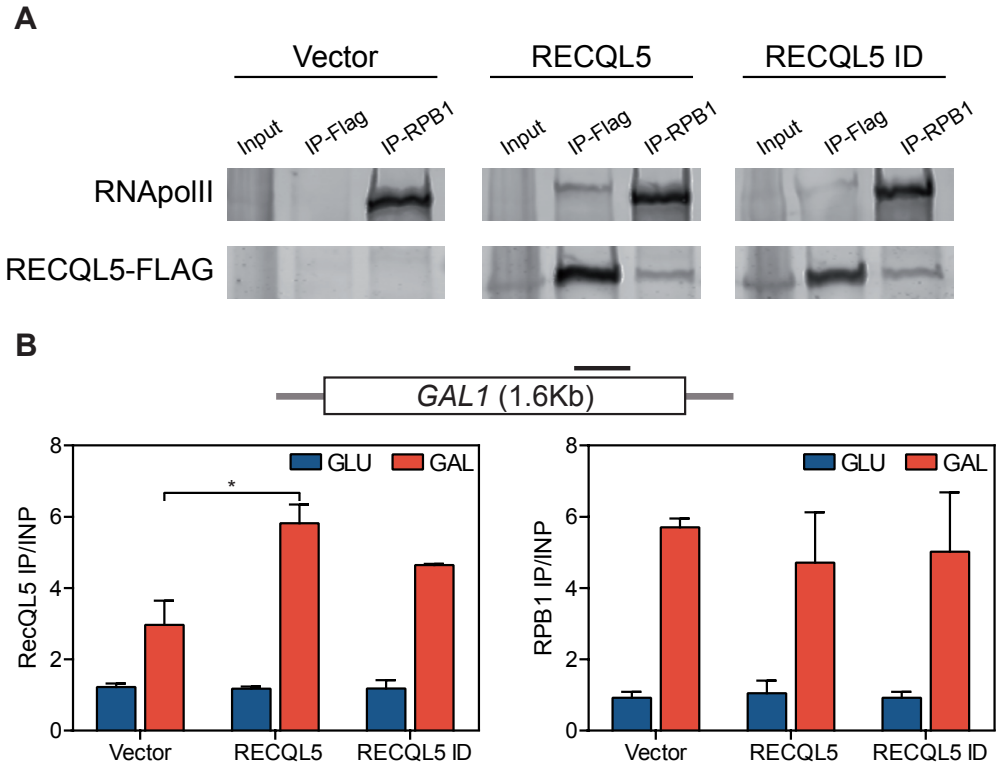


Figure R28. Interaction of the human helicase RECQL5 with yeast chromatin.

(A) Western Blot of different immunoprecipitations (IP) showing interactions between the two forms of the FLAG-tagged RECQL5 and RNAPII, using FLAG and RPB1 antibodies, respectively. Strain W303.1A was transformed with either pYES3 (Vector), pYES3:RECQL5 (RECQL5) or pYES3:RECQL5 ID (RECQL5 ID). (B) ChIP (Chromatin immunoprecipitation) of W303.1A cells transformed with either pCM184 (Vector), pCM184:RECQL5 (RECQL5) or pCM184:RECQL5ID (RECQL5 ID) plasmids. qPCR was performed in the GAL1 locus under conditions of active (GAL) or repressed (GLU) transcription. The black bar represents the amplified region.

4.2.1. Sensitivity of RECQL5-expressing yeast cells to genotoxic agents

The physiological role of RECQL5 in mammal cells had been associated with the maintenance of genome stability, presumably by acting as a regulating factor at the interface between transcription and replication. We wondered whether the expression of the human RECQL5 helicase in yeast could help the cells overcome the effect of the DNA damages produced by different genotoxic agents or, on the contrary, could interfere with the cellular response.

For this, we transformed the wild-type strain W303.1A with the transcriptional fusions (both wild-type RECQL5 and the RECQL5 ID mutant) expressed from the *GAL1p* or the *tetp* promoters. By plating serial dilutions of exponential cultures in petri dishes we assayed the effect on viability when cells were grown on synthetic media supplemented with different concentrations of three different chemicals. We tested the genotoxic agents hidroxiurea (HU), which interferes with replication by depleting the amount of deoxyribonucleotide triphosphate (dNTPs); methyl-methanesulfonate (MMS), an alkylating agent responsible of DNA methylation that leads to the formation of DSBs, presumably after the stalling of replication forks; and mycophenolic acid (MPA), a transcription elongation inhibitor that acts by interfering with the IMP dehydrogenase and, thus, the biosynthetic pathway of purine nucleotides impeding the transcription of newly synthesised mRNAs.

The results obtained **(Figure R29)** evidenced that the presence of either the wild-type or the ID mutant version of the RECQL5 helicase, indistinctly, produced a clear impact on the maintenance of the yeast's genome stability when cells were grown under genotoxic stress induced by HU or MMS, but not with the transcription elongation inhibitor MPA. Moreover, this effect was seen independently of the promoter strength, in both cases, either under the *tetp* or the *GAL1p* promoters. The presence of even low concentrations of HU or MMS in the media visibly affected the viability of the cells expressing either allele of the RECQL5 helicase in a similar way, compared to the yeast transformed with the empty vector, used as control in each condition. Surprisingly, MPA produced no effect in viability, no difference in growth rate was noticeable between the cells carrying the empty vector and those expressing RECQL5 gene at all MPA concentrations tested. This experiment

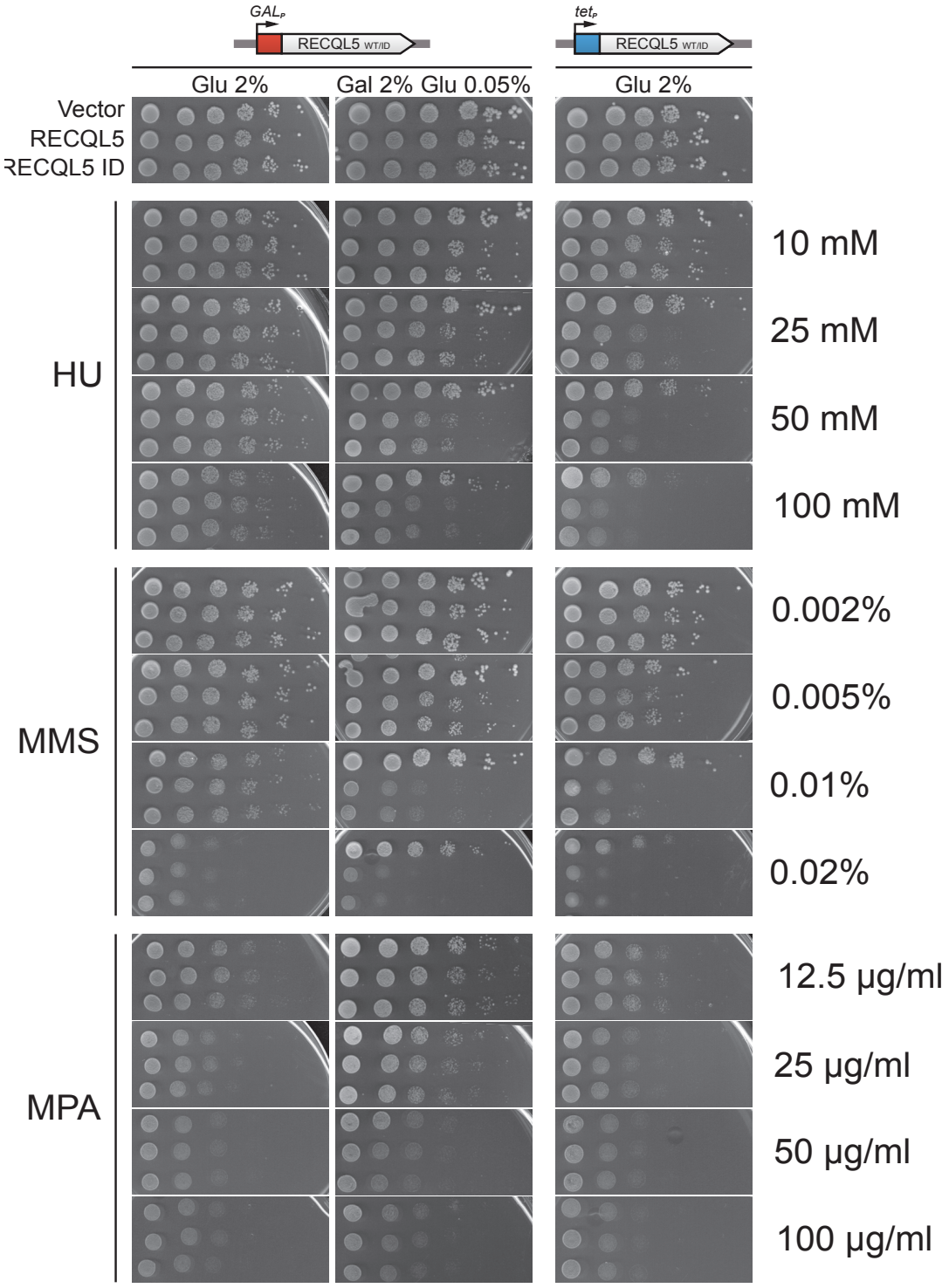


Figure R29. Effects of RECQL5 overexpression on the sensitivity to different chemical compounds.

Sensitivity of W303.1A strain transformed with indicated plasmids carrying the wild-type RECQL5 gene, or the mutated sequence of RECQL5 ID, or the empty vector, to different concentrations of genotoxic agents such as hidroxyurea (HU) and methylmetasulfonate (MMS) or the transcription elongation inhibitor mycophenolyc acid (MPA). 10-fold serial dilutions of exponentially growing cells in selective medium were plated on the indicated media. Photographs were taken after 3 days of growth at 30°C

grounded the idea that heterologous expression of this protein was interfering with cell proliferation, possibly with some aspect of replication. Consistent with the sensitivity caused by HU and MMS, the absence of a detectable effect on cell viability after MPA does not allow us to conclude whether or not RECQL5 functionally interacts with the transcriptional process of the yeast. However, we cannot discard at this point that RECLQ5 mRNA synthesis was significantly diminished due to MPA action.

4.3. Impact of RECQL5 expression on the yeast genome stability

In homologous recombination (HR), one of the first steps is the binding of Rad52 to the site of DNA breakage. This initial phase can be monitored with fluorescence microscopy and has been widely used to characterize the impact of mutations on genomic stability as a conssecuence of recombinogenic DSBs (Lisby *et al.*, 2001; Alvaro *et al.*, 2007). Taking advantage of this approach, we confirmed that overexpression of either RECQL5 or RECQL5 ID in a wild-type strain caused an increase in the accumulation of Rad52-YFP of about 4-fold above the control carrying the empty vector **(Figure R30A)**. This increase was dependent on the expression level of the helicase, since the increment was only of 2-fold when RECQL5 was expressed from the *tet* promoter (**Figure R30C**, green bars).

Taking into account the hypothesized roles of RECQL5 in a physiological environment, we next asked whether the helicase could be generating this phenotype of instability by means of a blockage in the replication machinery or a discoordination of transcription. Many transcription mutants, extensively studied in our laboratory such as *hprl1*Δ, accumulate DNA damage due to errors in the replication process derived from transcription failures. One of the main sources of these instability events is the formation of DNA:RNA hybrids, whose presence is evidenced by a Rad52 foci accumulation that is suppressed by over expression of RNase H, which degrades the RNA strand of a DNA:RNA hybrid. Alternatively, hybrids can be inferred by expressing the human cytidine deaminase AID, which exploits the presence of the displaced single stranded DNA of an R-loop, which is used as substrate and will lead to further increase of RNase H-dependent Rad52

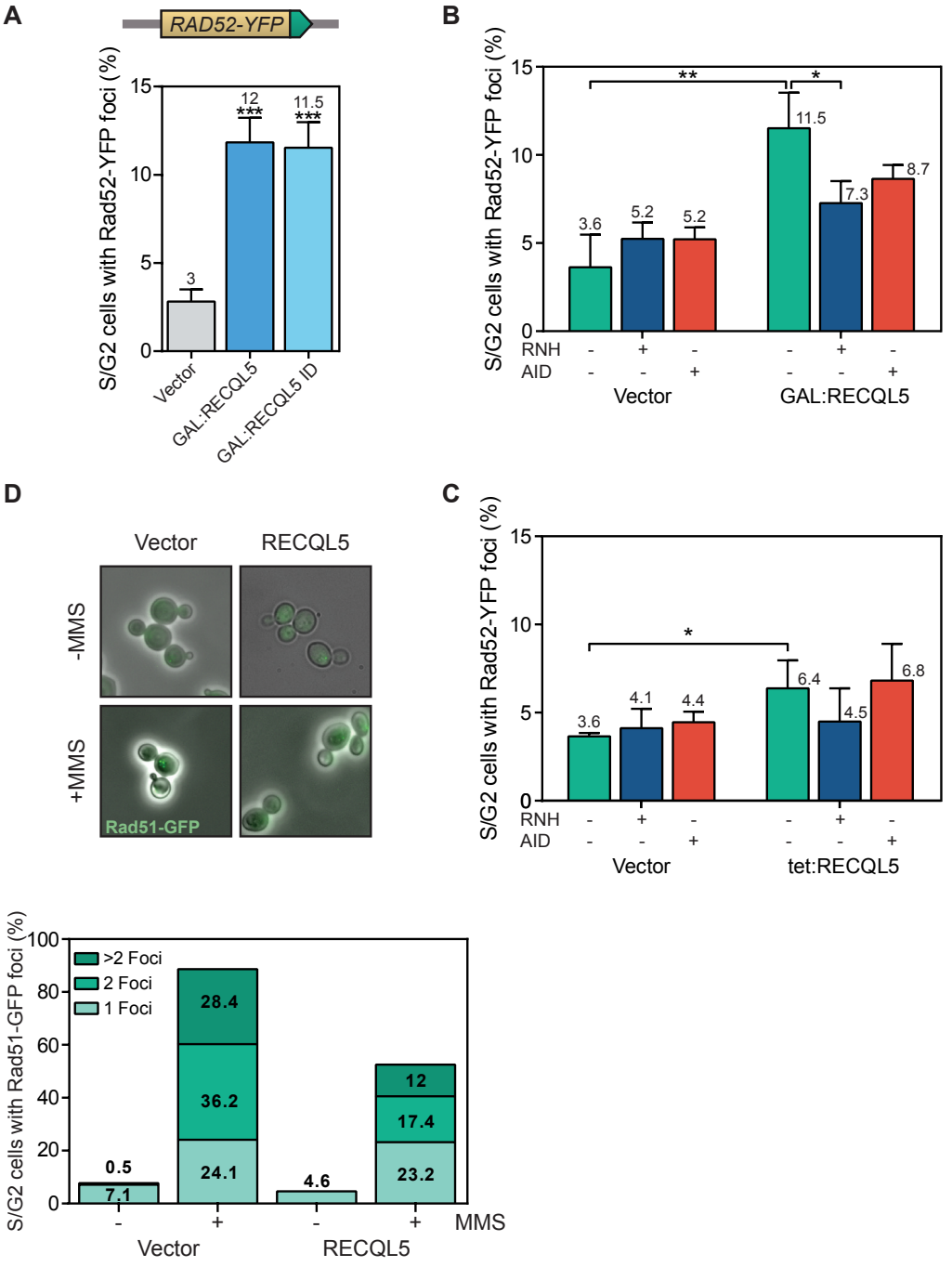


Figure R30. Genetic instability caused by RECQL5 expression in S. cerevisiae.

(A) Spontaneous Rad52-YFP foci formation in the W303.1A strain transformed with pYES3 (GAL1p) expression vectors. Representative images are shown. (B) Spontaneous Rad52-YFP foci formation in W303.1A strain transformed with pYES3 (GALp) or pCM184 (tetp). (AID). (D) Rad51-YFP foci formation in strain W303.1A transformed with pCM189 (Vector) or pCM189:RecQL5 (RecQL5). Cells were cultured to exponential state and MMS (0.01%) was added and incubated for two additional hours. Average of at least three independent experiments and representative images are shown. *, p≤0.05; **, p≤0.01; ***, p≤0.005. (Student's t-test).

foci. Since there were no consistent differences in the phenotypes when expressing the wild-type or the ID mutant version of RECQL5, we only assayed the effect of the full wild-type RECQL5. The results obtained by the simultaneous expression of the human helicase and either RNase H or AID gene did not allow to conclude that R-loops were the main source of instability generated by RECQL5 expression in yeast. Although a significative suppression of Rad52 foci formation was observed when overexpressing RNH1 in *GAL*:RECQL5 transformed cells **(Figure R30B)**, this effect was not statistically significant with lower expression levels of RECQL5 produced from the *tetp* construct **(Figure R30C)**. Co-expression of AID did not have the predicted additive effect over the instability as previously described in R-loop prone mutants (Gómez-González & Aguilera, 2007), failing to further support the hypothesis of R-loop accumulation in yeast cells expressing the human helicase.

It has been previously described that RECQL5 posses the capability to displace Rad51 filaments (Shimamoto *et al.*, 2000). These filaments are formed during HR pathway as part of the mechanism of strand exchange required in recombinational repair. Failure in Rad51-dependent strand exchange would render the cells unable to efficiently repair DSBs. For this reason, we decided to study if expression of the human RECQL5 helicase could interfere with the strand exchange step of HR. We transformed wild-type yeast strain W303.1A with either the empty plasmid pCM184 or the same backbone carrying the RECQL5 gene, and grew the cells in selective media with or without a low concentration of MMS (0.01%) in order to induce a mild replicative stress. As can be observed in **Figure R30D**, the results evidenced a lower accumulation of Rad51 foci in cells expressing RECQL5 compared with the wild-type control. This result suggests that the increased sensitivity of RECQL5-expressing cells to different genotoxic agents and the increase in Rad52 foci accumulation may be due to a capacity of RECQL5 to counteract Rad51 action or the HR process itself, either directly or indirectly.

We next determined the impact of the expression of RECQL5 on HR in yeast cells. To assay whether or not the accumulation of breaks resulted in an increase of

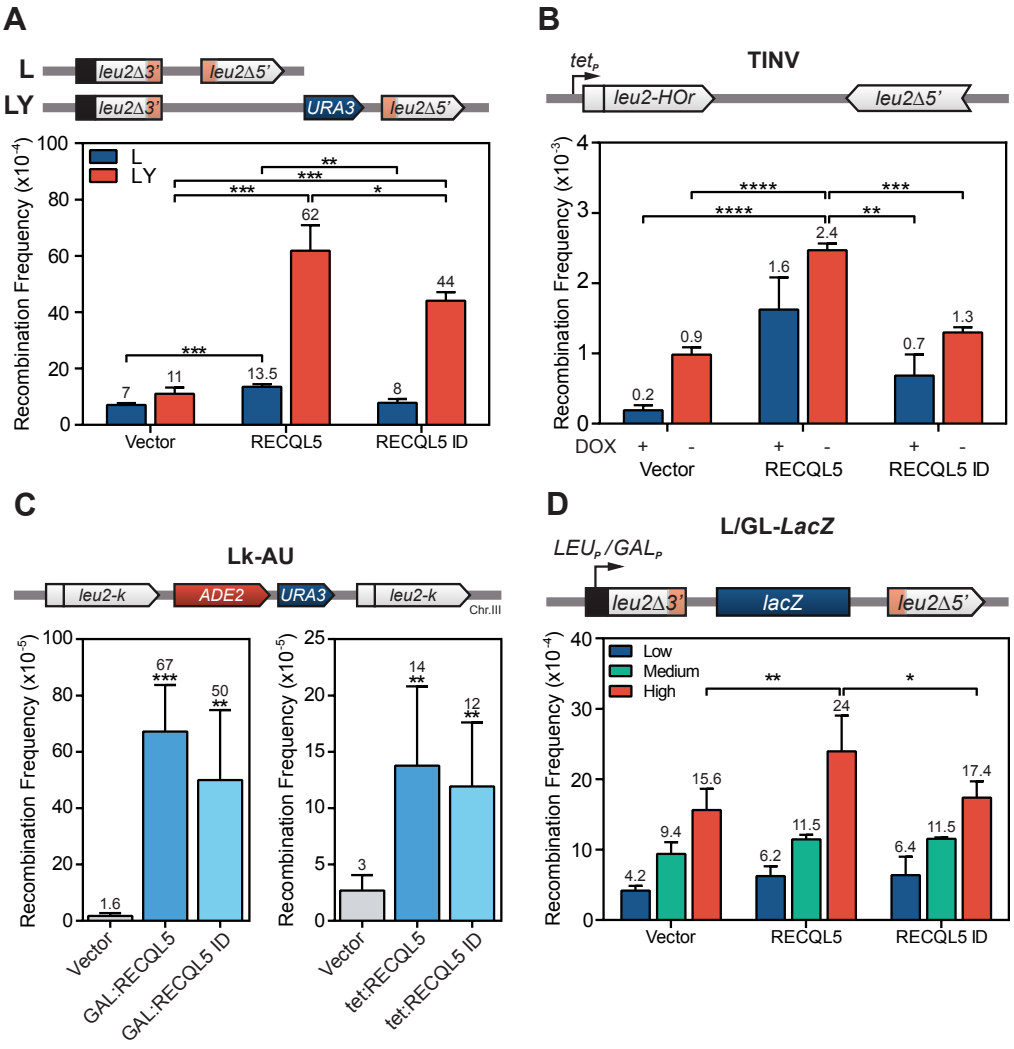


Figure R31. Effect of RECQL5 overexpression on S. cerevisiae recombination.

(A) Recombination analysis in W303.1A cells transformed with the L or the LY plasmid systems and either the empty vector pYES3 (Vector) or carrying RECQL5 or RECQL5 ID genes. (B) Recombination analysis in the AYW3-1B strain transformed with the indicated plasmids. (C) Transcription dependent recombination frequency in W303.1A cells transformed with either the plasmid pCM189 (Vector) or the constructs pCM189:RECQL5 (RECQL5) or pCM189:RECQL5ID (RECQL5 ID) using the plasmid-borne direct-repeat system GL-lacZ, expressed under the control of the GAL1 promoter in glucose or galactose (low and high transcription levels, respectively) and the L-lacZ system expressed under the control of the LEU2 promoter (medium transcription level). (D) Recombination analysis in cells carrying the TINV plasmid system, in which transcription is constitutively active (-DOX) but can be repressed with the addition of doxycycline 10µg/ml (+DOX) to the media. Average and SD of at least three different fluctuation tests from six independent colonies each one are shown. *, p≤0.05; **, p≤0.01; ***, p≤0.005; ****, p≤0.001. (Student's t-test). Orange boxes represent LEU2 repeats.

spontaneous recombination we took advantage of the plasmidic systems L and LY (Prado & Aguilera, 1995) and the chromosomal system *leu2-k::ADE2-URA3::leu2-k* (Lk-AU) (Aguilera & Klein, 1989) (**Figure R31A**-B). These systems are based on truncated repeats of 0.6 kb of the *LEU2* gene located in direct orientation and separated by intervening sequences of different length. Recombination frequencies

were measured as the frequency of Leu+ colonies in the plasmid-born systems or as FOA resistant colonies in the chromosomal one (further information on M&M 4.8). The results obtained show a strong impact on the recombination frequencies associated with the expression of both wild-type RECQL5 and RECQL5 ID respect to the control cells trasnformed with the empty plasmid, although the values are slightly lower for those expressing the mutant version. The instability associated with the expression of RECQL5 was dependent on the length of the intervening sequence between the repeats, as seen in the analisys with the plasmid-born systems **(Figure R31A)**. They were also dependent on the expression levels of the human helicase, as shown with the Lk-AU system **(Figure R31B)**, in which both trancription of the RECQL5 alleles, driven by either the *GAL* or the *tet* promoters, increased the recombination frequency. These values were lower when RECQL5 was expressed from the *tet* promoter.

We wondered whether the instability observed could be directly associated with transcription. For this, we assayed recombination with the plasmid-born systems L-*lacZ* and GL-*lacZ*, which are comprised of two *leu2* direct-repeats separated by the long and G+C rich *lacZ* gene under the constitutive *LEU2* promoter (*LEU2p*) or the inducible *GAL1-10* promoter (*GALp),* respectively (Gómez-González and Ruiz *et al.*, 2011). Recombination assays were performed in low transcription conditions (*GAL1p* grown in 2% glucose plates), medium transcription levels (*LEU2p*) and high transcription levels (*GAL1p* in 2% galactose plates). The recombination frequency values measured did not allow us to conclude any direct realtionship between the hyperrecobination observed and any possible deffect in the transcriptionl process in *tetp*:RECQL5-expressing yeast cells **(Figure R31C)**. Same outcome was achieved when perfoming the test with *GALp* produced RECQL5 in the inverted repeats system pTINV (González-Barrera *et al.*, 2002) **(Figure R31D)**, that also allows the study under low (plates supplemented with doxycycline) or high transcription conditions (plates without doxycicline), thanks to the presence of a *tetp* controlling the expression of the plamid system. Both experiments demonstrated that the expression of the human helicase RECQL5 induced high levels of genome instability to a greater extent in the wild-type RECQL5-expressing cells compared to those

carrying the ID mutant. Despite RECQL5 inducing high recombination frequencies, these increases did not seem to be directly connected to the trasncriptional status of the recombination systems, because the number of recombinants obtained was always higher under all conditions for RECQL5-expressing cells, independently of the transcription activation state.

4.4. Effect of RECQL5 expression on the transcriptional process

RECQL5 was the solely member of the RecQ family of helicases that had been found to directly interact with the RNAPII to date (Aygün *et al.*, 2008; Popuri *et al.*, 2013). We had established that the human helicase was also interacting with the yeast RNA polymerase II, but we did not evaluate yet if this interaction was generating any effect on the yeast transcriptional process.

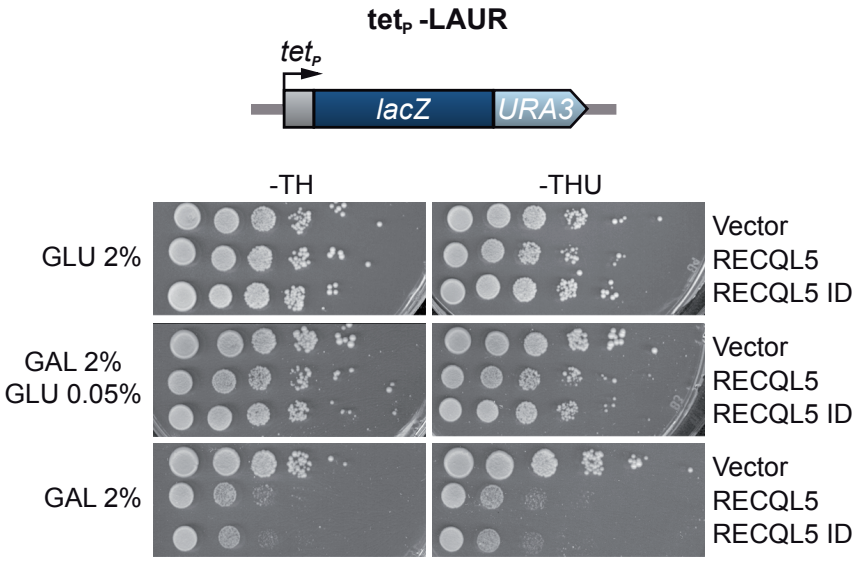


Figure R32. Effect of RECQL5 expression over S. cerevisiae transcription.

Top, scheme of the tetp::LacZ-URA3 translational fusion construct used for the analysis of transcription. Bottom, Growth analysis of the W303.1A strain transformed with either the plasmid pYES3 or the same vector carrying the RECQL5 and the RECQL5 ID genes. 10-fold serial dilutions plated on selective media with the indicated amounts of glucose or galactose. Photographs taken after 3 days of incubation at 30ºC.

We initially assessed the transcription proficiency of RECQL5-expressing yeast by taking advantage of the pLAUR system (Jimeno *et al.*, 2002). This is comprised of a transcriptional fusion *lacZ-URA3* under the control of the *tet* promoter. This construct does not allow distinguishing between initiation or elongation defects. It permitted to detect failures of the transcriptional process through the evaluation of the cell capability to grow on synthetic media plates without added uracil. Wild-type (W303.1A) transformed yeast with plasmids carrying either RECQL5 or RECQL5 ID were able to grow without any noticeable difference in plates with or without added uracil **(Figure R32)**. This result suggested that the presence of RECQL5 was not interfering with transcription, neither in the wild-type nor the ID mutant, at least to a detectable level in this assay, which requires a strong phenotype in order to distinguish between mild effects on transcription.

To bring the analysis one step further we moved to a molecular approach. We studied the kinetics of promoter activation and mRNA production in cells expressing RECQL5 helicase. The *GAL1-10* chromosomal locus is a well characterized region of the yeast genome, widely used as model in our group for transcription studies (Chávez & Aguilera, 1997; Piruat & Aguilera, 1998; Chávez *et al.*, 2000; Rondón *et al.*, 2003). Wild-type W303 yeasts, transformed with pYES2 plasmids carrying or not the human RECQL5 or the mutant RECQL5 ID, were cultured in synthetic media supplemented with raffinose until exponential. Then, galactose was added to a final concentration of 2% and samples were taken at the indicated times for whole RNA extraction and northern blot analysis. Results in **Figure R33** show that the promoter activation process was mostly unaffected but, interestingly, a larger mRNA than predicted was clearly accumulated in RECQL5 and RECQL5 ID-expressing yeasts. These larger mRNAs corresponded to a read-through transcript and are usually generated when an inefficient transcriptional termination of the upstream gene invades the neighbouring ORF inhibiting the initiation of transcription on the downstream gene (Shearwin *et al.*, 2005). Thus, read-through transcripts are made of the fusion of the gene targeted by the probe used in each experiment plus the added sequence of the following gene, *GAL1+FUR1* or *GAL10+GAL7* in our experiments **(Figure R33)**. Transcriptional interference may occur when an

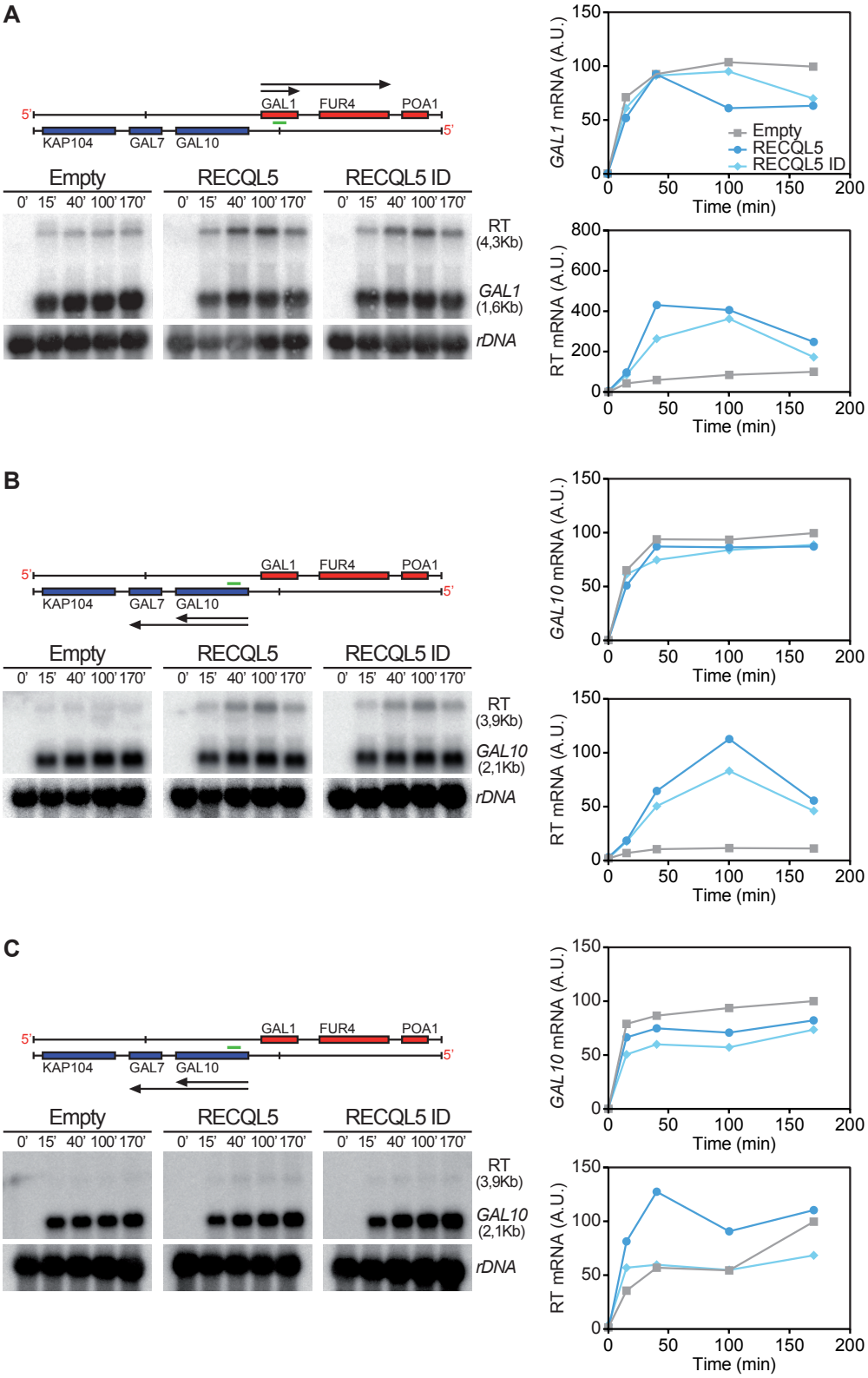


Figure R33. RECQL5 expression impact on active transcription at the GAL locus.

(A) Northern blot analysis of the GAL1 and (B) GAL10 mRNAs in W303.1A strain transformed with pYES2 expression vector carrying the indicated version of the RECQL5 gene. (C) Northern blot analysis of GAL10 mRNA in W303.1A strain transformed with pCM184 expression vectors carrying the indicated version of the RECQL5 gene. Schematic representation of the mRNAs detected and probe location (green bar) are shown. Graphs represent corresponding mRNA signal relative to rDNA. Mean of two experiments is shown.

elongating polymerase fails to terminate properly entering into a co-transcribed downstream transcription unit, reducing its promoter activity. Transcriptional interference also has been documented to affect the proper use of both ARS and CEN elements inhibiting either replication or chromosomal segregation in mitosis.

Our results evidence that both, the wild-type RECQL5 and the ID mutant, cause the same effect with comparable strength in the locus studied (**Figure R33A**-B), and this effect is also detected when the expression was driven from the *tet* promoter **(Figure R33C)**.

The *GAL7* and *GAL10* genes are part of the *GAL* gene cluster of *S. cerevisiae* genome, required for the metabolic pathway of galactose conversion into glucose-6-phosphate (Fridovich-Keil & Jinks-Robertson, 1993). Both genes are arranged in tandem orientation and their high transcription rate facilitates that polymerases continue transcribing after the *GAL10* poly(A) into *GAL7* producing long read-through transcripts, resulting in bi-cistronic mRNAs (Greger & Proudfoot, 1998). Promoter occlusion of *GAL7* produced by this mechanism abolishes the synthesis of Gal7 protein, rendering the cells inviable since *GAL7* is essential in the presence of galactose (Douglas & Hawthorne, 1964). For this reason, we wondered whether the viability defect observed when expressing RECQL5 from the *GAL1* promoter was due to the read-through effect. As seen in **Figure R34**, wild-type cells simultaneously overexpressing Gal7 protein and RECQL5 grew better than those

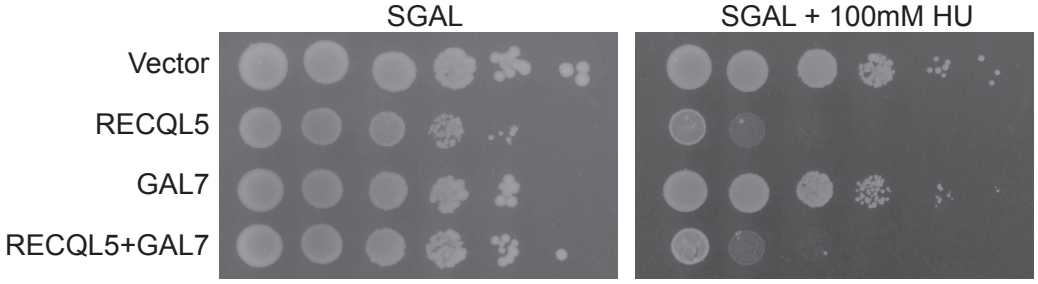


Figure R34. GAL7 deficit is not responsible for yeast sensitivity to genotoxic agents.

Viability test with W303.1A strain transformed with GAL1 inducible vectors carrying the RECQL5 and/or the GAL7 genes. Serial dilutions (1:10) of exponentially growing cultures spotted on plates. Photographs were taken after 3 days of growth at 30°C.

expressing only RECQL5 in a synthetic medium with galactose as the sole carbon source, demonstrating that the human helicase was interfering with the expression of the endogenous *GAL7* gene, but this phenomenon was not responsible of diminishing the sensitivity to genotoxic agents like HU, as cell viability appeared compromised to the same degree in RECQL5-expressing yeast cells whether or not overexpressing *GAL7*.

Aware that RECQL5 presence in *S. cerevisiae* cells had the capacity to alter the expression of some genes, we decided to perform a genome wide approach to search for transcripts whose levels would be affected and could explain the instability phenotypes observed.

To achieve this, wild-type strain W303.1A was transformed with the plasmid carrying RECQL5 gene under the control of the *tet* promoter. RECQL5-expressing cells under this promoter only showed visible phenotypes in the presence of genotoxic agents **(Figure R29)**, by studying genetic instability (**Figure R30**, **Figure R31**) or by molecular biology approaches **(Figure R33)**. Therefore, we hypothesized that RECQL5 could be affecting the expression of some important genes to overcome stress, but unless under this situation, the changes were not enough to explain the effects on viability. To that end, cells were cultured in SC media and recovered in mid-log. Whole RNA extractions were performed in three independent RECQL5-expressing transformants and one as control transformed with the empty vector. Total mRNA was determined using the GeneChip Yeast Genome 2.0 Array (Affimetrix). Control values were calculated averaging the data from our empty vector transformed cells with two other similar wild-type analyses from GEO accession GSE22644 (Tu *et al.*, 2011) (Pearson's R correlation coefficient of 0.914). Gene expression profiles of the control and RECQL5-expressing cells were quite similar (Pearson's R correlation coefficient of 0.972), suggesting that RECQL5 presence was not significantly impacting global expression levels in *S. cerevisiae*. To remove false positives, out of 5716 genes analysed in the microarray, we excluded as non-expressed those with values below the median of meiotic genes of the control experiments. Under this criteria, only 460 genes presented mRNA levels altered at least 1.5-fold respect to the control, and just 152 genes were

selected for analysis after passing the FDR (False Discovery Rate) *p*-value<0.1.

We found that the 58 genes that were up-regulated represented a group of significantly long genes (around 1kb longer than the genome median) and that their G+C content was slightly but significantly lower than the median value of the genome. Although up-regulated genes were overall highly expressed, this value had no statistical significance. On the other hand, the 98 down-regulated genes in RECQL5-expressing cells had very low transcription levels and were also characterized by a higher content in G+C respect to the genome median value, and slightly shorter in length than the genome median **(Figure R35)**. Gene Ontology (GO) analyses of up-regulated transcripts, performed using the yeastmine.org service, evidenced an enrichment just in ribosome biogenesis related genes (18 out of 58 genes, P-value 0.02003, GO:0042254). On the other hand, down-regulated transcripts appeared to be enriched just in metabolic genes, particularly related to the lysine biosynthesis pathway (11 out of 98, *p*-value 3.5251e-15, GO:009085).

Altogether, this analysis did not allow us to establish a direct association between the expression of RECQL5 in *S. cerevisiae* and a severe interference in the transcription process. The main reason behind this conclusion is the short number

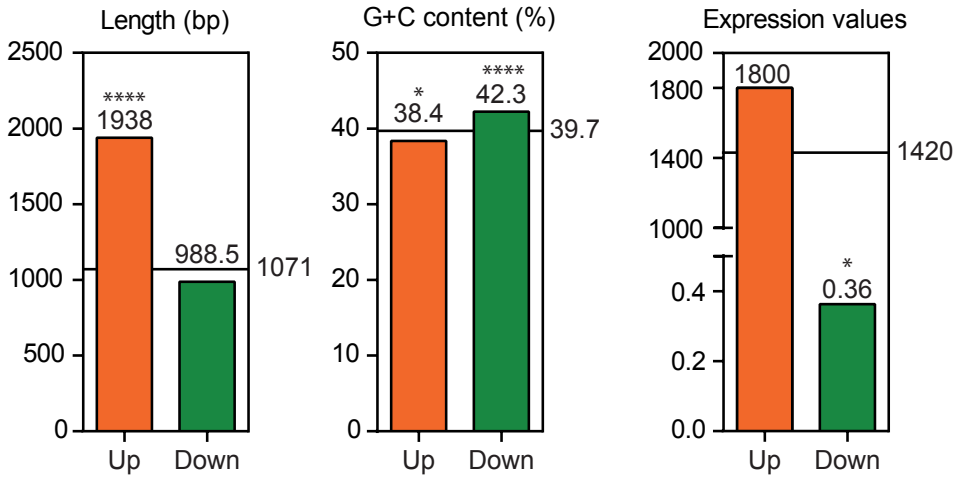


Figure R35. Structural and functional features analysis of deregulated genes in RECQL5 expressing yeast.

Statistical analysis of length, G+C content and model-based expression values of genes whose expression levels are significantly altered by RECQL5 expression in yeast. Median values are shown and line represents the genome median. *, p<0.05; **, p<0.01; ***, p<0.001. (Mann-Whitney's U-test)

of genes whose expression was significantly altered. However, by carefully analysing the lists of hits, we were able to find interesting genes with modified mRNA levels in RECQL5-expressing cells such as the ones compiled in **Table R5**. The fold change in none of the cases is enough to explain by itself the instability phenotypes, but could be indicative that RECQL5 affects transcription of a variety of genes, either as a response for possible damages that the helicase could be generating, or by directly impacting on their transcription.

Table R5. Genes altered in RECQL5-exrpressing cells

Up-Regulated genes				
Linear Fold	**Pvalue FDR**	**ORF**	**Alias**	**Function**
1.99	0.022	YOR346W	REV1	Deoxycytidyl transferase; involved in repair of abasic sites and adducted guanines in damaged DNA by translesion synthesis (TLS); forms a complex with the subunits of DNA polymerase zeta, Rev3p and Rev7p;
1.62	0.061	YLR430W	SEN1	Presumed helicase and subunit of the Nrd1 complex (Nrd1p-Nab3p-Sen1p); complex interacts with the exosome to mediate 3' end formation of some mRNAs, snRNAs, snoRNAs, and CUTs; homolog of Senataxin which causes Ataxia-Oculomotor Apraxia 2 and a dominant form of amyotrophic lateral sclerosis
1.6	0.054	YKR086W	PRP16	DEAH-box RNA helicase involved in second catalytic step of splicing; exhibits ATP-dependent RNA unwinding activity; mediates the release of Yju2p and Cwc25p in the second step; in the absence of ATP, stabilizes the binding of Cwc25p to the spliceosome in the first catalytic step
1.56	0.047	YJL092W	SRS2	DNA helicase and DNA-dependent ATPase involved in DNA repair and checkpoint recovery, needed for proper timing of commitment to meiotic recombination and transition from Meiosis I to II; blocks trinucleotide repeat expansion; affects genome stability
1.56	0.048	YDL031W	DBP10	Putative ATP-dependent RNA helicase of the DEAD-box protein family, constituent of 66S pre-ribosomal particles; essential protein involved in ribosome biogenesis

Down-Regulated genes				
Linear Fold	Pvalue FDR	ORF	Alias	Function
3.42	0.0028	YML116W	ATR1	Multidrug efflux pump of the major facilitator superfamily, required for resistance to aminotriazole and 4-nitroquinoline-N-oxide
1.91	0.029	YER139C	RTR1	CTD phosphatase; dephosphorylates S5-P in the C-terminal domain of Rpo21p; has a cysteine-rich motif required for function and conserved in eukaryotes; shuttles between the nucleus and cytoplasm
1.64	0.087	YOL086W-A	MHF1	Protein of unknown function; ortholog of human MHF1, and a component of the heterotetrameric MHF histone-fold complex that in humans interacts with both DNA and Mph1p ortholog FANCM, a Fanconi anemia complementation group protein, to stabilize and remodel blocked replication forks and repair damaged DNA; mhf1 srs2 double mutants are MMS hypersensitive
1.62	0.0543	YOR230W	WTM1	Transcriptional modulator involved in regulation of meiosis, silencing, and expression of RNR genes; required for nuclear localization of the ribonucleotide reductase small subunit Rnr2p and Rnr4p; contains WD repeats
1.56	0.0419	YBR095C	RXT2	Subunit of the histone deacetylase Rpd3L complex; possibly involved in cell fusion and invasive growth

4.4.1. RECQL5 expression does not interfere with RNAPII mutations

As another approach to analyse the relationship of RECQL5 with the transcription of *S. cerevisiae*, we decided to study the possible interactions in the case that the RNAPII holoenzyme was mutated. Two mutants of the major subunit Rbp1p, *rpb1-1* and *rpb1S751F*; another point mutant in the second subunit, *rpb2-10* and a complete deletion of one of the 12 subunits, the *rpb9*Δ mutant (Felipe-Abrio *et al.*, 2014). Because these four mutants are mainly characterized for presenting a variety of problems in the transcription elongation, we wondered if the effect of RNAPII stabilization observed in RECQL5-expressing cells as the accumulation of read-through transcripts **(Figure R33)** would help these strains to overcome their phenotypes. To answer this, we transformed these mutants with plasmids carrying or not the *GAL::*RECQL5 construction and plated serial dilutions of exponentially growing cultures on induction plates with galactose.

After four days of incubation at 30°C, the viability assay shown in **Figure R36** revealed some moderate differences in the mutants response to the to the overexpression of RECQL5 compared to the wild-type strain. The heat sensitive *rpb1-1* mutant showed a behaviour similar to that of the wild-type strain, revealing only a minor difference respect to the induction plate containing glucose, in which mutant cells expressing RECQL5 seemed less affected than wild-type cells. The mutant *rpb2-10,* either transformed with the empty plasmid or with the one expressing the RECQL5 helicase, grew similarly bad in all conditions, possibly as consequence of a poor transcription of the RECQL5 gene, this could also apply to the *rpb9*Δ mutant, which did not reflect any difference viability impact after RECQL5 expression. Notably, a slight improvement in the viability of the mutants *rpb1S751F* and *rpb9*Δ carrying the RECQL5 gene was observed respect to their control, maybe explained by a low expression of RECQL5 in glucose.

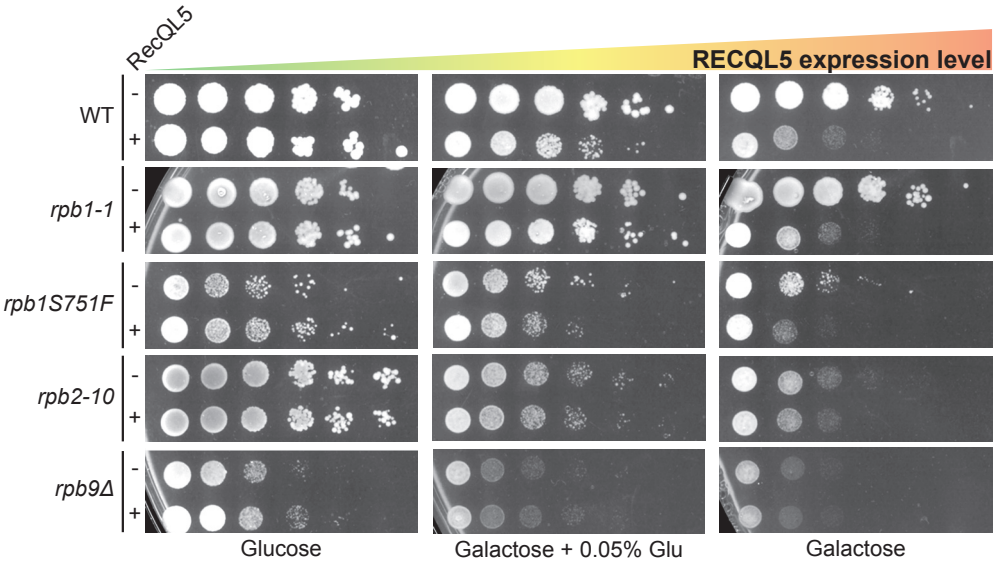


Figure R36. Effect of RECQL5 expression in RNAPII mutants.
Viability assay with WT (W303.1A), rpb1-1 (WRP1-12A), rpb1S751F (WSR8-5A), rpb2-10 (WRP2) and rpb9Δ (WRP9-3C) strains transformed with expression vector pYES2:RECQL5.10-fold serial dilutions of exponentially growing cultures plated in selective media with indicated concentrations of carbon source to regulate the expression of the protein. Photographs were taken after 4 days of growth at 30°C.

4.5. Impact of RECQL5 helicase on cell cycle progression

We know that transcription is one of the major sources of recombinogenic DNA damage (Gaillard & Aguilera, 2016). This threat is mostly derived from the encounter of transcriptional and replication machineries that usually lead to impairment of the latter process. The proposed role of the human helicase RECQL5 was to regulate transcription in the event of collisions with the replication machinery (Popuri *et al.*, 2013; Saponaro *et al.*, 2014). In our model, RECQL5-expressing yeasts showed elevated levels of genomic instability but subtle defects in the transcription, although we demonstrated a positive interaction with the transcription machinery. Therefore, we wondered whether RECQL5-associated impact in the *S. cerevisiae* genomic stability could be more related to replication fork stalling instead of transcription-associated defects. For this, We transformed

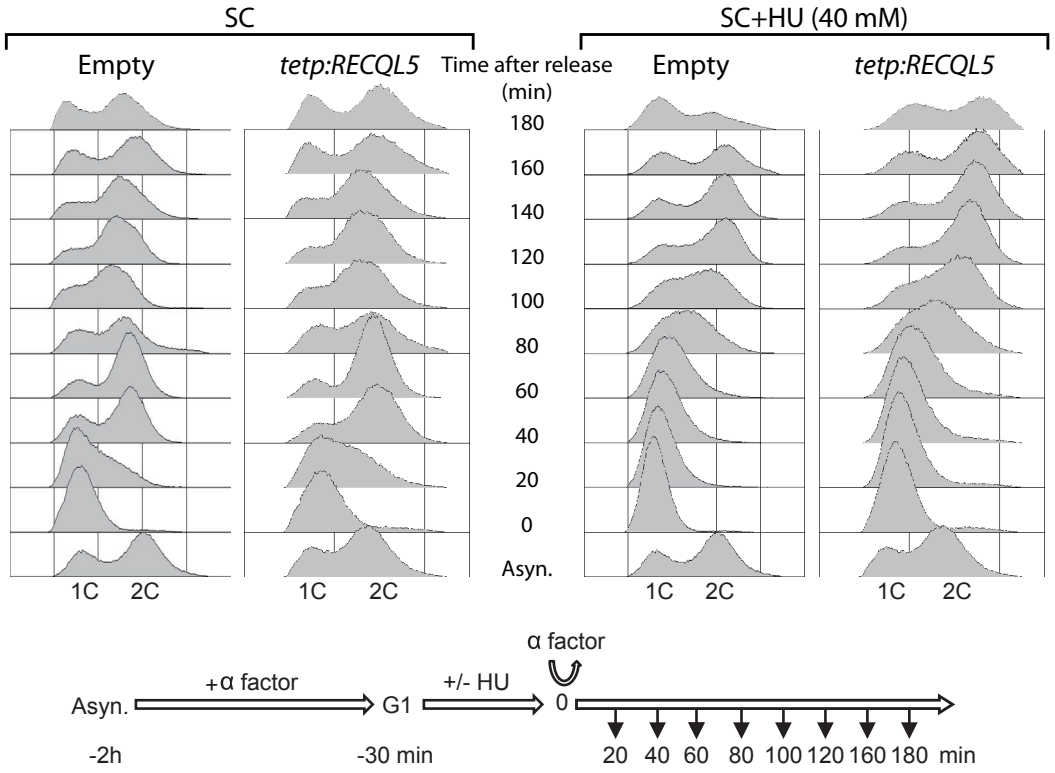


Figure R37. RECQL5 expression does not lead to replication impairment.

FACS analysis of S-phase progression of wild-type strain (W303.1A) transformed with indicated plasmids based on pCM189 backbone. Cells were synchronized in G1 with α-factor and released in presence (right) or absence (left) of 40 mM hydroxyurea by changing the culture medium. A diagram of the experiment is shown at the top.

the wild-type strain W303.1A with the *tet*:RECQL5 construct or its empty backbone plasmid and monitored their S-phase progression in the presence or absence of HU by FACS **(Figure R37)**. After α-factor synchronization, cells were released from G1 into fresh medium supplemented or not with 40mM of HU, in order to slow down replication progression and also, to slightly stress the cells in order to manage a clearer RECQL5-associated phenotype. As show in the result, we found that cells expressing RECQL5 did not present any delay in S-phase progression neither in the absence or the presence of the assayed concentration of HU. The only observable difference was at 180 minutes after α-factor release, when cells were under replication stress (+HU condition). At this point RECQL5-expressing yeasts showed a slight arrest in G2-phase compared to the strain transformed with the empty plasmid. Therefore, it seems that the instability phenotypes associated to RECQL5 expression in yeasts were not accompanied by a strong replication impairment.

4.6. Genome-wide distribution of RECQL5 through the *S. cerevisiae* genome

To explore possible mechanisms of action for RECQL5 on *S. cerevisiae*, we performed ChIP-chip analysis using the RECQL5-FLAG tagged protein. The data were analysed and compared to Rpb3-HA and Hprl-FLAG ChIP-chip results previously published (Gómez-González *et al.*, 2011) **(Figure R38)**. Statistical analysis of the IP signal allowed to identify regions or clusters where RECQL5 recruitment was statistically significant (genome regions where enrichment signal difference over the input had a *p*-value lower than 0.05). These positive intervals were distributed all over *S. cerevisiae*'s genome.

Overall, the normalized signal values obtained for RECQL5 recruitment were slightly lower to those acquired for Rpb3 and Hprl. Annotation of RECQL5 enriched regions (SGD_features.tab for sacCer3 genome version) showed that half (51.8%) of the yeast features where RECQL5 could be found were ORFs. This number was slightly below compared to that of yeast proteins directly associated with transcription like Rpb3 or Hprl (66.9% and 65.6% of total hits mapping to

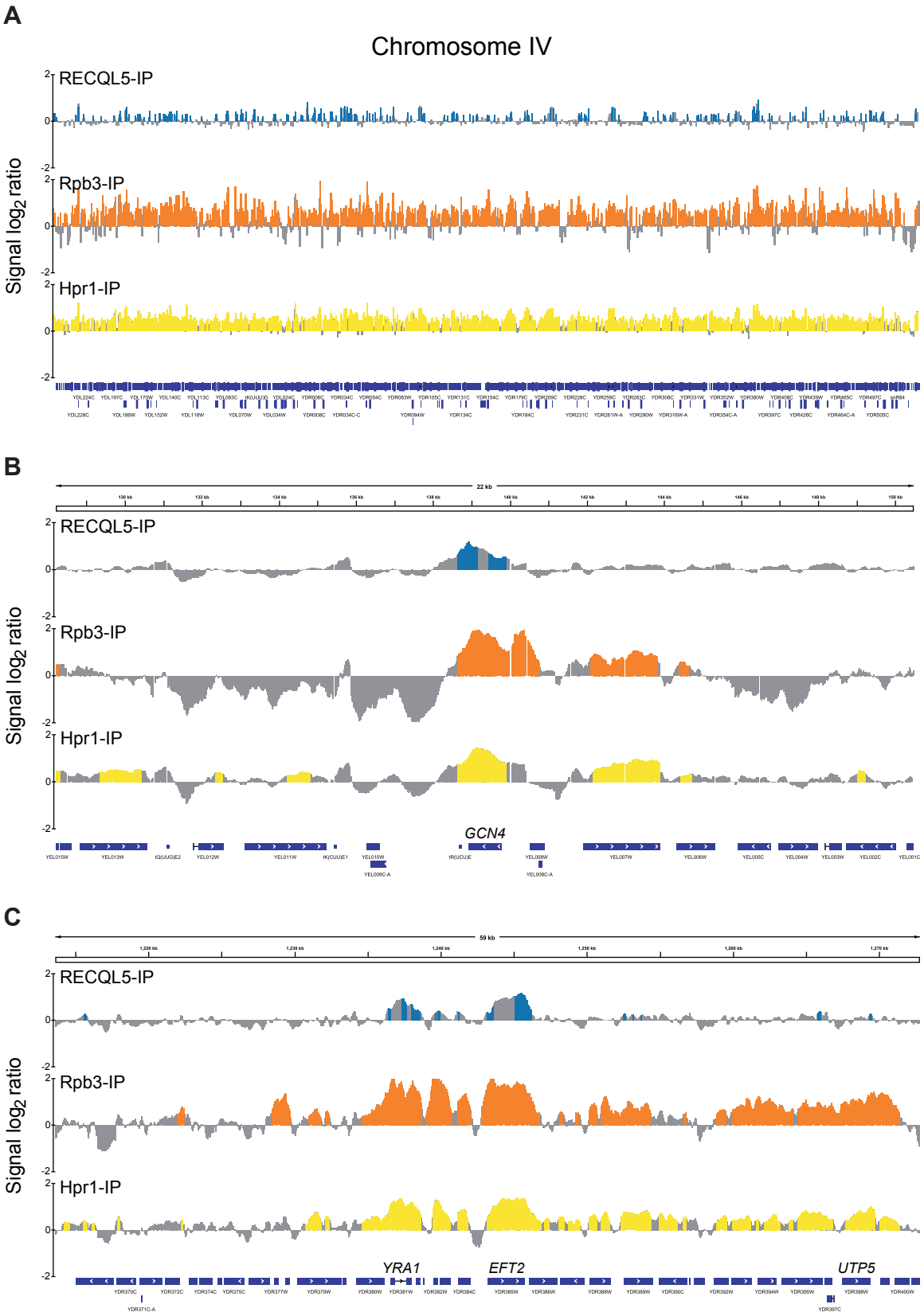


Figure R38. RECQL5 recruitment to the genome.

Genomic views of RECQL5 (blue), Rpb3 (orange) and Hpr1 (yellow) recruitment. (A) Full view of chromosome IV. (B) GCN4 ORF and adjacent 10 kb of the surrounding region. (C) YRA1 locus and surrounding 30 kb are shown. Signal log2 ratio values are plotted. SGD ORFs are represented below as blue bars and white arrows according to the direction of transcription.

ORFs, respectively). Whereas RECQL5 mapped to less ORFs, 79.5% of these were actively transcribed genes (Fig 13B). Besides the fact that we also found RECQL5 bound to other non RNAPII-transcribed regions (i.e. tRNAs, introns, or snRNAs and snoRNAs), RECQL5 occupancy profile showed differences to those of RNAPII or Hpr1. Interestingly, RECQL5 was found in 9 out of the 16 yeast centromeres and, proportionally to the total number of regions identified as positive, the binding to telomeres (1.45% of total RECQL5 positive regions) was higher than in Rpb3-IP or Hpr1-IP (0.76 and 1.18, respectively). These numbers support the hypothesis that the human helicase binding to the yeast chromatin was not exclusively dependent of transcription and suggest the possibility that RECQL5 could be interacting with a number of yeast factors **(Table R6)**.

Table R6. Genomic features mapped by different ChIP-chip binding clusters

	Genome	RECQL5	RPB3	HPR1
ORFs	6607	1296	3255	4324
Centromeres	16	9	0	0
Telomeres	32	27	22	30
ARS	337	35	37	78
Introns	376	73	144	136
snRNAs/snoRNAs	83	24	81	35
Retrotransposons	50	35	46	50
ncRNA	15	1	5	8
rRNA	27	0	25	25
tRNA	299	16	13	50
Total Hits		2504	4865	6592

To determine if ORFs identified to accumulate RECQL5 shared common characteristics, we analysed the structural and functional features of the 561 RECQL5-enriched genes found. Statistical analyses of their median revealed that RECQL5 recruited to long and highly expressed genes, but not rich in G+C content compared to the genome median **(Figure R39C)**. The GO analysis revealed an overrepresentation of genes associated with ribosome biogenesis and RNA processing, in agreement with the results obtained from the analysis of UP-regulated genes in the Expression Array. Although only 14 RECQL5-erniched genes were coincident out of the 58 classified as up-regulated genes in the Expression Array,

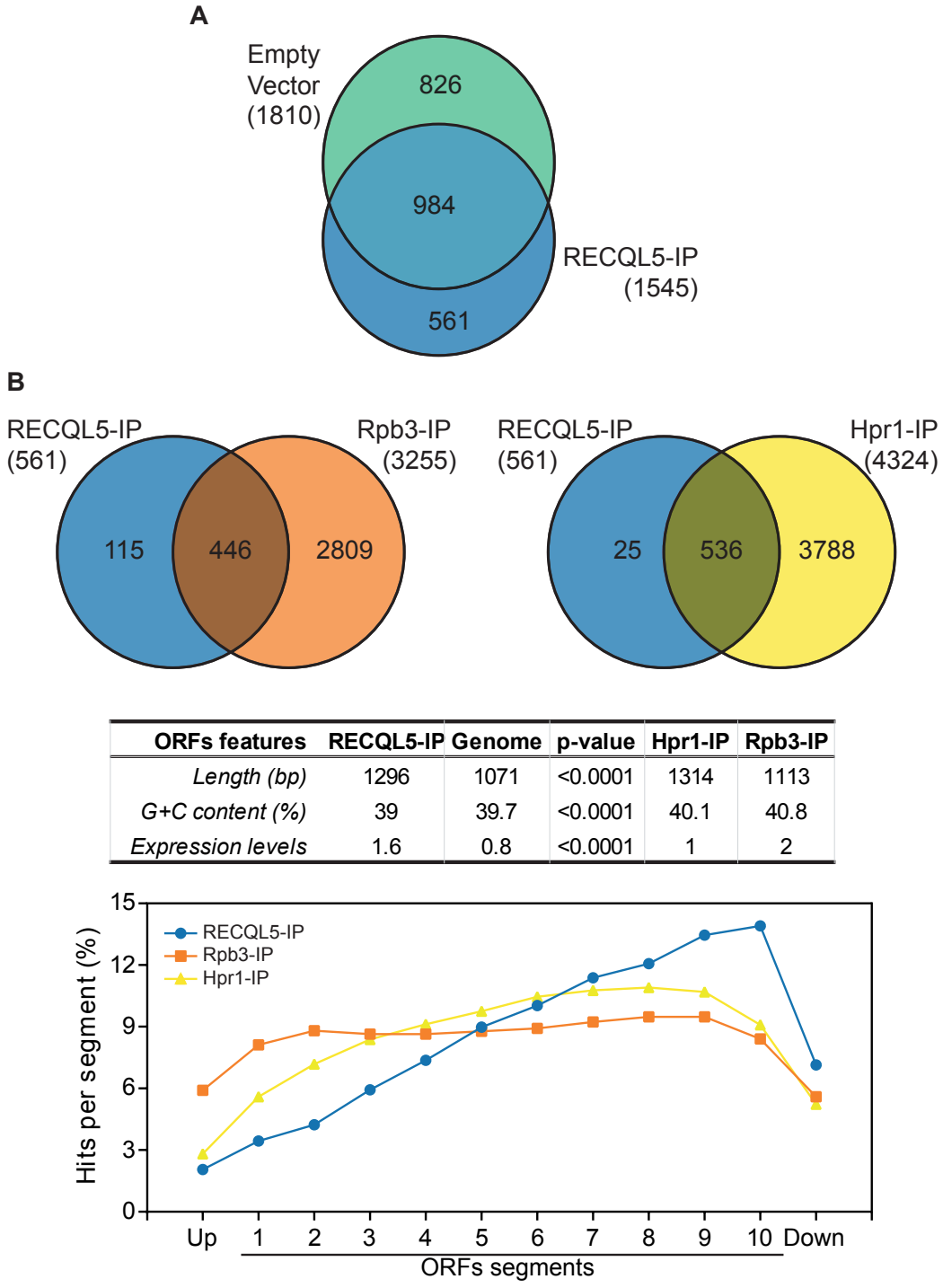
this result is consistent with those genes being highly transcribed in yeast.

We next analysed the relative distribution of RECQL5 along the length of the genes, as explained below. Selected as positive genes were subdivided into 10 segments independent of their length, and additional upstream and downstream segments of the same size were included. The percentage of RECQL5-entriched clusters mapping on each segment determined the occupancy profile of the human helicase along the ORFs. Interestingly, the results showed that RECQL5 distribution increased gradually toward the 3' end of the genes **(Figure R39D)**, similar to Hprl and other transcription factors previously related to elongation (Gómez-González *et al.*, 2011). The low occupancy observed at the promoter region and the beginning of the ORF, and the clear accumulation toward the 3' end of the gene is consistent with previous analyses in human cells, where the interaction of RECQL5 is shown to be dependent on the phosphorylation of the CTD tail of an elongating RNAPII (Kanagaraj *et al.*, 2010).

4.7. Yeast mutants resistant to RECQL5 expression

Our results evidenced that expression of the human helicase RECQL5 impacted *S. cerevisiae* genome stability. To try to understand how RECQL5 affects DNA metabolism in yeast, we screened for mutants that overcame the damage generated by the human helicase, whether or not in genes involved in the generation of RECQL5-dependent DNA damage.

We designed a reduced version of the KO collection of yeast deletion mutants (EUROSCARF). The selection and manual compilation of the mutants as well as the plate design work was performed by Francisco Garcia Benitez in our lab, who kindly provided a copy of the selected collection. It comprised 400 knock-out mutants out of the original 5105 strains, selected on the criteria that their functions were associated with DNA or RNA metabolism. Therefore, all selected strains were KO for genes related to replication, transcription, recombination or chromatin modification. We transformed the collection with plasmid pYES2:RECQL5 following the transformation protocol for 96-well plates detailed in Materials and Methods. The screening was performed growing the transformed strains in synthetic media



ORFs features	RECQL5-IP	Genome	p-value	Hpr1-IP	Rpb3-IP
Length (bp)	1296	1071	<0.0001	1314	1113
G+C content (%)	39	39.7	<0.0001	40.1	40.8
Expression levels	1.6	0.8	<0.0001	1	2

Figure R39. RECQL5 recruitment to ORFs

(A) Overlap between enriched genes identified in the control sample with an empty plasmid, and the annotated genes in the RECQL5-IP analysis. Subsequent analyses were performed only with positive hits unique to RECQL5-IP. (B) RECQL5-enriched genes overlap with similarly identified ORFs in Rbp3-IP or Hpr1-IP. (C) Statistical analysis of length, C+G content and expression levels of the RECQL5-enriched genes. Median values for RECQL5-IP, the genome Rpb3-IP and Hpr1-IP plotted. p-value calculated by Mann-Whitney's U-test. (D) Composite profile of RECQL5 and Hpr1 occupancy detected by ChIP-Chip across the average ORF plotted as RECQL5, Rpb3 or Hpr1 percentage of ChIP clusters per segment (see Materials and Methods).

plates containing either glucose 2% or galactose 2%, supplemented with 100mM HU or not. All plates included triplicate positons with wild-type strain (BY4741) transformed with the empty plasmid pYES2 or carrying RECQL5 gene as controls. This method allowed us to compare growth with the rest of the mutants in the same plates. **Figure R40** shows an example of the resulting plates. In it, key positions have been signalled to help identification with green circles, wild-type transformed with empty plasmid pYES2; orange circles, wild-type transformed with plasmid pYES2:RECQL5; red circles, candidates growing slower than RECQL5 expressing wild-type strain and light green circles, candidates growing better than RECQL5 expressing wild-type cells.

From this analysis we identified 37 mutants that appeared to grow better than the wild-type strain expressing RECQL5 in the presence of 100mM of HU, and 31 mutants that seemed to be hyper-sensitive to the expression of RECQL5 in synthetic media plates without HU. We elaborated a list with this category, but due to time and technical limitations we only kept working with the selected candidates

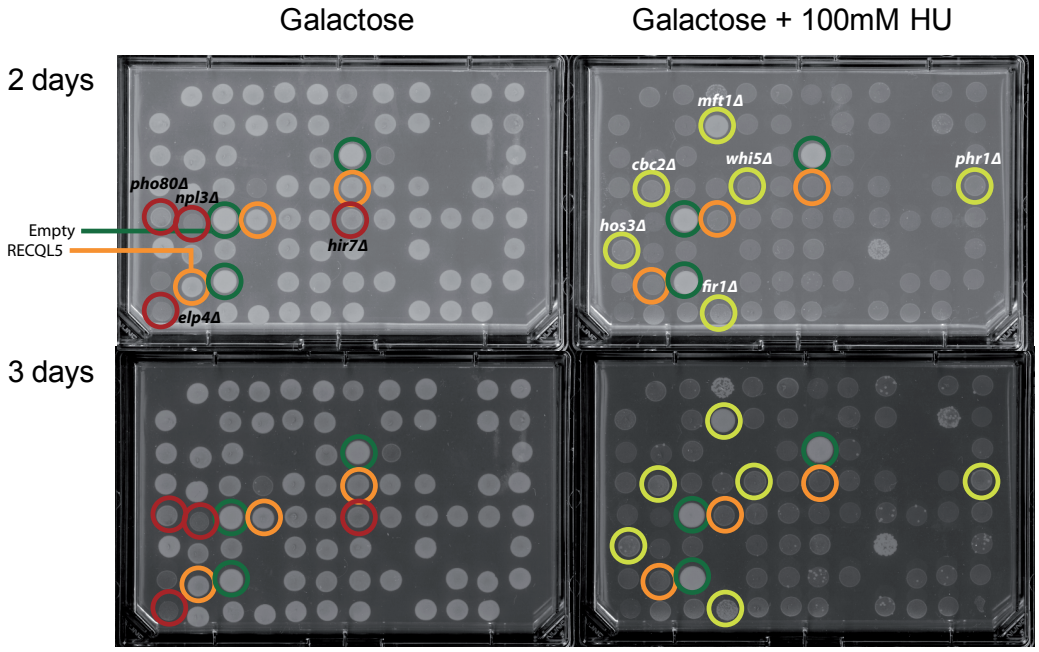


Figure R40. Example of plates used in the screening of selected KO-collection

Subset of mutants from the mini-KO collection, transformed with GAL:RECQL5 and grown in synthetic media plates with galactose and HU or not. Orange circles: Wild-type (BY4741) transformed with plasmid pYES2:RECQL5; Green circles: Wild-type transformed with empty plasmid pYES2; Red circles: Candidates growing slower than the WT expressing RECQL5; Yellow circles: Candidates growing better than wild-type cells expressing RECQL5.

from the initial goal of the screening. Next, we transformed each candidate strain, re-isolated from the original collection, with the plasmid carrying or not RECQL5 under the *GALp* promoter and compared the growth between cells expressing or not RECQL5. After this filter step, in which many candidates dropped off the list, we selected 4 genes whose deletion appeared to partially supress the viability defect generated by RECQL5 expression in presence of HU: MFT1, ISY1, SPT4 and SEM1 **(Table R7)**.

In principle, since all selected candidates were related to transcription, the results support the hypothesis that RECQL5 interference was mediated by hampering transcription. We then analysed the candidates by checking the expression levels of the human helicase, since it was possible that deletion of the gene candidates could affect the transcription of RECQL5. By taking a quick look

Table R7. Genes which mutation confers resistance to RECQL5 expression

Gene	Name	Function	
YML062C	**MFT1**	Subunit of the THO complex; THO is a nuclear complex comprised of Hpr1, Mft1, Tho2 and Thp2, that is involved in transcription elongation and mitotic recombination; involved in telomere maintenance	ELONGATION
YJR050W	**ISY1**	Member of the NineTeen Complex (NTC); NTC contains Prp19p and stabilizes U6 snRNA in catalytic forms of spliceosome containing U2, U5, and U6 snRNAs; interacts with Prp16 to modulate splicing fidelity; isy1 syf2 cells have defective spindles	SPLICING
YGR063C	**SPT4**	Component of the universally conserved Spt4/5 complex (DSIF complex); the complex has multiple roles in concert with RNA polymerases I and II, including regulation of transcription elongation, RNA processing, quality control, and transcription-coupled DNA repair; Spt4 also localizes to kinetochores and heterochromatin and affects chromosome dynamics and silencing; required for transcription through lengthy trinucleotide repeats in ORFs or non-protein coding regions	ELONGATION TCR
YDR363W-A	**SEM1**	Component of lid subcomplex of 26S proteasome regulatory subunit; involved in mRNA export mediated by THSC/TREX-2 complex (Sac3-Thp1); assumes different conformations in different contexts, functions as molecular glue stabilizing the Rpn3/Rpn7 regulatory heterodimer, and tethers it to lid helical bundle; ortholog of human DSS1; protein abundance increases in response to DNA replication stress	PROTEASOME, mRNA EXPORT

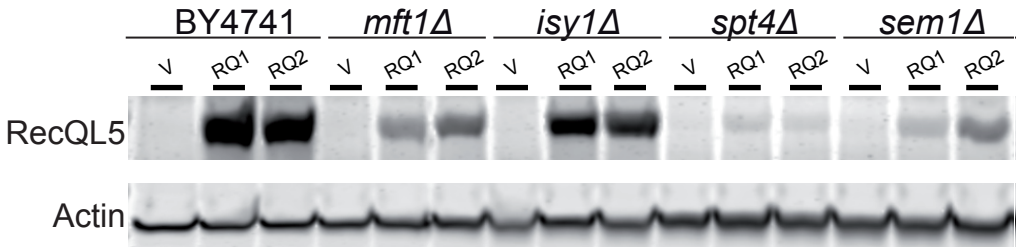


Figure R41. Suppression of phenotypes in the selected candidates is due to lower levels of RECQL5 expression.

Western blot of flag tagged RECQL5 expressed in WT (BY4741), mft1Δ (YML062C), spt4Δ (YGR063C), sem1Δ (YDR363W-A) and isy1Δ (YJR050W) strains. Exponentially growing cultures in S-raffinose were gown four additional hours after addition of galactose to final a concentration of 2%. Immunodetection of actin is shown as loading control.

at **Figure R41**, it is appreciable that, in fact, the expression of RECQL5 was clearly affected in the four mutants, being almost absent in *spt4*Δ.

This dramatic impact in the expression of RECQL5 was also supported by the fact that the human gene is 2957bp long and its sequence contains 59% of GC, making it quite difficult for the yeast transcriptional machinery to transcribe, even more if any of the important genes is mutated as in our selected strains. Therefore, we could only conclude from the screening that expression level of RECQL5 is key in order to generate noticeable phenotypes in yeast.

4.8. Genetic interaction of RECQL5 with yeast helicase mutants

SGS1 is the only member of the RecQ family of helicases present in *S. cerevisiae*, and its function is homologous to that of the human helicase BLM (Ashton & Hickson, 2010). It was also interesting that Sgs1 was a multicopy suppressor of *srs2*Δ mutant (Mankouri *et al.*, 2002). Next, we tested the overexpressing RECQL5 in the mutant for the only yeast recq helicase, and because of the tight relationship with SRS2 we analysed both mutants. For this, we transformed the strains wild-type (BY4741), *sgs1*Δ and *srs2*Δ from the EUROSCARF collection of deletion mutants with the plasmid pYES2 (GALp) carrying one of the following genes: RECQL5, RECQL5 ID, SGS1, BLM, WRN or none. Transformants were grown until exponential and serial dilutions were plated on synthetic media with different carbon sources supplemented or not with HU.

Figure R42 presents the results obtained. As expected, the wild-type

RECQL5-expressing cells grew worse than the empty plasmid control in the galactose plates, and effect being exacerbated in the presence of HU. Interestingly, overexpression of *SGS1* and BLM also impacted the growth capability of the yeast, although to a lower extent. We confirmed that the *SGS1* expression plasmid was functional since it complemented the sensitivity conferred by *sgs1*Δ in the presence of HU. The effect of complementation was clear in the glucose plates with HU, where *GALp* was repressed or activated at very low levels. This observation is likely due to promoter leakage, so that low levels of expression in this condition were enough for complementation. In addition, we confirmed that *SGS1* is a multicopy suppressor of *srs2*Δ mutant (Mankouri *et al.*, 2002). Cells transformed with the *GALp*:*SGS1* plasmid grew better than those with the empty plasmid in galactose plates. Interestingly, we observed a similar effect to that described previously with the *sgs1*Δ mutant in glucose plates. Again, low transcription of RECQL5 allowed the complementation in the *srs2*Δ mutant. However, overexpression of RECQL5 in the plates with galactose was highly deleterious, as expected. This new result

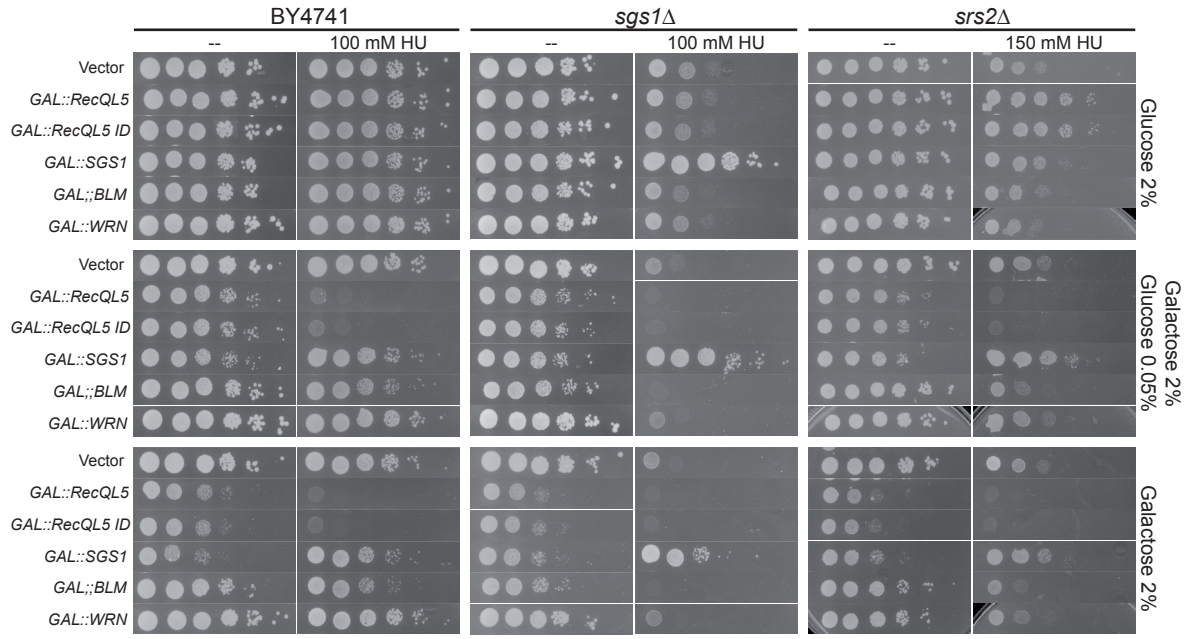


Figure R42. RECQL5 complementation of yeast helicase mutants.

RECQL5 overexpression in sgs1Δ (YMR190C) and srs2Δ (YJL092C). Ten-fold serial dilutions of WT (BY4741) and mutant strains transformed with plasmids carrying indicated human and yeast recq helicases, under the control of a GAL promoter, were plated on minimal selective medium with either glucose, galactose or a combination of both in the presence or not of hydroxyurea (HU). Photographs were taken after 3 days of growth at 30ºC

broadened our view of the possible actions of RECQL5 in yeast.

Next, taking into account that overexpression of *SRS2* impairs cell growth (León Ortiz *et al.*, 2011), we wondered whether although *SRS2* is not a member of the RecQ family of helicases, it acted as an orthologue of the human RECQL5. Since high levels of expression from both these proteins seemed to affect the viability of the cells, we decided to test if they could affect the cell cycle progression in a similar way. For this, we transformed the wild-type strain W303.1A with the *GALp* expression vectors carrying genes RECQL5, *SRS2* or none. We monitored the S-phase progression of the transformed yeast after overexpressing the indicated genes by FACS analysis **(Figure R43)**. Cells were grown to exponential phase in synthetic media with raffinose, this sugar does not repress *GALp* as glucose does, and can be maintained in the media after induction with galactose to be used as an alternative source of carbon. After a first round of α-factor synchronization, galactose was added to the media to a final concentration of 2% and cells were incubated for

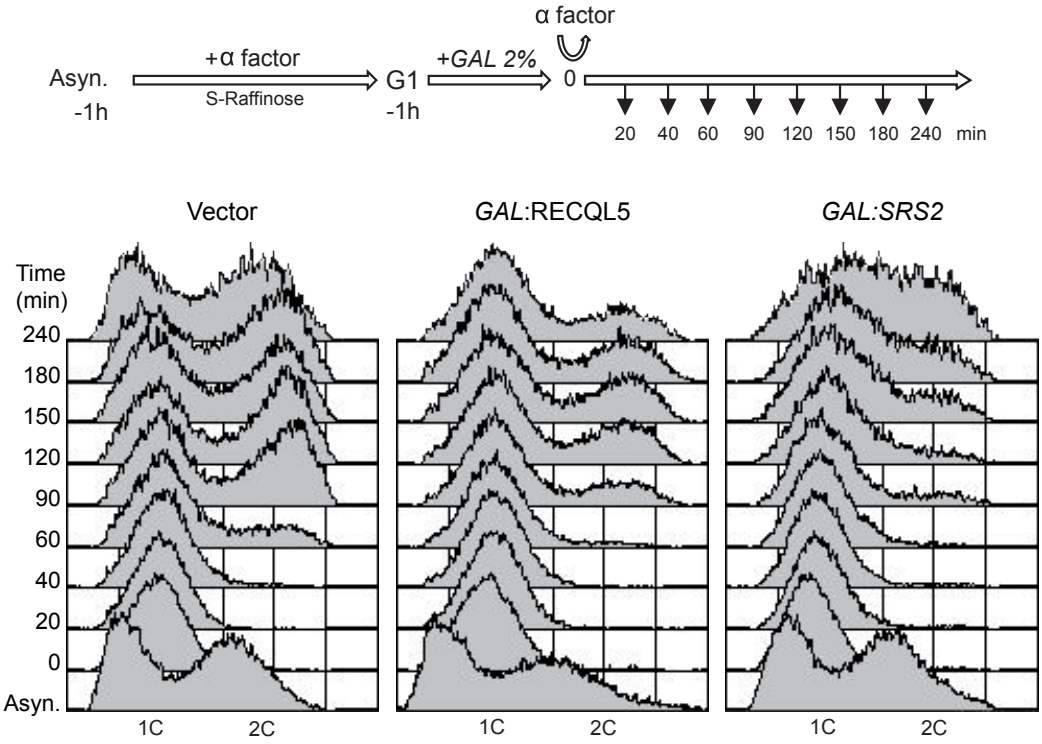


Figure R43. Helicases overexpression impact on cell cycle.

S-phase progression of wild-type strain (W303.1A) transformed with respective pYES2 vectors. Cells were synchronized in G1 with α-factor and released by changing the culture medium. A diagram of the experiment is shown at the top.

an additional hour. After 2 hours of synchronization, cells were released from G1 by shifting to fresh media with galactose and raffinose and samples were taken every 20 minutes. FACS analysis of the samples revealed that both RECQL5 and *SRS2* expressing cells presented a clear delay in entering S-phase, more acute in the *SRS2* expressing ones. This indicates the presence of problems in the cell cycle progression, probably derived from impairment of replication.

Further analyses are needed to understand the replication failure of cells expressing RECQL5 or SRS2. Our results support that there may exist shared functions between these proteins, and working with both RECQL5 and SRS2 as models could help to understand each other's functions.

Discusion

This thesis is focused on the study of transcription-replication collisions, natural events that occur inside the nucleus. This type of events is one of the main sources of replicative stress and increased genomic instability, which in turn are hallmarks of cancer and some human degenerative diseases (Gaillard *et al.*, 2015). Specifically, we study R-loops and their role in the generation of genomic instability associated with transcriptional and replicative stress.

We have studied different aspects such as where they are formed, how are they formed and which are the factors implicated in the regulation of their presence.

1. Genome wide study of RNA:DNA hybrids

We studied the genome-wide distribution of RNA:DNA hybrids in budding yeast using DRIP, with the S9.6 antibody, followed by sequencing of immunopurified DNA fragments (DRIP-seq). Our results are related to data from similar previously published studies, which we use as reference for our analysis.

We performed DRIP-seq on wild-type cells and a *hprl*Δ mutant strain, and started analyzing our sequencing results by comparing them side by side with those from previous similar publications (Chan *et al.*, 2014; El Hage *et al.*, 2014; Wahba *et al.*, 2016). We found that, although in our hands the protocol is still open to major optimization, the already available data about R-loop localization must be interpreted with discretion. The three works used as reference were performed with different methodologies, and the general conclusions dragged from these studies are essentially the same. All three studies propose that rDNA, transposons and telomeres constitute rich RNA:DNA hybrid-prone loci along the budding yeast's genome. Additionally, they mostly agree about the ORFs prone of RNA:DNA hybrid formation to be highly transcribed and GC rich. However, Chan *et al.* (2014) suggested that around one third of the yeast genome is prone to hybrid formation, while the other two studies limit their incidence to a much more reduced number of regions, covering only 8% of the genome according to Wahba *et al.* (2016). This brings to light that the differences in the methodology may influence the general outcome of the analysis and limit the possibility of extracting specific conclusions

and mechanistic hints that could potentially help understand R-loop physiological functions.

We were able to detect R-loop enrichment in a wild-type strain at rDNA locus and telomeres, which are regions that had been identified initially by qPCR-based studies (El Hage *et al.*, 2010; Balk *et al.*, 2013), as well as in transposons and a subset of RNAPII-transcribed genes **(Table R1)**. The analysis resulted in a low number of positive regions compared to previous works, but showed a high correlation with the values from the most recent study (Wahba *et al.*, 2016), which was more astringent in their analysis thanks to revisiting the methodology from the precedent studies. We put special emphasis in the analysis of the open reading frames identified as hybrid-prone regions, since the actual literature widely supports the formation of R-loops as co-transcriptional by-products (further reviewed in Aguilera & García-Muse, 2012; Santos-Pereira & Aguilera, 2015). We took all the positive ORFs identified in each of the analysis and generated a list with 235 genes that showed DNA:RNA hybrid enrichment in all 4 datasets. These genes had in common to be long, GC-rich and highly expressed **(Figure R3)**, which is consistent with previously reported features for genes prone of R-loop formation.

1.1. Accumulation RNA:DNA hybrids towards the 3' end of the gene

We also analized the signal distribution along all RNA:DNA hybrid-positive genes in our wild-type strain, compared to the data available for the wild-type used by El Hage *et al.*, (2014). The profiles show a signal distribution that spreads along the length of the transcription unit of genes, with a slight enrichment towards the 3' end **(Figure R4)**. This is consistent with the generally accepted co-transcriptional nature of the R-loops, for which it has been speculated a function in the deceleration of the RNAPII to facilitate the co-transtricptional splicing process and the correct termination and poly-adenilation of the nascent mRNA (Wahba *et al.*, 2016; Proudfoot, 2016). However, there are other interpretations that can not be ruled out yet with the data available. These include the accumulation of negative supercoiling behind the transcribing RNA polymerase, which could potentially favor the formation of an R-loop due to a higher probability of ssDNA able to

associate with the nascent mRNA as it has been observed in yeast *top1* mutants (El Hage *et al.*, 2010), or that failures in the formation of a functional mRNP could leave the RNA molecule unprotected increasing its chances to hybridize with the template DNA as observed in THO mutants (Gómez-González *et al.*, 2011; Aguilera & García-Muse, 2012).

1.2. RNA:DNA hybrids in *hprl*Δ

We performed DRIP-seq on the *hprl*Δ mutant. Cells lacking Hprl are defective in the formation of the THO complex, and consequently in the biogenesis of mRNPs (Aguilera, 2005), which leads to genomic instability associated with an elevated presence of R-loops (Huertas & Aguilera, 2003; Aguilera & García-Muse, 2012). Contrary to what could be initially expected, our results show little differences in the overall number of R-loop-prone regions compared to the wild-type **(Table R2)**. This suggests the possibility that a higher or lower accumulation of R-loops is not critical in the origin of genomic instability, but likely the R-loop nature. Indeed other determinants could be their abnormal stabilization or other processes associated with R-loops, as for example chromatin changes as suggested in the study by Castellano-Pozo *et al.*, (2013). According to our results, R-loop-accumulating ORFs are almost equally abundant in the wild-type as in the *hprl*Δ mutant (1276 in wild-type vs 1286 in *hprl*Δ strain). Interestingly, only one third of them are common to both strains. More biological replicates are needed to stablish conclusions, the results could imply that RNAPII-transcribed genes are not equally susceptible to form R-loops in the absence of the THO complex. Previous genome-wide studies (Gómez-González *et al.*, 2011) showed that Hprl is preferentially recruited to RNAPII-transcribed genes that are long, C+G rich and highly expressed, and that Hprl binding follows a gradient that increases towards the 3' end of ORFs. Similar features were found for the recruitment in different THO mutants of Rrm3, which is a DNA helicase required for the progression of the replication fork through obstacles in the DNA (Ivessa *et al.*, 2000; Azvolinsky *et al.*, 2009), suggesting that the absence of Hprl leads to such obstacles and, consequently, replicative stress (Gómez-González *et al.*, 2011).

Although *hpr1*Δ presents a lower RNA:DNA hybrid signal than the WT, similar results were previously shown by Chan *et al.* (2014). A posible explanation could be that DNA:RNA hybrids are more persistent in the *hpr1*Δ mutant. One critial feature could be R-loop length, so that this would make the hybrid more estable, thus increasing their half-life which in turn would facilitate the collapsing of incoming replication forks. This hypothesis would be consistent with the trasncription defects and the increase of Rrm3 clusters genome-wide associated to the absence of HPR1 (Gómez-González *et al.*, 2011). Nevertheless, other unexplored possibilities may explain this phenomenon, such as the differences produced in chromatin structure.

It will be interesting to compare results from studies on RRM3, histone marks and other chromatin associated factors with more definitive results obtained from DRIP-seq mapping of R-loops in the *hpr1*Δ strain, in order to stablish possible correlations with R-loop-prone regions over the genome. Such studies should open new possibilities to discover correlations woth the presence of different proteins at different sites over the genome.

2. Prevalence of R-loops in *trans*

Our analyses support the idea that RNA:DNA hybrid presence is more common than previously foreseen before the application of genome-wide methodologies to the field. Most R-loop-prone sites are associated with regions in the genome were transcription is active (i.e. ORFs, tRNAs, rDNA and telomeres) (Chan *et al.*, 2014; El Hage *et al.*, 2014; Wahba *et al.*, 2016; Ginno *et al.*, 2012 and this study), supporting the hypothesis that most R-loops if not all are formed co-transcriptionally. However, recent studies have raised the possibility that R-loops can generate by other mechanisms non associated with transcription in cis (Wahba *et al.*, 2013; Keskin *et al.*, 2014). R-loops would form in *trans* when a RNA molecule anneals with a homologous DNA strand different from the locus where it was originally synthesized. We sought to assay whether the formation of these RNA:DNA hybrids form abundantly in the cell and if so, whether they have a measurable impact on genome instability. Henceforth our study stablishes a system that could

help us understand possible mechanisms underlying the generation of R-loops.

To assay the formation of R-loops in *trans* we stablished an indirect system relying on the ability of R-loops to increase instability. The assay was based on the co-transformation of yeast cells with two plasmids. One plasmid carried the recombination system GL-LacZ which contains the whole LacZ ORF between the DNA repeats (Piruat *et al.*, 1997) and the other one would carry different sequences of the LacZ gene expressed from a tet promoter, depending on the experiment performed. We used this approach in a wild-type strain and in R-loop accumulating mutants, including two mutants of the THO complex (*mft1*Δ and *hpr1*Δ) and the double mutant *rnh1*Δ *rnh201*Δ. The THO complex, conserved through eukaryotes, is responsible of proper mRNP formation and mRNA processing and export (Luna *et al.*, 2012). Absence of different components of this complex leads to variable degrees of R-loop accumulation, defects in transcription elongation and increased levels of genomic instability (Huertas & Aguilera, 2003; García-Rubio *et al.*, 2008), mostly associated with the nascent mRNA being unprotected and failures in its post-transcriptional processing. On the other hand, Rnh1 and Rnh201 are responsible for the degradation of most RNA:DNA hybrids that are generated in the cell (Cerritelli & Crouch, 2009). Therefore, their inactivation leads to a widespread signal of S9.6 antibody detecting R-loops throughout the nucleus as seen by cytological studies (Wahba *et al.*, 2011; Chan *et al.*, 2014). However, this effect does not impair growth and only reflects in minor genomic instability phenotypes (Huertas & Aguilera, 2003; Wahba *et al.*, 2011).

In general terms, our results did not show a measurable impact of the mRNA in *trans* on the genome stability of the wild-type strain in any of the conditions assayed (**Figure R10** to **Figure R14**), suggesting that R-loops do not form easy in *trans* in the wild-type strain. We tested the capacity of different transcripts in *trans* to induce recombination in the LacZ gene (full length LacZ, inverted LacZ and 400bp fragment from the LacZ 3' end). None of the mutants showed a consistent significant increase in recombination when the only source of LacZ mRNA was in *trans* (cis transcription OFF, *trans* transcription ON), further supporting the hypothesis that R-loops in *trans* are not frequent even under conditions that could

presumably favor them. Only the *mft1*Δ showed a significant increase in one of the experiments when using the full length LacZ as template **(Figure R11)**, but this effect did not repeat in other experiments.

Additionally, when transcription was active in both plasmids (cis transcription ON, *trans* transcription ON), the only effect observed was a partial but significant suppression of the hyper-recombination in the THO mutants. Neither the wild-type nor the double *rnh1*Δ *rnh201*Δ mutant showed any significant change in their recombination levels due to the presence of an additional homologous transcript in trans. This effect was independent of the features of the *trans* RNA, either the length or the sense of the sequence.

Since our cis and *trans* mRNAs were in two different plasmids, we were limited by the marker availability in order to add other variables to the analysis, like the expression of RNH or AID. For this, we generated a plasmid-chromosome system, in which we integrated the recombination system (in cis mRNA) into the yeast genome, providing the option to analyze the effect of introducing plasmids expressing either RNH1 or AID, to determine its effect in the genome instability caused by putatively formed RNA:DNA hybrids in trans. This new system corroborated previous results, in which the presence of a mRNA in *trans* does not have an impact on recombination frequency when transcription in the recombination system was OFF. More variables and mutants must be assessed with this system in order to know why our results do not fit with those of Wahba *et al.* (2013) about the formation of R-loops in trans. However, our results do not favor the idea that R-loops form in trans. The fact that when the RNA generated in *trans* is antisense with respect to the one generated in the recombination system suggests that it hybridizes with the nascent RNA, in particular in THO mutants where they have a suboptimal mRNP structure. As a consequence, the antisense RNA is impeding the cis-ARN to hybridize back with the DNA template.

The possibility of R-loop formation starting with a molecule of RNA that anneals to a homologous DNA sequence is genuine, always if the conditions required are met. In an in vivo environment, where many factors work in the protection of the double helix structure and the regulation of RNAs, it is difficult to envision the

formation of RNA:DNA hybrids in *trans* without the intervention of proteins that positively work in the generation of a RNA:DNA hybrid. Indeed the CRISPR-Cas9 system is a good example of specific RNA:DNA hybrids that may form in trans, but mediated by proutein factors. The absence of any impact on recombination of RNA produced in *trans* in the wild-type strains supports our conclusion that there are low chances of finding a general factor implicated in the active formation of R-loops. This opposes to the work of Wahba *et al.* (2011) that claimed Rad51 to have an active role in the formation of RNA:DNA hybrids formed in trans. By working with different transcription mutants (*leo1*Δ, *med12*Δ, *sin3*Δ, *kem1*Δ and *rrp6*Δ) prone to the accumulation of RNA:DNA hybrids, they found that the deletion of Rad51 led to a complete suppression of R-loop formation and, in consequence, reduced the genomic instability phenotypes associated, measured as the loss rate of an artificial yeast chromosome (YAC). It is intriguing why they didn't include in their work any of the classical mutants prone to R-loop formation such as *hpr1*Δ (Huertas & Aguilera, 2003), *sen1-1* (Mischo *et al.*, 2011) or the double mutant *rnh1*Δ *rnh201*Δ (Cerritelli & Crouch, 2009). The function of Rad51 cannot be as critical as they show in their IF analyses, because that would imply that the Rad51 recombinase would account for almost all R-loops formed in the cell. It would be interesting to corroborate these results with other mutants prone of the R-loop formation, but so far the IF results are certainly unclear.

We sought to test our hypothesis with our own tools, using the *hpr1*Δ mutant, and see if the absence of Rad51 would also lead to the suppression of phenotypes associated with an increased presence of RNA:DNA hybrids. We studied the impact on the recombination frequency in our two-plasmid system, with the GL-LacZ (Piruat *et al.*, 1997) to measure the recombination events, which occur by the SSA (single strand annealing) repair pathway, and are independent of Rad51. In contrast to the YAC instability assay used by Wahba *et al.* (2011), which is based on a functional Rad51-dependent homologous recombination process to monitor the occurrence of GCR, that therefore study instability at loci different to where the R-loop was formed (Ray *et al.*, 1989; Pardo *et al.*, 2012). With our assay we did not observe any impact caused by the presence of an mRNA in *trans* when the

transcription of the system (in cis) was off, neither a decrease in the recombination frequency in the double mutant *hpr1*Δ *rad51*Δ compared to the *hpr1*Δ alone **(Figure R15)**. We corroborated that R-loop formation was not dependent on Rad51 by analyzing the levels of Rad52-foci in *hpr1*Δ. Repair foci accumulation in *hpr1*Δ is suppressed by RNH over-expression. However, this effect was not achieved when we combined RNH overexpreesion with the *rad51* deletion **(Figure R16)**, separating the DNA damage generated by the absence of Rad51 from the increased presence of RNA:DNA hybrids.

3. New helicases putatively involved in R-loop metabolism.

In addition to RNase H-mediated hybrid degradation, multiple helicases have been implicated in RNA-DNA hybrid removal via displacement of the RNA strand, including yeast Pif1 (Boule and Zakian, 2007) and Sen1 (human SETX) (Mischo *et al.*, 2011; Sollier *et al.*, 2014; Skourti-Stathaki *et al.*, 2011), AQR (Sollier *et al.*, 2014), and recently during the work on this project, DDX19 and DDX23 (Sridhara *et al.*, 2017). The THO complex is a transcription-coupled RNA processing complex that limits the formation of RNA-DNA hybrids in a co-transcriptional manner (Huertas and Aguilera, 2003) by allocating to the nascent RNA and ensuring that it gets efficiently exported. When R-loop removal factors are inactivated, many DNA repair genes (namely those involved in homologous recombination) become essential, underlining the propensity of R-loops to promote DNA damage as a result of replication stress (Aguilera and Garcia-Muse, 2012; Santos-Pereira and Aguilera, 2015). Mutations in SETX have been linked to juvenile amyotrophic lateral sclerosis (ALS) (Chen *et al.*, 2004) and ataxia with oculomotor apraxia type 2 (AOA2) (Moreira *et al.*, 2004), indicating the importance to identify all player in the maintenance of R-loop metabolism.

In our pursuit of understanding how R-loop homeostasis is controlled in eukaryotic cells, we set out to identify novel helicase factors that could be involved in RNA:DNA hybrid metabolism through a small-scale genetic screen looking for mutants with accumulation of DNA damage (Rad52 foci) that was sensitive to RNH1

overexpression. Four mutants were selected out of the 25 candidates selected for the screening. We finally choose just one of them to get deeper insight into its possible role as a novel factor involved in RNA:DNA hybrid metabolism.

Mutants that were positive for Rad52-foci accumulation that were suppressed by RNH overexpression were in the genes encoding the following proteins:

- **Prp28**, a RNA binding protein with no putative helicase activity (Strauss & Guthrie, 1994). Its absence renders cells inviable pointing to a role in the processing of pre-mRNA that is essential. We did not proceed with its characterization, but the fact that its human homolog, DDX23, has been linked with R-loop suppression (Sridhara *et al.*, 2017) and was found in a Co-IP with BRCA2 (V. Bhatia unpublished results, (Bhatia *et al.*, 2014)), place this protein as an interesting candidate to look forward in future projects.

- **Mrc1**, which is involved in an S-phase checkpoint required for DNA replication (Alcasabas *et al.*, 2001; Katou *et al.*, 2003). Is not essential, but its function is key for cells to survive replicative stress (Duch *et al.*, 2013). We ended setting this mutant aside because the RNH suppression was not consistent between experiments, leading us to think that if there was any possible role for Mrc1 in RNA:DNA hybrid metabolism this would be indirect. Mrc1 is likely involved in R-loop disappearance due to its known functions in the response to replication arrest, which is a notorious consequence of uncontrolled R-loop formation (Aguilera & García-Muse, 2012).

- **Dbp1**. Little is known about this candidate. This helicase has an ortholog, DED1, due to the whole genome duplication in *S. cerevisiae* (Byrne & Wolfe, 2005). Their human homolog, the human DEAD-box helicase 3 (DDX3) was found to interact with BRCA2 and co-IP with the HBD construct (V. Bhatia personal communication and (Bhatia *et al.*, 2014)). Dbp1 is a putative ATP-dependent RNA helicase, required for translation of mRNAs (Berthelot *et al.*, 2004). Despite it is known that its protein abundance increases in response to DNA replication stress (Tkach *et al.*, 2012), and that we obtained positive results for S9.6 immunostaining and consistent suppression of RAD52-foci with RNH overexpression, we discontinued its characterization. This was due to its role in translation outside the nucleus,

which represented an important drawback to explain how this helicase could be participating in the nuclear metabolism of R-loops, leaving further characterization of Dbp1 for a following project.

- **Mph1**. Finally, Mph1 is a DNA helicase, not an RNA helicase, that belongs to the DEAH/DExH family of RNA helicases (Scheller *et al.*, 2000). It is involved in translesion and RF restart. Although budding yeast cells lack a canonical Fanconi Anemia pathway, Mph1 is accepted as the homolog of the FANCM protein (Whitby, 2010). It has been recently observed that the Fanconi Anemia repair pathway, which is involved in the repair of DNA inter-strand crosslinks (ICLs) and obstacles that impede replication fork progression, participates in the dissolution of R loops (Garcia-Rubio *et al.*, 2015; Schwab *et al.*, 2015) suggesting a relevant role of DNA repair in R loop prevention and removal. Human FANCM has the ability to branch migrate replication structures, to resolve RNA-DNA hybrids in vitro and to prevent RNA:DNA hybrid accumulation in vivo in human and chicken cells (Schwab *et al.*, 2015). Additionally, the loss of Mph1 has been reported to cause a synthetic growth defect in the absence of RNH203 (Allen-Soltero *et al.*, 2014). Both the deletion and overexpression of yeast Mph1 are associated with increased replication stress and genome instability (Banerjee *et al.*, 2008; Schurer *et al.*, 2004), indicating that its activity is important when facing replication stress, but must be tightly regulated in order to prevent toxicity.

All this prompted us to advance in the characterization of the yeast Mph1 role in the control of R-loop presence. Thus, we showed that absence of Mph1 led to an increase in genomic instability that could be suppressed by overexpressing RNH **(Figure R19)**. Another way to associate Rad52-foci accumulation to the presence of R-loops is by expressing the human cytidine deaminase AID, which exacerbates the instability phenotype by introducing mutations in the ssDNA strand displaced in by the RNA:DNA hybrid (Gómez-González & Aguilera, 2007). However, we did not observe the predicted effect **(Figure R20)**, suggesting the possibility that the R-loops formed in an *mph1*Δ background are in some way different from those generated in other R-loop prone mutants such as *hpr1*Δ or *sen1-1*. This could be due to the possibility that the single stranded DNA is not unprotected.

Recently, it has been shown that FANCM, the Mph1 homolog in higher eukaryotes, is able to branch migrate and thereby unwind R-loop structures in vitro (Schwab *et al.*, 2015). Moreover, chicken DT40 cells lacking a functional FANCM translocase domain accumulate RNA-DNA hybrids (Schwab *et al.*, 2015). We demonstrate the accumulation of RNA:DNA hybrids in *mph1*Δ cells by S9.6 IF **(Figure R21)** and by DRIP analysis at transcribed genes followed by qPCR **(Figure R26)**, and found that Mph1 forms nuclear foci when RNA-DNA hybrids accumulate, using the *hpr1*Δ mutant as model **(Figure R25)**. We show that Mph1, and in particular its helicase function, suppresses the accumulation of RNA:DNA hybrids **(Figure R26)**. It is interesting, that although the spontaneous Rad52-foci accumulation is indicative of more DNA damage in Mph1 lacking cells, this is not associated to an increase in recombination **(Figure R22)** not even with the transcriptional state as it has been usually characterized in classical R-loop-prone mutants (Huertas & Aguilera, 2003; Mischo *et al.*, 2011; Santos-Pereira *et al.*, 2013). Although we cannot exclude that Mph1 may directly unwind R-loop structures in vivo, we favor a role for Mph1 at the replication fork. A possibility is that Mph1 role is not to avoid R-loop formation, but to repair by recombination of replication forks that have encountered them. This could explain the increase in Rad52-foci, which can be suppressed by RNH, but do not lead to productive recombinants. The partial suppression of the recombination frequency in the *hpr1*Δ mutant supports this hypothesis **(Figure R24)**, and if Mph1 foci were to be found only in S/G2 phase cells, this would support that Mph1 is acting at replication forks and not necessarily at RNA:DNA hybrids, per se. However, the fact that there is only a partial suppression, that *mph1*Δ does not generate any visible impact on replication fork progression **(Figure R23)** and that there is no apparent genetic interaction between the absence of Mph1 and mutants such as *hpr1*Δ or *sen1-1*, suggests that Mph1 works together with other factors in R-loop associated DNA damage.

Taking the in vitro enzymatic activities into account, we speculate that Mph1 could promote fork reversal and thereby contribute to fork restart when replication forks stall at RNA:DNA hybrids. Restart of replication forks is especially crucial when DNA replication proceeds unidirectionally, i.e. at the telomere or in the

rDNA, because converging replisomes do not exist to complete replication. In this regard, it is interesting that the SMC5/6 complex, the negative regulator of Mph1, has been proposed to play a particularly important role at sites of unidirectional replication (Murray and Carr, 2008). Additionally, Mph1 has been shown to interact with RPA and Mte1. RPA specifically binds to ssDNA that has been unwound at the fork and Mte1 binds to branched DNA structures and recruits Mph1 in S/G2 phase (Silva *et al.*, 2016). Thus, we hypothesize that if Mph1 were to unwind RNA-DNA hybrids in a more general manner, we would expect Mph1 to also be essential in *sen1-1* and THO-TREX mutants, but we did not see such genetic interactions.

Together, here we identified Mph1 as a factor preventing DNA damage at RNA:DNA hybrids. We analyzed the effects of Mph1 absence in RNA:DNA hybrid turnover in more detail and found that Mph1 not only forms foci when RNA-DNA hybrids accumulate but it may also be important in the prevention of the replication fork stalling due to the presence of R-loops.

4. Understanding the role of RECQL5 at the interface between transcription and replication

RECQL5 is one of the five members present in humans from the RecQ family of helicases. These proteins are evolutionarily well conserved and have critical roles in the maintenance of the genome integrity (Chu & Hickson, 2009; Croteau *et al.*, 2014). RECQL5 is the only one to show direct interaction with RNAPII (Aygün *et al.*, 2008; Islam *et al.*, 2010) and it was proposed to have a regulatory function at the interface between transcription and replication trough the inhibition of transcription (Saponaro *et al.*, 2014). To date, RECQL5 role in transcription hasn't been linked to R-loop regulation in a straightforward way, but recent studies point to the direction that its presence is required for the coordination of different transcription factors, like TOP1, whose function is alleviating negative supercoiling behind the transcription machinery, contributing to the suppression of genome instability caused by R-loops (Li *et al.*, 2015; García-Muse & Aguilera, 2016). In this

thesis we worked with *S. cerevisiae* as a model organism to further advance in the understanding of the biological role of RECQL5 as genome caretaker.

We showed for the first time that the expression of the human RECQL5 or its mutant version, RECQL5 ID, in yeast cells generated a growth defect that was equivalent for both proteins **(Figure R27)**. Both versions of the human helicase were found to interact with the *S. cerevisiae* RNAPII **(Figure R28)**, despite being described by Aygün *et al.*, (2009) that the KIX domain (mutated in RECQL5 ID) was essential for the interaction of the helicase with the human RNAPII in vitro. However, the posterior work by Islam *et al.*, (2010) demonstrated the existence of a second RNAPII-interaction domain in the sequence of RECQL5, the SIR domain. The KIX domain is present in several other RNAPII-transcription factors (Parker *et al.*, 1996), and the SRI domain takes its name from the Set2 Rpb1 interaction domain, found in the histone methyltransferase SetD2 which also regulates transcription (Kizer *et al.*, 2005). The RNA polymerase II undergoes different phosphorylation states at its CTD (C-terminal domain) tail during transcription. Initiation state is associated with the absence of phosphorylation (RNAPIIa) while the transcription machinery is being fully assembled. Subsequent phosphorylations of the CTD mean the beginning of the elongation phase of the transcription, in which the RNAPII stays hyper-phosphorylated (RNAPIIo) until reaching the transcription termination step. RECQL5 associates to both RNAPIIa and RNAPIIo thanks to the KIX domain, while the SRI domain is only able to interact with the phosphorylated form of the polymerase (RNAPIIo). Islam *et al.* (2010) demonstrated that these associations are maintained between the human RECQL5 and the chicken RNAPII through their experiments with DT40 cells. Since RNAPII is known to be one of the most highly conserved proteins in all eukaryotes, we predicted that the human helicase was affectively interacting with the *S. cerevisiae* polymerase.

We found that expressing RECQL5 in yeast generated high levels of genomic instability, measured by increases in recombination frequency **(Figure R31)**, in the accumulation of Rad52 foci **(Figure R30)** and sensitivity to genotoxic agents **(Figure R29)**. These phenotypes were dependent on dosage as a lower impact was produced when RECQL5 was expressed from a centromeric plasmid and placed

under control of a tetp promoter. This was confirmed in the search for yeast mutants able to suppress the sensitivity to hidroxiurea when expressing RECQL5 that we performed, in which all initial positive candidates demonstrated to be impairing RECQL5 expression **(Figure R41)**.

The hyperrecombination phenotype is characteristic of yeast mutants for RNA metabolism factors like *hpr1*Δ (Prado *et al.*, 1997), *yra1*Δ (Gavaldá *et al.*, 2016) or *npl3*Δ (Santos-Pereira *et al.*, 2013). This increase is largely derived from failures in the transcriptional process that lead to replication stress and DNA damage (Gaillard & Aguilera, 2016). In the case of RECQL5-expressing cells, we could not confirm the observed instability to be associated to the transcriptional state of the template **(Figure R31)**. Moreover, we could not detect any impact on transcription itself, as cells expressing RECQL5 were able to properly transcribe the LacZ –URA construct **(Figure R32)**, which was previously used as a tool to link transcription defects and genomic instability (Jimeno *et al.*, 2002). In spite of this, by looking at transcription at the molecular level we found that expression of RECQL5 generated a visible defect on termination, with the accumulation of mRNA read-through at the GAL10 locus **(Figure R33)**. We identified this to be the reason for the growth defect observed in cells expressing RECQL5 from a GALp, but not for the sensitivity to genotoxic agents **(Figure R34)**. It is interesting though, that the physiological role in transcription of RECQL5 in mammalian cells has been described to be as transcription inhibitor (Saponaro *et al.*, 2014), and the effect that we were observing in yeast cells closely resembles a termination bypass.

We found that RECQL5 expressed in RNAPII mutants, *rpb1*S751F and *rpb9*Δ, showed a subtle complementation with the leaky expression from the GALp in glucose. These mutants are specifically affected at transcription elongation, due to a diminished capability to avoid or repair ribonucleotide miss-incorporations into the nascent RNA (Strathern *et al.*, 2013; Koyama *et al.*, 2010). The phenotypes of these mutants have been shown to be at least partially suppressed by the overexpression of the transcription elongation factor TFIIS (DST1), which is recruited to the elongating RNAPII via the interaction with Rpb9 subunit and stimulates the backtracking and editing activity of Rpb1, which is affected in the *rpb1*S751F mutant (Koyama *et al.*,

2007; Strathern *et al.*, 2013). Recent in-vitro studies (Kassube *et al.*, 2013) proposed that RECQL5 controls transcription regulation through structural mimicry with the TFIIS factor, blocking RNAPII by competing with TFIIS. Our results would partially support this similarity, but more experiments are needed to determine the levels of RECQL5 leaky expression to explain why RECQL5 seems to positively affect transcription elongation in *S. cerevisiae*, contrary to experiments performed in-vitro or with mammalian cell cultures (Aygün *et al.*, 2009; Saponaro *et al.*, 2014).

To this point, an arising possibility is that RECQL5 could interfere with other yeast transcription factors that are recruited after elongation started, leading to failures in the proper termination process. Specific analyses are required to asses if the presence of RECQL5 was affecting the transcription rate of the yeast RNA polymerase, but to asses if the transcription defects are general and could affect mRNA production of other yeast genes we analyzed the expression profile of yeasts expressing RECQL5. Few genes were significantly affected, and most of them were just 1.5 to 2 fold compared to the control. We consider that these changes were not sufficient to explain the instability phenotypes observed. Analysis of genome-wide recruitment of RECQL5 to chromatin in yeast cells supported the interaction with the RNAPII, as many of the hits corresponded with actively transcribed genes (52% of all positive regions). Interestingly, no hits mapped to the rDNA locus, confirming the interaction specificity with the RNAPII and not with RNAPI. Detailed examination of RECQL5 distribution along the positive ORFs revealed a pattern characterized by a low occupancy at the promoter region and the beginning of the ORF, followed by an increasing accumulation toward the 3' end of the gene. This is consistent with previous analyses in human cells, where the interaction of RECQL5 is shown to be dependent on the phosphorylation of the CTD tail of an elongating RNAPII (Kanagaraj *et al.*, 2010). In physiological conditions, RECQL5 is supposed to associate to RNAPII complex 1:1 (Aygün *et al.*, 2008), such high accumulation at gene termination regions as observed in **(Figure R39)** could be explained by the accumulation of subsequently blocked RNA polymerases. However, this would dramatically impact on the viability of the cell caused by a great replication impairment and it is not the case. Accordingly, we were not able to find by FACS

analysis a cell-cycle delay that could support the idea of RECQL5 causing a defect on replication progression. It would be of great interest though, to perform Rrm3-FLAG ChIP-chip analysis like it was done before to study the global defects of THO and TREX mutants in replication progression (Gómez-González *et al.*, 2011; Santos-Pereira *et al.*, 2014). Rrm3 is a helicase required for the progression of the replication fork through obstacles in the DNA and its accumulation at specific DNA sites has been used to identify replication fork pauses or stalls (Ivessa *et al.*, 2000; Azvolinsky *et al.*, 2009). Simultaneous analyses of Rrm3 accumulation or Rpb3 binding while RECQL5 is expressed could add relevant information to understand this phenomenon.

On the other hand, RECQL5 was enriched in nine centromeres and was also present at telomeres. This could be indicative of RECQL5 interacting with other yeast proteins that lead to its recruitment to other regions of the genome different from RNAPII transcribed genes. It is known that RecQ helicases have roles in multiple DNA processes. For RECQL5, transcription is one of them due to its association with the RNAPII, but looking at other known interactors, like Rad51 (Hu *et al.*, 2007), Topoisomerase 3α and 3β (*Topo3*α and *Topo3*β) (Shimamoto *et al.*, 2000) proliferating cell agent (PCNA) (Kanagaraj *et al.*, 2006), Mre11-Rad50-Nbs1 (MRN) complex (Zheng *et al.*, 2009) it is evident that RECQL5 also participates in repair pathways as the other members of the RecQ family. Since we confirmed the interaction between the human RECQL5 and the yeast RNAPII, other yeast proteins homologous to human factors known to associate with RECQL5 could conserve this interaction in the model organism. We found that the human RECQL5 have the capability of suppressing the formation of Rad51 foci in yeast cells **(Figure R30)**. Rad51 is a protein required in initial steps in all homologous recombination reactions. HR is an important mechanism for the maintenance of genome integrity, which functions to repair double-strand breaks (DSBs) and single-strand gaps formed during replication or created by DNA damaging agents or from processing DNA lesions. In addition, HR is implicated in the restart of damaged replication forks and functions in the telomere length maintenance in cells lacking telomerase. This result suggested that RECQL5-dependent instability phenotypes could be associated to an

interference in the DNA damage repair pathway.

Nakayama *et al.* (2004) did a complementation work in *S. cerevisiae* with the RECQL5/QE from *D. melanogaster*. They concluded that the fly's helicase was able to complement some functions of the yeast unique RecQ helicase, Sgs1. However, this complementation was not observed in our yeast system. Through the years it is more accepted that the likely human homolog of Sgs1 is the RecQ helicase BLM, although it is known that BLM and RECQL5 share some functionalities in mammal cells (Hu *et al.*, 2005; Lu *et al.*, 2011). Interestingly, we did observe that human RECQL5 could partially complement *srs2*Δ yeast mutants **(Figure R42)**.

A potent and well-studied mechanism to control recombination in yeast involves the Srs2 protein (Sung & Klein, 2006). Srs2 is an helicase with similarities to the bacterial UvrD/Rep helicase (Rong & Klein, 1993), which functions to restrict recombination by its ability to bind Rad51 (Krejci *et al.*, 2003; Le Breton *et al.*, 2008). Double mutants *sgs1*Δ *srs2*Δ are extremely sick (Gangloff *et al.*, 2000), and over-expression of Sgs1 is able to partially rescue *srs2*Δ sensitivity to drugs but Srs2 does not complement *sgs1*Δ (Mankouri *et al.*, 2002). Moreover, Srs2 overexpression is known to impair cell growth (Krejci *et al.*, 2003; Mankouri *et al.*, 2002) but not by means of dysregulation of HR but because of replication fork impairment through its interaction with PCNA (León Ortiz *et al.*, 2011), which is also an identified partner of RECQL5.

More experiments need to be performed in order to understand the effect of RECQL5 expression in cells. We have demonstrated that it is able to interact with the yeast RNAPII, that it does not interfere with transcription to a degree that explains the genomic instability phenotypes observed, but it does significantly affect the formation of Rad51 filaments. In agreement, it is able to complement for Srs2 deficiency at very low concentrations, establishing a possible functional homology. An interesting approach could be to validate more known interactors, like for example PCNA, or find new ones that could uncover new clues to understand the role of RECQL5.

Conclusions

Conclusions

1. The presence of R-loops in yeast wild-type strains as determined by DRIP-seq is widespread along the genome, with special enrichment found at rDNA, telomeres, transposons, tRNAs and specific RNAPII-transcribed ORFs, consistent with other reports.
2. R-loop-prone regions mainly correlate with genomic loci that are GC-rich and highly transcribed, consistent with the idea that transcription is the main source of R-loop formation.
3. Absence of Hprl does not lead to a significant increase in the number of R-loop-prone regions. However, these regions do not entirely correlate with the ones found in the wild-type. Still, our data suggest that there are technical issues yet to be resolved to stablish definitive conclusions.
4. We did not find evidence for the formation of R-loops in trans. Anyhow, if that were the case we show that is not a phenomenon with an impact on genomic instability.
5. Deletion of Rad51 in a *hprl*Δ strain does not suppress the instability phenotype of the THO mutant, suggesting that there is no role for the recombinase in the formation of R-loops.
6. Absence of Mphl or its helicase function leads to an increase in R-loop levels inside the nucleus, suggesting a role for this protein in the maintenance of RNA:DNA hybrid homeostasis.
7. Mphl, which works with the replication fork, could potentially resolve R-loops encountered by the replication machinery, by a mechanism yet to decipher.
8. The human helicase RECQL5 can be expressed in yeast, leading to an accumulation of genomic instability that is not apparently dependent on transcription or on replication defects.
9. RECQL5 does interact with *S. cerevisiae* RNAPII and Rad51 and partially complements yeast *srs2*Δ mutants, suggesting that it may partially complement the Srs2 function, and act in yeast independent of transcription.

Conclusiones

1. La presencia de bucles R (R-loops) en fondos silvestres está distribuida por el genoma según los datos de DRIP-seq, con especial concentración en regiones específicas como el ADN ribosómico, los telomeros, transposones, secuencias de ARN transferente y algunos genes transcritos por la polimerasa de ARN II. Nuestros resultaos coinciden con otros estudios ya publicados.
2. Las regiones que presentan mayor enriquecimiento de R-loops se caracterizan por tener un alto contenido en C+G y por estar sometidas a altas tasas de transcripción, lo que apoya la idea de que, por lo general, los R-loops se generan como subproducto de la transcripción.
3. La ausencia de la proteína Hpr1 en la levadura no provoca un incremento significativo en el número de híbridos de ARN:ADN con respecto a una cepa silvestre, aunque la distribución de estos híbridos ARN:ADN no coincide en su totalidad entre ambos fondos genéticos. Sin embargo, los datos obtenidos sugieren que hace falta mejorar la técnica antes de que podamos arrojar conclusiones definitivas.
4. No hemos encontrado evidencias que indiquen que los R-loops pueden formarse en trans. En cualquier caso, si esto fuera posible, nuestros datos apuntan a que este tipo de estructuras no produce un impacto en la estabilidad del genoma.
5. La eliminación de la proteína Rad51 en células mutantes *hpr1*Δ no suprime el fenotipo de inestabilidad genómica característico de estos mutantes del complejo THO, lo que sugiere que no existe un papel activo de esta recombinasa en la formación de R-loops.
6. La ausencia de Mph1 o defectos en su función como helicasa dan lugar a la acumulación de híbridos de ARN:ADN en el núcleo, lo que sugiere un papel para esta proteína en el mantenimiento de la homeostasis de los híbridos.
7. Mph1, que actúa en conjunto con la horquilla de replicación, podría tener un papel en la resolución de R-loops que interfieran con la progresión de la replicación, mediante un mecanismo que todavía tendrá que ser descifrado.

8. La helicasa humana RECQL5 puede ser expresada en células de levadura, esto genera un fenotipo de inestabilidad genética que no parece estar asociado de forma directa a interferencias en la trascripción o la replicación.
9. RECQL5 interacciona con la polimerasa de ARN II y la recombinasa Rad51 de *S. cerevisiae*. También complementa parcialmente a los mutantes *srs2*Δ, lo que sugiere que puede cubrir parcialmente la función de Srs2 y actuar de forma independiente a la transcripción.

Materials&Methods

1. Growth media and conditions

1.1. Yeast culture media

– **Rich medium (YPAD):** 1% yeast extract, 2% bacto-peptone, 2% glucose, 20 mg/L adenine.

– **Minimum medium (SD):** 0.17% yeast nitrogen base (YNB) without ammonium sulfate nor amino acids, 0.5% ammonium sulfate, 2% glucose.

– **Complete medium (SC):** SD medium supplemented with amino acids leucine, tryptophan, histidine, lysine, methionine, aspartate and threonine and the nitrogen bases adenine and uracil. Final concentrations as described (Shermann et al., 1986). The absence of one or more of the requirements is specified when required as in SC-Ura, meaning complete medium without uracil. Alternatively, complete medium can be prepared with 2% galactose (SGal medium), 2% raffinose (SRaff medium) or 3% glycerol and 2% sodium lactate (SGL medium) instead of glucose as carbon source when required by experimental conditions.

– **Complete medium with FOA (SC+FOA):** SC medium supplemented with 500 mg/L 5-FOA, with half concentration of uracil (10 mg/L) and 0.1% L-proline instead of ammonium sulfate as nitrogen source. 5-FOA was added to autoclaved medium after cooled down to 60°C.

– **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.

1.2. Bacteria culture media

– **Rich medium (LB):** 0.5% yeast extract, 1% bacto-triptone, 1% NaCl. When necessary, it was supplemented with 100 µg/L of sodium ampicillin after autoclaving.

– **SOB medium:** 0.5% yeast extract, 2% bacto-triptone, 0.005% NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, 10 mM $MgCl_2$.

1.3. Growth conditions

Yeast strains were incubated at 30°C, except when specified. Bacteria strains were incubated at 37°C in all cases. Liquid cultures were incubated with shaking at 200rpm. Diploid yeast strains were sporulated at 26°C in SPO medium for 3-4 days.

2. Antibiotics, drugs, inhibitors, enzymes and antibodies

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. Concentration 100 µg/mL.

– **Doxycycline, Dox (SIGMA):** antibiotic from the tetracycline family that inhibits bacterial protein synthesis. Used to regulate transcription from Tet promoter in yeast strains carrying the appropriate constructs. Concentration 5 µg/ml, except when specified.

– **Geneticin, G418 (USB):** aminoglycoside antibiotic that inhibits bacterial protein synthesis. It was used in yeast strains to select, follow and maintain the kanamycin resistance gene KAN. Concentration 100 µg/ml.

– **Hygromycin B, Hyg (ROCHE):** aminoglycoside antibiotic from Streptomyces hygroscopicus that inhibits protein synthesis. It was used in yeast strains to select, follow and maintain the hygromycin resistance gene HPH. Concentration 250 µg/ml.

– **Nourseothricin, Nat (Werner BioAgents, Germany):** streptothricin-class antibiotic, extracted from Streptomyces noursei. It was used in yeast strains to select, follow and maintain the nourseothricin resistance gene NAT1. Concentration 100 µg/ml

2.2. Drugs

– **5-fluorotic acid, FOA (USB):** toxic analog of uracil that only allows growth of Ura3 mutants (Boeke et al., 1984). Concentration 500 mg/l.

– **Mycophenolic acid, MPA (SIGMA):** compound used as transcription inhib-

itor. It stops GTP nucleotides synthesis blocking inosine monophosphate dehydrogenase in the guanosin monophosphate pathway.

– **6-Azauracil, 6AU (SIGMA):** drug used to interfere in the transcription elongation process. It inhibits enzymes implicated in purine and pyrimidine biosynthesis, thus decreasing UTP and GTP intracellular pools.

– **Hydroxyurea (USB):** inactivates ribonucleoside reductase by forming a free radical nitroxide that binds a tyrosyl free radical in the active site of the enzyme. This blocks the synthesis of deoxynucleotides, which inhibits DNA synthesis.

– **Methyl methanesulfonate (Sigma):** is an alkylating agent that acts as a mutagen by altering DNA, adding methyl groups to DNA at 7-guanine preferentially, but also 3-adenine and 3-guanine.

2.3. Inhibitors

– **Phenylmethanosulfonyl fluoride, PMSF (SIGMA):** inhibitor of serine (trypsin and chymiotrypsin) and cysteine proteases. Concentration 1 mM.

– **Complete Protease Inhibitor cocktail (Roche):** it inhibits serine and cysteine proteases and metalloproteases. Used according to manufacturer's recommendations.

2.4. Enzymes

– **Spermidine (SIGMA):** polyamine involved in cell metabolism. It binds and precipitates DNA and protein-bound DNA. Concentration 0.5 mM.

– **Spermine (SIGMA):** polyamine involved in cell metabolism present in all eukaryotic cells. It binds nucleic acids and contributes to stabilize the helix structure. Concentration 0.15 mM.

– **Klenow (Roche):** major fragment of the E. coli DNA polymerase I, with 5'-3' polymerase and 3'-5' exonuclease activities.

– **Alkaline phosphatase, AP (ROCHE):** hydrolyzes 5'-monophosphate groups from DNA ends generated after an enzymatic cut. Dephosphorylation hampers the religation of the cut vector, favoring the cloning of the fragment of interest.

– **T4 phage DNA ligase (Roche):** it catalyzes the covalent union of dsDNA ends.

– **Expand® High-Fidelity DNA polymerase (Roche):** mix of Taq (from Thermus aquaticus) and Pwo (from Pyrococcus woesei). It was used for high fidelity PCRs with 5'-A overhang ends.

– **Phusion® High-Fidelity DNA polymerase (Finnzymes):** a Pyrococcus-like polymerase fused with a processivity-enhancing domain. It was used for high fidelity PCRs with blunt ends.

– **Go-Taq® Flexi DNA polymerase (Promega):** it was used for DNA probes and checking PCRs.

– **iTaq™ Universal SYBR® Green Supermix (Biorad):** 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 as passive reference, in an optimized buffer for the qPCR reaction.

– **Pronase (SIGMA):** Streptomyces griseus proteases.

– **Proteinase K (ROCHE):** very efficient serine protease from Pichia pastoris with no significant cleavage specificity.

– **Restriction enzymes (New England Biolabs and Takara):** sequence-specific DNA endonucleases.

– **RNase H (New England Biolabs):** endoribonuclease that specifically hydrolyzes the phosphodiester bonds of RNA which is hybridized to DNA. This enzyme does not digest single or double-stranded DNA.

– **RNase A (Sigma):** endonuclease that degrades single-stranded RNA.

– **Zymolyase 20T & 100T (USBiological):** mix of enzymes from Arthrobacter luteus able to digest S. cerevisiae cell wall. The two preparations differ in their lytic activity, being 20,000 units/g and 100,000 units/g respectively. Zymolyase®-20T is ammonium sulfate precipitate while Zymolyase®-100T is a further purified preparation by affinity chromatography.

– **Glusulase (PerkinElmer):** preparation of the intestinal juice of the snail Helix pomatia. It contains a mixture of enzymes including β-glucuronidase, sulfatase, and a cellulase. The activities of these enzymes have proven particularly useful for breaking open yeast cell walls

– **Lysozyme (SIGMA):** enzyme from chicken egg white that acts hydrolyzing

bacterial peptidoglycans.

– **Protein A Dynabeads (Life Technolgies):** it binds to the Fc region of IgG, IgA and IgM immunoglobulins. It was used for immunoprecipitation assays against various antibodies.

3. Antibodies

Antibodies used in this thesis are listed in **Table M1** and **Table M2**.

Table M1. Primary antibodies used in this study

Antibody	Source	Epitope	Reference	Use
Anti-FLAG M2	Mouse	N-Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys-C	F3165 (SIGMA-Aldrich)	ChIP (3 µl), WB (1:2000) BO 0.1% Tween-20 or BB, CoIP (10µl)
RNA Pol II (8WG16)	Mouse	C-terminal heptapeptide repeat present on the largest subunit of Pol II	MMS-126R (Covance)	ChIP (20 µl), WB (1:10000) BO 0.1% Tween-20
β-Actin	Mouse/ Rabbit	Synthetic peptide conjugated to KLH derived from within residues 1-100 of human β-actin	ab8224 (Abcam)	WB (1:200) BO 0.1% Tween-20
S9.6	Mouse	DNA-RNA hybrids	Hybridoma cell line HB-8730	DRIP (10 µg)

WB: Western blot; ChIP: Chromatin immunoprecipitation; IF: Immunofluorescence; BB: Blocking Reagent (ROCHE); TBS+T: 1x TBS + 0.1% Tween*20*; BO: Blocking Odissey.

Table M2. Secondary antibodies used in this study

Specificity	Conjugation	Reference	Use
Mouse	Peroxidase	A4416 (Sigma)	WB (1:6000)
Rabbit	Peroxidase	A6154 (Sigma)	WB (1:10000)
Rabbit	IRDye 800CW	925-32211 (LI-COR)	WB (If primary ab was used ≤1:1000, secondary was used 1:5000, if it was higher then seconary concentration was 1:15000)
Mouse	IRDye 680RD	925-68074 (LI-COR)	WB (If primary ab was used ≤1:1000, secondary was used 1:5000, if it was higher then seconary concentration was 1:15000)
Mouse	Cy3	#115-165-003 (Jackson Laboratories)	IF(1:1000)

3.1. Antibody preabsorption

Spheroplasts were prepared from 100 ml of exponentially growing yeast. Strain BY4741 was cultured in YPAD at 30°C. Cells were collected centrifuging 5 minutes at 3000 rpm, resuspended in 50 ml formaldehyde 3,7% and incubated overnight with gently rocking at 4 °C. Afterwards, cells were washed three times with 50 ml of KPi 0.1 M pH 6.4 buffer and resuspended in 10 ml sorbitol-citrate 1.2 M. For the cell wall digestion, 100 µl of zymoliase 100T (USBiological) and 1 ml of glusulase (PerkinElmer) were added and incubated for 1 hour at 30 °C with shaking. To stop the reaction, spheroplasts were softly centrifuged 5 min at 1000 rpm and washed 3 times with 50 ml of sorbitol-citrate 1.2M. Finally, pellet was resuspended in 5 ml of sorbitol-citrate and distributed in 1ml aliquots that can be stored at -20 °C.

For preabsorption of the antibodies, 200 µl of the antibody of interest were mixed with 200 µl of yeast spheroplasts and incubated at RT for 20 min with gently rocking. Mixture was centrifuged 5 min at 1000 rpm, and supernatant was mixed with another 200 µl of fresh spheroplasts in a new 1.5 ml tube followed by incubation in the same conditions described before. This process was repeated for a total of 5 times, taking last supernatant as final product. Because of this procedure, the antibody is already diluted 1:6 compared to the original concentration and so, this has to be taken in account when using the antibody for an experiment. Preabsorbed antibody was aliquoted and stored at -20 °C.

4. Strains

All experiments with E. coli were carried out using the DH5a strain: F- endA1 *gyr96* hsdR17 ΔlacU169(f80lacZΔM15) recA1 relA1 supE44 thi-1 (Hanahan, 1983).

Yeast strains used are shown in **Table M3** and were obtained from the referenced sources. Single, double and triple mutants were obtained by backcrosses and tetrad dissection using a SINGER MSM 200 micromanipulator or by PCR-mediated gene replacement using the short flanking homology (SFH) method (Wach *et*

al., 1994) (primers listed in Table **Table M5**.

Table M3. Yeast strains used in this study

Strain	Description	Reference
BY4741	*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	EUROSCARF
BY4742	*MATα his3Δ1 leu2Δ0 ura3Δ0 lys2Δ0*	EUROSCARF
Y10775	*BY4741 sgs1Δ::KAN*	EUROSCARF
Y01331	*BY4741 srs2Δ::KAN*	EUROSCARF
YJR050W	*BY4741 isy1Δ::KAN*	EUROSCARF
YML062C	*BY4741 mft1Δ::KAN*	EUROSCARF
YKR059W	*BY4741 tif1Δ::KAN*	EUROSCARF
YJL138C	*BY4741 tif2Δ::KAN*	EUROSCARF
YIL084C	*BY4741 sds3Δ::KAN*	EUROSCARF
YDL160C	*BY4741 dhh1Δ::KAN*	EUROSCARF
YKR024C	*BY4741 dbp7Δ::KAN*	EUROSCARF
YPL119C	*BY4741 dbp1Δ::KAN*	EUROSCARF
YGL078C	*BY4741 dbp3Δ::KAN*	EUROSCARF
YGR063C	*BY4741 spt4Δ::KAN*	EUROSCARF
YDR363W-A	*BY4741 sem1Δ::KAN*	EUROSCARF
YBR018C	*BY4741 gal7Δ::KAN*	EUROSCARF
Mini-KO collection	Mutants shown in Apendix 1	Francisco García Benitez
R1158	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0*	Winzeler et al. 1999
TH_2831	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pDED1::KAN-tetO7TATA*	Winzeler et al. 1999
TH_7495	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pDBP9::KAN-tetO7TATA*	Winzeler et al. 1999
TH_2295	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pMAK5::KAN-tetO7TATA*	Winzeler et al. 1999
TH_2297	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pFAL1::KAN-tetO7TATA*	Winzeler et al. 1999
TH_3021	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pROK1::KAN-tetO7TATA*	Winzeler et al. 1999
TH_2803	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pPRP5::KAN-tetO7TATA*	Winzeler et al. 1999
TH_2771	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pSPP382::KAN-tetO7TATA*	Winzeler et al. 1999
TH_4953	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pPRP28::KAN-tetO7TATA*	Winzeler et al. 1999
TH_5621	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pMOT1::KAN-tetO7TATA*	Winzeler et al. 1999
TH_6131	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pMCM5::KAN-tetO7TATA*	Winzeler et al. 1999
TH_7499	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pSEN1::KAN-tetO7TATA*	Winzeler et al. 1999
TH_7555	*MATa URA::CMV-tTA MATa his3Δ1 leu2Δ0 met15Δ0 pRVB2::KAN-tetO7TATA*	Winzeler et al. 1999

Table M3. Yeast strains (Continuation)

Strain	Description	Reference
AYW3-1B	*MATα ade2 can 1-100 ura3-1/52 his3-200/11,15 trp1 rad5-G35R leu2k::URA3-ADE2::leu2k*	Piruat & Aguilera, 1996
W303.1A	*MATa ade2-1 leu2-3,112 trp1-1 can1-100 ura3-1 his3-11,15 rad5-G35R*	R. Rothstein
W303.1B	*MATα ade2-1 leu2-3,112 trp1-1 can1-100 ura3-1 his3-11,15 rad5-G35R*	R. Rothstein
Ybp249	*MATa ade2-1 leu2-3,112 trp1-1 can1-100 ura3-1 his3-11,15 bar1Δ RAD5*	B. Pardo
Ybp250	*MATα ade2-1 leu2-3,112 trp1-1 can1-100 ura3-1 his3-11,15 bar1Δ RAD5*	B. Pardo
W303.1AR5	*MATa ade2-1 leu2-3,112 trp1-1 can1-100 ura3-1 his3-11,15 RAD5*	M. Moriel-Carretero & Aguilera, 2010
W303.1BR5	*MATα ade2-1 leu2-3,112 trp1-1 can1-100 ura3-1 his3-11,15 RAD5*	M. Moriel-Carretero & Aguilera, 2011
HPBAR1-R	*MATa W303.1A RAD5 bar1Δ::HYG hpr1A3::HIS3*	M. SanMartin-Alonso
HPBAR1-R2	*MATα W303.1B RAD5 bar1Δ::HYG hpr1A3::HIS3*	M. SanMartin-Alonso
SEN1-R	*MATa W303.1A RAD5 bar1Δ::HYG sen1-1*	M. SanMartin-Alonso
RNH-R	*MATa W303.1A RAD5 bar1Δ::HYG rnh1::KAN rnh201::KAN*	M. SanMartin-Alonso
WMPH1.1A	*W303 MATα mph1Δ::KAN*	This study
WMPH1.2B	*W303 MATa mph1Δ::KAN*	This study
WMPH1.8B	*W303 MATa mph1Δ::KAN*	This study
T597-1	*MATα W303 mph1-Q603D-YFP::HIS5*	Chen et al. 2009
T617	*MATα W303 mph1-E210Q-YFP::HIS5*	Chen et al. 2009
WMPHP-5A	*MATα W303 mph1Δ::KAN hrp1ΔHIS3*	This study
WMPHP-6B	*MATa W303 mph1Δ::KAN hrp1ΔHIS3*	This study
WMPSEN-1C	*MATα W303 mph1Δ::KAN sen1-1*	This study
WMPSEN-2C	*MATa W303 mph1Δ::KAN sen1-1*	This study
ML8-9A	*MATa BAR1 LYS2 ADE2 can1-100 ura3-1 his3-11,15 leu2-3, 112 trp1-1 RAD5*	Brian Luke
ML66-11A	*MATa ML8-9A MPH1-YFP*	Silva et al. 2016
YBG722	*ML66-11A hpr1Δ::KAN*	B. Gomez-González
WbMPH.L	*Mata WMPH1.2B bar1ΔHYG*	This study
WbMPH.M	*Mata WMPH1.2B bar1ΔHYG*	This study
WbDBP.2B	*Mata Ybp249 dbp1ΔNAT*	This study
WbDBP.2A	*Mata Ybp249 dbp1ΔNAT*	This study
ML149-84	*MATa ADE2 trp1-1 LYS2 YFP-RAD51 RAD5*	M. Lisby
WRP1-12A	*MATα ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5 rpb1-1*	Felipe-Abrio et al, 2014
WSR8-5A	*MATα ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5 rpb1S751F*	Felipe-Abrio et al, 2014
WRP2	*MATa ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5 rpb2-10*	Felipe-Abrio et al, 2014
WRP9-3C	*MATα ade2-1 can1-100 his3-11,15 leu2-3,112 trp1-1 ura3-1 RAD5 rpb9 ::HIS3*	Felipe-Abrio et al, 2014

Table M3. Yeast strains used in this study (Continuation)

Strain	Description	Reference
WHP.20A	*Matα W303.1BR5 hpr1DHIS3*	I. Felipe-Abrio
WGLZN.1	*W303.1AR5 (YREC56-24) GAL::LEU2 3'Δ-lacZ-leu5'Δ::NAT @ChIII*	This study
WGLZN.5	*W303.1AR5 (YREC56-24) GAL::LEU2 3'Δ-lacZ-leu5'Δ::NAT @ChIII*	This study
WGLZN.3B	*MATa W303 GL-LacZ::NAT*	This study
WGLZN.10B	*MATα W303 GL-LacZ::NAT*	This study
WHPGLZN.1A	*MATα W303 GL-LacZ::NAT hrp1ΔHIS*	This study
WHPGLZN.13D	*MATa W303 GL-LacZ::NAT hrp1ΔHIS*	This study
WRBb-9D	*MATa ade2-1 bar1Δ can1-100 his3-11,15 leu2-3,112 RRM3::FLAG::KAN trp1-1 ura3::URA3/GPD-Tk(7x)*	Felipe-Abrio et al, 2014
HRBb.12	*WRBb-9D hpr1Δ::NATnt2*	This study
HRBb.23	*WRBb-9D hpr1Δ::NATnt2*	This study
MRBb.1	*WRBb-9D mph1Δ::HYGnt1*	This study
MRBb.8	*WRBb-9D mph1Δ::HYGnt1*	This study
U678-1C	*W303.1A hpr1A3::HIS3*	R. Rothstein
U678-4C	*W303.1B hpr1A3::HIS3*	R. Rothstein
WMK.1A	*W303.1A mft1::KanMX4*	S.Chavez et al. 2002
HRN2.10C	*MATa his3 leu2 trip1-1 ura3 rnh1::KAN rnh201::KAN*	Huertas & Aguilera, 2003
WSR51.4A	*MATα ade2-1 can1-100 his3-11,15 leu2Δ::SFA1 trp1-1 ura3-1 rad51Δ::kanMX4*	Gonzalez-Barrera et al, 2002
HPR51.15A	*Mat α trp- ade- leu- his- lys+ met+ hpr1ΔHIS rad51ΔKAN rad5-*	This study
F4	*MATa thr4*	G. Fink
F15	*Matα thr2 arg4*	G. Fink

5. Plasmids

Plasmids used in this study are listed in **Table M4** and were obtained from the indicated sources.

Table M4. Plasmids used in this study.

Plasmid	Description	Source
pFA6-kanMX4	pFA6 containing the kanMX4 cassette	Wach et al., 1994
pFA6-hphNT1	pFA6 containing the hphMX4 cassette with TTEF terminator exchanged for a TCYC	Janke et al., 2004
pFA6-natNT2	pFA6 containing the natMX4 cassette with TTEF terminator exchanged for a TADH	Janke et al., 2004
pCM184	YCp containing TRP1 as marker and 7 tetO boxes as promoter. Doxycycline repressible.	Gari et al., 1997

Table M4. Plasmids used in this study (Continuation)

Plasmid	Description	Source
pCM189	YCp containing URA3 as marker and 7 tetO boxes as promoter. Doxycycline repressible.	Gari et al., 1997
pCM190	YEp containing URA3 as marker and 7 tetO boxes as promoter. Doxycycline repressible.	Gari et al., 1997
pCM179-tet::LacZ	YEp containing URA3 as marker and 7 tetO boxes that control the expression of the lacZ gene from E. coli. Doxycycline repressible.	Gari et al.,1997
pCM190-tet::LacZi	YEp containing URA3 as marker and 7 tetO boxes that control the expression of the lacZ gene from E. coli, cloned in inverted orientation with respect to the promoter.	This study
pCM190-tet::LacZ400	YEp containing URA3 as marker and 7 tetO boxes that control the expression of a KpnI-BamHI 400bp fragment of the 3' from the lacZ gene.	This study
pCM189:RNH1	YCp pCM189 containing RNH1 gene under the Ptet	Castellano-Pozo et al., 2012
pCM189:AID	YCp pCM189 containing AID gene under the Ptet at the NotI site.	D. García-Pichardo
pCM184:RNH1	YCp pCM184 containing RNH1 gene under the Ptet	J.M. Santos-Pereira
pCM184:AID	YCp pCM184 containing AID gene under the Ptet at the NotI site.	J.M. Santos-Pereira
pCM184:lacZ-URA3	YCp containing the Ptet::lacZ::URA3 fusion and TRP1 as a marker	Jimeno et al., 2002
pRS314	YCp containing TRP1 as marker	Sikorski y Hieter, 1989
pRS314-GL-lacZ	YCp p314GLB containing the GL-lacZ recombination system under the PGAL1 with the 3Kb-fragment BamHI from lacZ inserted between the two leu2 direct repeats	Piruat & Aguilera, 1997
pRS314-L	pRS314 containing two direct repeats of LEU2 gene sharing 600 bp of homology	Prado & Aguilera, 1995
pRS314-LY	pRS314-L containing the complete YIp5 sequence (5.6kb) inserted at the BglII site located between the repeats	Prado & Aguilera, 1995
pSCh204	YCp pRS314-LB containing the L-lacZ recombination system under the PLEU2 with the 3Kb-fragment BamHI from lacZ inserted between the two leu2 direct repeats	Chavez & Aguilera, 1997
pRS314-GL-lacZ	YCp p314GLB containing the GL-lacZ recombination system under the PGAL1 with the 3Kb-fragment BamHI from lacZ inserted between the two leu2 direct repeats	Piruat & Aguilera, 1997
pCM184-TL-lacZ	YCp pCM184 containing the TL-lacZ recombination system under the Ptet with the 3Kb-fragment BamHI from lacZ inserted between the two leu2 direct repeats	Santos-Pereira et al., 2013
pRS316-TINV	YCp pRS316 containing two leu2 inverted repeats: the allele leuOHr under the Ptet and the leu2 Δ5' allele	González-Barrera et al., 2002

Table M4. Plasmids used in this study (Continuation)

Plasmid	Description	Source
pRS316	YCp containing URA3 as marker	Sikorski y Hieter, 1989
pRS316-L	pRS316 containing two direct repeats of LEU2 gene sharing 600 bp of homology	Prado & Aguilera, 1995
pRS316-LY	pRS316-L containing the complete YIp5 sequence (5.6kb) inserted at the BglII site located between the repeats	Prado & Aguilera, 1995
pRS316-LYΔNS	pRS316-LY in which a 1.92-kb SphI-NsiI deletion has been made, removing the URA3 gene.	Prado et al., 1997
pRS316-L-LacZ	Fragment ApaI-SacI from pSCh204 cloned in pRS316	This study
pRS316-GL-LacZ	Fragment ApaI-SacI from pRS314-GL-LacZ cloned in pRS316	This study
pRS315	YCp containing LEU2 as marker	Sikorski y Hieter, 1989
pRS315-GAL:RNH1	YCp pRS315 containing the GALp::GALRNH1 fusion	Gómez-González & Aguilera, 2007
pWJ1213	YCp HIS containing the Rad52::YFP fusion	Lisby et al.,2001
pWJ1344	YCp LEU containing the Rad52::YFP fusion	Lisby et al., 2001
pYES3	Multicopy plasmid (2μ) with TRP1 marker and a PGAL1	SIGMA
pYES2	Multicopy plasmid (2μ) with URA3 marker and a PGAL1	SIGMA
pYES2::RECQL5 WT	pYES2 with human RECQL5 orf obtained from pCMV::FLAG-RECQL5 (Aygun et al., 2009). PCR amplified with restriction site adapters and a FLAG epitope at the 3' end. Cloned between HindIII/XhoI	J. Svejstrup
pYES2::RECQL5 ID	Same construct as before, containing the RECQL5 ID (Interaction Domain mutation (9aa deletion between 542-552)) from pCMV::FLAG-RECQL5ID (Aygun et al., 2009)	J. Svejstrup
pYES3::RECQL5 WT	Same cosntruct as pYES2::RecQL5 ID but with pYES3 backbone	J. Svejstrup
pYES3::RECQL5 ID	Same cosntruct as pYES2::RecQL5 WT but with pYES3 backbone	J. Svejstrup
pCM184::RECQL5 WT	pCM184 with HindIII-XhoI fragment from pYES2::RECQL5 WT blunted and cloned at StuI	This study
pCM184::RECQL5 ID	pCM184 with HindIII-XhoI fragment from pYES2::RECQL5 ID blunted and cloned at StuI	This study
pCM189::RECQL5 WT	Same cosntruct as pCM184::RecQL5 WT but with pCM189 backbone	This study
pCM189::RECQL5 ID	Same cosntruct as pCM184::RecQL5 ID but with pCM189 backbone	This study
pYEs2:SRS2	pYES2 carrying the SRS2 gene	Mankouri et al, 2002
pYES2:SGS1	pYES2 carrying the SGS1 gene	Mankouri et al, 2002
pYES2::GAL7	Gal7 ORF PCR amplified with adapters, cloned at PvuII XhoI of pYES2.	This study
pRS425-GPD	YEp LEU2with PGPD	Mumberg et al., 1995

Table M4. Plasmids used in this study (Continuation)

Plasmid	Description	Source
pRS425-GPD::RNH1	Fragment BamH1-PstI from pCM184:RNH1 cloned in pRS425-GPD	This study
pRS425-GPD::AID	Fragment BamH1-PstI from pCM184:AID cloned in pRS425-GPD	This study

6. Yeast methodology

6.1. Yeast transformation

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

6.2. Cell cycle synchronization and FACS analysis

For Fluorescence-Activated Cell Sorting (FACS) overnight mid-log cultures of *bar1*Δ strains were synchronized in G1 with 0.125 µg/ml α-factor (Biomedal) for 2.5 h and released from G1 in the present or not of 20 mM HU. Samples at each time-point were processed for the different experiments. For FACS analysis, 1 mL of the desired culture was centrifuged and washed with 1 ml distilled water, then resuspended in 1 ml 70% ethanol and stored at -20 °C. Cells were washed with 1 ml 1x PBS, resuspended in 100 µl 1x PBS-RNase A 1 mg/ml and left for overnight incubation at 37°C. Next they were washed again with 1x PBS and resuspended in 1 ml of 5 µg/mL Propidium Iodide in 1x PBS, incubated in darkness for 30 minutes, sonicated 4 seconds at 10% amplitude and scored in a FACScalibur (Becton Dickinson, CA).

6.3. Genotoxic damage sensitivity assay

Mid-log cultures growing in YPAD or SC medium were 10-fold diluted in sterile water. For drop assays 3 or 7 µl of the culture were plated on solid YPAD or SC medium, respectively, containing the concentration of drugs indicated in the figures. For UV irradiation drops were dry before the irradiation. Plates were incubated during 2-6 days (in darkness in the case of UV) at the indicated temperature.

6.4. pLAUR transcription assay

Cells were transformed with the pCM184-LAUR plasmid containing the pLAUR system (Jimeno *et al.*, 2002) and selected in SC-trp. Transformed cells were cultured in SC-trp and mid-log cultures were plated by 10-fold serial dilution on SC-trp plates to observe the growth of each strain, and on SC-trp-ura to analyze the capacity of the cells to transcribe the fusion construct lacZ-URA3 and, hence, to growth in SC-ura.

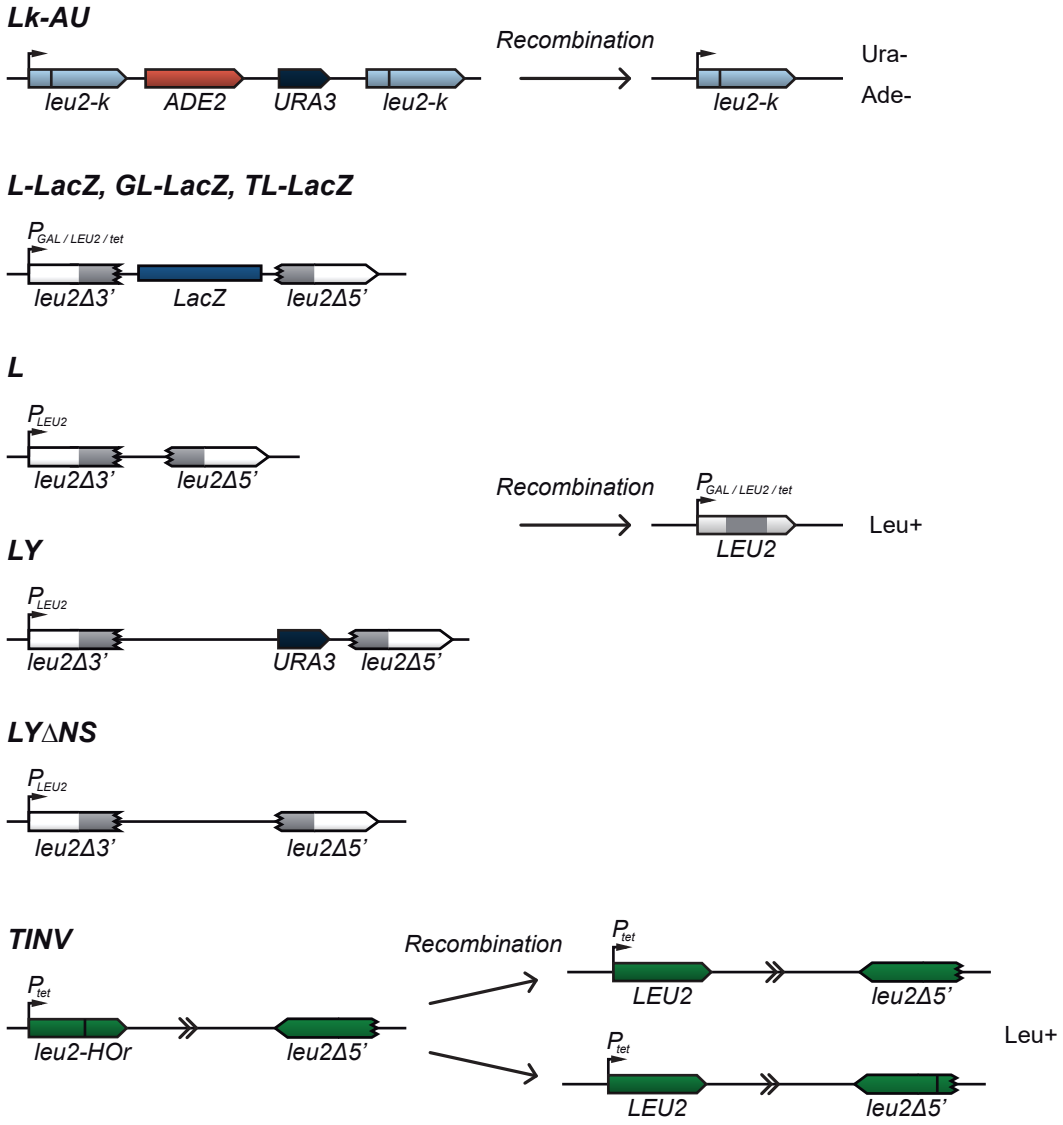


Figure M1. Recombination systems

The outcome of the recombination even and the marker used for recombinant selection are shown

6.5. Recombination assays

Recombination frequencies were calculated as means of at least 3 median frequencies obtained each from 6 independent colonies isolated in the appropriate SC medium for the selection of the required plasmids. Recombinants were obtained by platting appropriate dilutions in applicable selective medium. To calculate total number of cells, plates with the same requirements as for the original transformation were used. All plates were grown for 3-4 days at 30°C unless otherwise noted

6.5.1. Direct-Repeat systems

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

Recombination system integrated in the yeast genome at chromosome III. Consistent in two direct-repeats of the *leu2*-k allele, which lacks a 7 bp region that contains a KpnI target. Between the two repeats there is an 11 kb fragment containing ADE2 and URA3 markers. This system allows studying deletions, since the loss of URA3 confers resistance to FOA and the loss of the ADE2 gene causes accumulation of a red pigment that makes hyper-recombinants generate red-white sectored colonies. (Aguilera & Klein, 1989) **(Figure M1)**

- ***Plasmid L-lacZ, GL-lacZ and TL-lacZ systems***

Starting with an LY backbone, the sequence of the 3 kb long lacZ gene from E. coli was cloned between the direct repeats, resulting in the L-lacZ system. In the GL-lacZ the promoter driving the expression of the construction comes from the GAL1 gene, and in the case of the TL-lacZ system, the promoter sequence was replaced by a doxycycline repressible promoter. The diversity of promoters allows studying the influence of transcription in the instability: with PGAL, the use of glucose in the media renders a low transcription level, while galactose induces a high level of transcription. PLEU is constitutive and considered as medium level of transcription. The TL-lacZ is generally used when strains do not grow in galactose containing media as carbon source, and so Ptet is repressed with the addition of 5 μg/ml of DOX to obtain a low level of transcription and the high levels are achieved without the addition of DOX. In all cases, recombinants are selected in Leu- plates. (Chávez & Aguilera, 1997; Piruat & Aguilera, 1998; Santos-Pereira *et al.*, 2013) **(Figure M1)**.

- ***Plasmid L, LY and LYΔNS systems***

All three systems are based on two fragments of the LEU2 gene that share 600 bp of homology, *leu2*Δ3' and *leu2*Δ5'. The difference resides in the length of the sequence that has been placed between the repeats. L system has only 31 bp, LY system has a complete YIp5 open at BamHI inserted and LYΔNS system is the same construction as LY but in which 3.7 kb have been eliminated and the URA3 gene of YIp5 alongside. Recombinants are selected in Leu- plates. (Prado & Aguilera, 1995; Prado *et al.*, 1997) **(Figure M1)**

6.5.2. Inverted-repeat system

- ***Plasmid TINV system***

This system is comprised of two inverted *leu2* sequences. One copy is the *leu2*-HOr allele, which contains a modified target of the HO endonuclease. The other copy is the *leu2*Δ5' fragment. Both repeats are separated by 560 bp. Transcription of *leu2* is driven from Ptet. Leu+ recombinants can arise by gene conversion of the *leu2*-HOr without an associated inversion or by crossover occurring upstream of the HO site, whether or not associated with gene conversion. (González-Barrera *et al.*, 2002). **(Figure M1)**

7. Detection of Rad52-YFP foci

Spontaneous Rad52-YFP foci from mid-log growing cells carrying plasmid pWJ1344 were visualized and counted by fluorescence microscopy in a Leica DC 350F microscope, as previously described (Lisby *et al.*, 2001). More than 200 S/G2 cells where inspected for each experimental replica.

8. Chromosome spreads Immunofluorescence

Cells were grown to mid-log phase (OD660 ≈0.5) in YPAD at 30°C, washed in cold spheroplasting solution (1.2 M sorbitol, 0.1 M potassium phosphate, 0.5 M MgCl2, pH 7) and digested in spheroplasting solution with 10 mM DTT and 150 mg/

ml Zymolase 20T at 37°C for 10 minutes. The digestion was halted by addition of ice- cold stop solution (0.1 M MES, 1 M sorbitol, 1 mM EDTA, 0.5 mM MgCl2, pH 6.4). Next, spheroplasts were lysed with 1% vol/vol Lipsol and fixed on slides using 4% wt/vol paraformaldehyde 3.4% wt/vol sucrose, spreading was performed with a glass rod before letting the slides to dry overnight in the extraction hood. Chromosome spread slides were incubated with the mouse monoclonal antibody S9.6 (1 mg/ml in blocking buffer of 5% BSA, 0.2% milk and 1x PBS) for 2h at 23°C in a humid chamber. After washing with PBS for 10 minutes, the slides were further incubated with a preadsorbed (See Materials & Methods 4.11) secondary Cy3-conjugated goat anti-mouse antibody (Jackson Laboratories, #115-165-003, diluted 1:1000 in blocking buffer) during 1h at 23°C in an humid chamber kept in the dark. Slides were mounted with 50 µl of VectaShield® (Vector Laboratories, CA) with 1x DAPI and sealed with nail polish. For each replicate, at least 100 nuclei were visualized and manually counted to obtain the fraction with detectable DNA:RNA hybrids. Each mutant was assayed in triplicate similar to (Chan *et al.*, 2014).

9. DNA analysis

9.1. Southern BLOT

Yeast genomic DNA was digested, separated in an agarose gel and transferred to Hybond-N or –XL nitrocellulose membranes (Amersham), which were hybridized with 32P-labelled DNA probes, and the radioactive signal was read using a FLA-5100 Imager Fluorescence Analyzer (Fujifilm), and quantified when appropriate using the MultiGauge 2.0 analysis software (Science Lab).

9.2. Polymerase chain reaction (PCR)

- **Non-quantitative PCR**

DNA amplification with temperature-stable polymerases for probe generation, strain verification or amplification of cloning fragments were performed following standard and manufacture's protocols with the polymerases described in Materials and Methods 2.3. DNA primers used are listed in **Table M9**.

- **Real-time quantitative PCR (qPCR)**

This technique allows the measurement of the DNA quantity present in a sample during the reaction thanks to the fluorescence emitted by the SYBR® Green reactive. For this thesis, real-time qPCRs were performed using the SYBR® Green PCR Master Mix (Biorad) and a 7500 Fast Real Time PCR System (Applied Biosystems). Reactions were set with 6 µl H2O, 2 µl primer mixture (each 0.1 mM), 2 µl DNA and 10 µl SYBR. Runs were always performed in "FAST" mode with the following program: 1 cycle (10 min. 95°C), 40 cycles (15 s. 95°C and 1 min. 65°C) with a final dissociation stage (15 s. 95°C, 1 min. 65°C, 15 s. 95°C and 15 s. 60°C). Results were analysed with 7500 System Software V2.0.6. A calibration curve consisting of five 10-fold serial dilutions of a standard DNA sample was calculated to allow absolute quantification. Real-time qPCR primers were designed using Primer Express 3.0 Software (Applied Biosystems).

Primers used in this thesis, both for non-quantitative and quantitative PCR, are described in **Table M3**.

10. Chromatin immunoprecipitation (ChIP)

10.1. Chromatin extraction

Asynchronous mid-log cultures (OD660 ≈0.5) growing in indicated mediums at 30 °C were collected. Sample processing was performed as described (Hecht & Grunstein, 1999) with some modifications. 50 ml of culture was cross-linked by incubating with formaldehyde to a final concentration of 1% under soft agitation for 15 minutes at RT. Reaction was stopped adding glycine to a final concentration of 125 mM for 5 minutes. After that, samples were washed twice with cold PBS and stored at -80°C. For cell extract preparation, pellets were resuspended in 300 µl of lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA pH 8, 1% Triton X-100, 0.1% sodium deoxycholate) supplemented with protease inhibitors (1x Complete Protease Inhibitor Cocktail (Roche) and 1 mM PMSF). Next, 1 volume of glass beads were added in 2 ml tubes and cells were shredded at 4 °C in a multi-beads Shocker (MB400U, Yasui Kikai, Japan) or a IKA Vibrax VXR orbital shaker for 45

minutes. Liquid was recovered by perforation of the bottom of the tubes and samples were centrifuged at 3000 rpm for 1 minute to eliminate soluble proteins. The pellet was resuspended in 300 µl of lysis buffer supplemented with protease inhibitors and sonicated to fragments of 400 bp using a Bioruptor (Diagenode, Belgium) alternating 30'' high intensity - 30'' pause pulses for 30 minutes. Then, samples were centrifuged for 5 and 15 minutes at 13000 rpm to eliminate cell debris. 20 µl of the final volume (300 µl) were used as a control of total DNA (Input) and the rest was processed for immunoprecipitation.

10.2. Immunoprecipitation

3 µl of mouse anti-FLAG antibody M2 (SIGMA-Aldrich) were coated to (previously 1x PBS/0.5% BSA washed x 2) Dynabeads Protein A (Invitrogen) overnight at 4°C rocking at V:12. Next day, antibody excess was washed and chromatin samples obtained in the previous step were put to incubate rocking overnight at 4°C. Subsequently, samples were washed several times: 2 washed with 1 ml of lysis buffer, 1 ml of lysis buffer supplemented with 500 mM NaCl, 1 ml of buffer III (10 mM Tris-HCl pH 8, 1 mM EDTA pH 8, 250 mM LiCl, 0.5% IGEPAL, 0.5% SDS) and 1 wash with 1x TE (10 mM Tris-HCl, 1 mM EDTA pH8). Samples were dried and chromatin was eluted by incubation of the samples in 60 µl 1x TE – 1% SDS at 65°C for 10 minutes. 20-60 µl of the supernatant were taken as precipitate (IP).

Purification of the DNA in Input and IP was performed after the descrosslinking and elimination of proteins in the samples. 20 µl of 1x TE – 1% SDS were added to each sample (20 µl) and samples were treated with 1.4 µl of 50 mg/ml pronase for 2 hours at 42°C and 6 hours at 65°C. The WizardTM SV DNA clean-up system (Promega) was used for the last DNA purification step. DNA was eluted in 150-200 µl of TE.

11. ChIP hybridized to Tilling Arrays (ChIP-Chip)

GeneChip *S. cerevisiae* Tiling 1.0R Arrays were provided by Affymetrix. The high-density oligonucleotide arrays used are able to analyse yeast chromosomes

at a 300-bp resolution, each of the 300-bp region being covered by at least 60 probes. ChIP-chip of asynchronously growing cells was carried out as described (Katou *et al.*, 2003; Katou *et al.*, 2006; Bermejo *et al.*, 2007; Bermejo *et al.*, 2009; Gómez-González *et al.*, 2011). Briefly, we disrupted 1.5x107 cells by multi-beads shocker (MB400U, Yasui Kikai, Japan), which was able to keep cells precisely at lower than 61°C during the process. Anti-FLAG monoclonal antibody M2 (Sigma-Aldrich) was used for ChIP. ChIP DNA was purified and amplified by random priming using a WGA2 kit (SIGMA-Aldrich) following the manufacturer's procedure. A total of 4 µg of amplified DNA was digested with DNase I to a mean size of 100 bp, purified, and the fragments were end-labelled with biotin-N6-ddATP[23]. Microarray hybridization, washing, labelling, and scanning were performed by the CABIMER's Genomics Unit according to manufacturer's instructions.

12. DNA:RNA Hybrid Immunoprecipitation (DRIP)

100ml mid-log cultures grown in YPAD at 30°C were collected and washed with chilled water and carefully resuspended in 1.4 ml of Spheroplasting buffer (1 M sorbitol, 10 mM EDTA pH 8 with freshly added 0.1% β-mercaptoethanol and 2 mg/ml Zymoliase 20T) and incubated at 30°C for 30 min. under soft agitation. Spheroplast were pelleted, rinsed with water and homogeneously resuspended in 1.65ml of Buffer G2 (800mM Guanidine HCl, 30 mM Tris-Cl pH 8, 30 mM EDTA pH 8, 5% Tween-20, 0.5% triton X-100) before adding 10µl 10mg/ml RNase A and incubating for 30 minutes at 37°C. Next, samples were treated with 75 µl of 20 mg/ml proteinase K (Roche) 1 hour with gentle agitation. DRIP was performed mainly as described (Ginno *et al.*, 2012) with some modifications. DNA was extracted gently with chloroform:isoamyl alcohol 1:24. Isopropanol precipitated DNA was spooled with a glass rod, washed in 70% EtOH, resuspended gently in TE and digested overnight with 50 U of *HindIII, EcoRI, BsrGI, XbaI and SspI,* 2 mM spermidine and BSA. For the negative control, half of the DNA was treated with 8 µl RNase H (New England BioLabs) overnight at 37°C. Both treated and untreated samples were bound to 10 µl of S9.6 antibody (1mg/ml) in 500 µl binding buffer (10 mM $NaPO_4$

pH 7.0,140 mM NaCl, 0.05% triton X-100) in TE, overnight at 4°C. Hybrid-antibody complexes were immunoprecipitated using Dynabeads Protein A (Invitrogen) during 2 h at 4°C and washed 3 times with 1x binding buffer. DNA was eluted in 100 µl DRIP elution buffer (50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS) with 7 µl proteinase K, for 45 min at 55°C, maintaining constant shaking. Eluted samples were purified with Quiagen DNA purification kit (QUIAGEN).

13. ChIP and DRIP data quantification and normalization

Sample quantifications were performed by quantitative PCR (qPCR) as described in Material and Methods 37.2. Means and SDs were calculated from at least 3 independent experiments. Sample quantifications by qPCR were performed in triplicate. Primers used are listed in Table **Table M5**.

– **ChIP experiments:** 1/50 and 1/4 dilutions of the Input and the IP were typically used. DNA ratios in the different regions were calculated from the IP with respect to the Input. An experiment using an equivalent amount of Dynabeads Protein A or G was included as the background control and background values were subtracted. The intergenic region at positions 9716-9864 of chromosome V was used as a negative control.

– **DRIP experiments:** 1/25 and no dilution for Input and IP were used for que qPCR. The relative abundance of DNA:RNA hybrid immunoprecipitated in each region was normalized to the signal obtained in the Inputs.

14. RNA analysis

14.1. Northern blot

RNA was extracted from mid-log cultures using acid phenol (Köhrer & Domdey, 1991) and northern was performed following standard procedures. Solutions were previously treated with 0.1% DEPC. Cells were cultured in SRaff medium and transferred to galactose 2%-containing medium for analysis of GAL genes activation. The DNA probes used in the hybridization experiments are listed in **Table M5**.

14.2. Microarray analysis of gene expression

Global expression analyses were performed with Affymetrix platform GeneChip Yeast Genome 2.0 Array. Total RNA was isolated from 50 ml of mid-log growing cultures using the RNeasy Midi kit (Qiagen). The relative RNA levels for all yeast genes were determined using an Affymetrix microarray scanner. For the mutants, microarray analysis was conducted in triplicate and the values presented represent the average of these three determinations, the wildtype was performed only once and standardised with published data of experiments performed in the same conditions. A total of 7.5 μg of RNA were used for cDNA synthesis, labelling and microarray hybridization, performed by the CABIMER's Genomics Unit according to the manufacturer's instructions.

15. Statistical analyses

- **Microarray data analysis**

Microarray data were normalized by RMA (robust microarray average) and statistically analyzed by LIMMA (Linear Models for Microarray Analysis) comparing the mutant expression profile with its isogenic wild-type strain. Genes with expression levels below 350 (the median expression value of non-expressed meiotic genes in these experiments) were removed from the analysis to reduce false positive. Genes showing at least a 1.5-fold expression change with a P-value<0.01with a false discovery rate (FDR) correction were considered as altered. GO analyses were performed using the GO Term Finder tool from the Saccharomyces Genome Database (http://www.yeastgenome.org). A P-value<0.01 was established to consider GO terms as significantly enriched.

- **ChIP-chip data analysis**

ChIP-chip data were analyzed using the Tiling Array Suite software (TAS) from Affymetrix. TAS produces, per each probe position, the signal and the change P-value, taking into account the probes localized within a given bandwidth around the inspected probe. Two experiments per each strain were normalized using the quartiles method. Protein chromosomal distribution was then analyzed by detect-

ing binding clusters, which were defined as ranges within the chromosome satisfying the following conditions: estimated signal (log_2 IP/SUP binding ratio) positive in the whole range; P-value<0.01, minimum run of 100 bp and maximum gap of 250 bp. The results were visualized with the UCSC Genome Browser, developed and maintained by the Genome Bioinformatics Group (Center for Biomolecular Science and Engineering at the University of California, Santa Cruz) (http://genome.ucsc.edu/).

For statistical analysis of the functional and structural features of the genes, expression levels were taken from microarray of wild-type cells. Distribution of binding sites along genes was carried out as previously described (Gómez-González *et al.*, 2011). For this and other analyses, Perl scripts developed by Prof. Antonio Marín (Genetics Department, University of Seville) were used. Top-500 Rrm3-bound ORFs were selected as the ORFs with the highest signal log_2 ratio average in a sliding window of 200 bp. Average binding intensities along different genomic regions was calculated using tools from the Galaxy website (https://usegalaxy.org/).

- **Statistical tests**

Statistical tests (Student's t-test, Mann-Whitney U-test, Binomial test, Pearson's correlation) were calculated using GraphPad Prism software. In general a P-value<0.05 was considered as statistically significant.

16. Primers and Probes

– **Radioactive signal quantification:** Radioactive signals were acquired using PhosphorImager Fujifilm FLA-5100 and were quantified using ImageGauge.

– **Northern blot analysis:** Signals were normalized to the signal of the *SCR1* or *rDNA* genes, which are very stable and transcribed by the RNAPIII or RNAPI, respectively, and were expressed as arbitrarily units (A.U.).

– **PFGE analysis:** The percentage of the signal of each well with respect to all the signals in the same lane was calculated.

Table M5. Primers and probes used in this study

Name	Sequence	Usage
	qPCR	
GCN4-3 Fw	TTGTGCCCGAATCCAGTGA	DRIP. 3' end of GCN4 gene
GCN4-3 Rv	TGGCGGCTTCAGTGTTTCTA	
PDC1 Fw	GAAGGTATGAGATGGGCTGGTAA	DRIP. 5' end of the PDC1 gene
PDC1 Rv	CCTTGATACGAGCGTAACCATCA	
GAL1 Reg2 Fw	AAACAGGGCTTTAGTGTTGACGAT	ChIP. Middle region of GAL1 gene
GAL1 Reg2 Rv	CTCTTGTGAATTCTTCGCGAGAA	
V1	TGTTCCTTTAAGAGGTGATGGTGAT	ChIP. Non-transcribed region at positions 9716-9864 of ChrV
V2	GTGCGCAGTACTTGTGAAAACC	
	Standard PCR	
HPR1-MX6nt-Fw	TAACAATTCAAGAGGCATTAAAACTTG-GGCAAAGGAGTAATAATGGATCCCCGG-GTTAATTAAGG	Universal primer for cassettes MX6nt to remove HPR1 gene
HPR1-MX6nt-Rv	ATGAATTTCTTATCAGTTTAAAATTTC-TATTAAGAGGATAATTTATAGTGGATCT-GATATCATCG	
MPH1-MX6nt-Fw	CATTCCGGTTCTGTTTTATTTTAGT-GTCCTTTTTTCTCTCTGATGCGGATC-CCCGGGTTAATTAA	Universal primer for cassettes MX6nt to remove MPH1 gene
MPH1-MX6nt-Rv	AGCGTTATTTTTGTATAGACGCCGACG-TATAAGAGTCTCCTATCATAGTGGATCT-GATATCATCG	
DBP1-MX6nt -Fw	TTAAGAAAAACCCCTTTGAGTGAAAG-TATTACAAGAAAAACGGATCCCCGGGT-TAATTAA	Universal primer for cassettes MX6nt to remove DBP1 gene
DBP1-MX6nt-Rv	TTAAAGGAGTTCTATATTTGGATT-AGTCTTTTATTCTTTCTAGTGGATCTGA-TATCATCG	
BAR1C	TTAGAGATGCGTTGTCCCTG	Primers to remove BAR1 gene
BAR1D	CGTCATCCTAAACGTCCGTA	
lacZ 400 Fw +KpnI	TGTACGGTACCATGGCGATTACCGTTG	Primers to amplify lacZ fragment
lacZ 400 Rv +BamHI	TAGTAGGATCCTTATTTTTGACACCAG	
	Probes	
ADE5,7-5	TACCTAAGCGTTAAGAAATCGTCTA	ADE5,7 probe
ADE5,7-3	AGAGACACCTGAGTCTGCGTATGTG	
rDNA Fw	TTGGAGAGGGCAACTTTGG	rDNA probe
rDNA Rv	CAGGATCGGTCGATTGTGC	
SCR1 Fw	AGGCTGTAATGGCTTTCTGG	SCR1 probe
SCR1 Rv	GTTCAGGACACACTCCATCC	
GAL10-A	CTATTGAGGGTACGGAGATTATGGTGCG	GAL10 probe
GAL10-B	CCGCCGAGTACATGCTGATAGATAATGA	
GAL1 Fw	TGCACCGGAAAGGTTTGC	GAL1 probe
GAL1 Rv	CTCTTGTGAATTCTTCGCGAG	
Kan probe Fw	GCGCCAGATCTGTTTAGCTT	KAN probe
Kan probe Rv	CACTGGATGGCGGCGTTAGT	

Table M5. Primers and probes used in this study (Continuation)

Name	Sequence	Usage
RECQL5 Fw	TCCTGACTACTTGCGTCTGG	RECQL5 probre
RECQL5 Rv	CTTGCGGCCTCCCTCGTACA	

17. Protein analysis

17.1. Protein extraction

– TCA protocol

For protein extraction from mid-log yeast culture 10 mL of each strain culture were recovered and kept in ice. In the case of UV treatment, the rest of the culture was centrifuged, resuspended in distilled, sterile water and irradiated in plates as described above (Material and Methods 7 and 12). They were immediately re-suspended in fresh medium and were recovered at the indicated times. Proteins were extracted from pellets by adding 200 µl of cold 10% TCA and 200 µl of glass beads, then vortexing 7 times 20 seconds each time at 4°C. Supernatant was recovered and beads washed twice with 200 µl of cold 10 % TCA. Samples were centrifuged 10 minutes at 3000 rpm and supernatant discarded. The remaining pellet was resuspended using 100 µl of 2x Loading Buffer (62.5 mM Tris-HCl pH 6.8, 25% glycerol, 2% SDS, 0.01% water-diluted Bromophenol Blue, 5% β-mercaptoethanol) supplemented with protease inhibitors (1mM PMSF, 66 µg/ml chymostatin), 50 µl of water and 50 µl of 1 M Tris (not-adjusted pH). Prior to gel loading samples were boiled for 5 minutes and centrifuged 10 minutes at 3000 rpm at RT.

– Urea protocol

When posttranslational modifications were not essential for the analysis, extraction method based in a 8M Urea solution was used. Briefly, pellets from 10ml of culture were placed at 95°C 5minutes, immediately after that, cells were resuspended in 150µl of urea solution (8M UREA, SIGMA, dissolved in miliQ water) and placed back at 95°C 5min. Next, 200 µl of glass beads were added and vortexed at RT for 10 minutes at maximum speed. Following, 60µl of SDS 10% and 30µl of

Tris-HCl pH 7.5 were added, mixed and placed at 95°C 5minutes. Two new vortexing cycles where performed for 15 second to finally centrifuge the samples for 5 minutes at max speed. Supernatant was recovered to a clean eppendorf. Samples could be kept at -20°C or prepared for electrophoresis, usually taking 4µl of sample, 4µl of 4x LB and 8µl of water, boiling for 5 minutes, consolidating by centrifugation and keeping on ice until loading.

– Co-IP protocol

For protein complex precipitation 50 ml of cultured cells were harvested and washed. 100 µL of glass beads and 166 µL of Lysis Buffer (50mM Tris HCl pH 7.5, 100mM NaCl, 1.5mM MgCl2, 0.0075% NP40 10%, 100mM DTT) were added and samples were vortexed in the FastPrep (Intensity 5m/s) 4x20''. After separating the beads, samples were centrifuged 10' at 13000 rpm. Protein amount in supernatant (crude extract) was quantified using Bradford reagent. 1µl of each protein extract was diluted 1:10 and 1µl, 2µl and 3µl were diluted to a final volume of 50µl of Lysis Buffer, added to 950 µl of Bradford 1:4 and protein concentration was determined by spectrophotometry at 598 optical density by comparison to a standard calibrated curve of 0 – 0.8% BSA protein dilutions. Extracts were homogenised to 5mg/ml, 10µl were separated for INPUT and the remainder was incubated with 10µl of a FLAG antibody attached to Dynabeads Protein A for 3h at 4°C. Magnetic beads were centrifuged, washed 3x with ice-cold PBS 1x and resuspended in 40µl Laemmly Buffer 1x. Samples were boiled 5' before loading the gel.

17.2. Sodium dodecyl sulfate Polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated in 29:1 acrylamide:bis-acrylamide gels with concentrations appropriate to the molecular size of the proteins to study. SDS-PAGE was performed according to previously described method (Laemmli, 1970). For Rad53 detection, 8% 37.5:1 acrylamide:bis-acrylamide SDS-PAGE or a 4-20% gradient SDS-PAGE CriterionTM TGXTM Precast Gels (BioRad) were used. Electrophoreses were performed in a Mini-PROTEAN 3 Cell or in a Criterion Cell (BioRad) with Running Buffer (25 mM Tris base pH 8.3, 194 mM glycine, 0.1 % SDS buffer) at

100 V. Page RulerTM (Fermentas, CA) was used as a protein marker

17.3. Western Blot Analysis

For Western blot, proteins were wet-transferred using Trans-Blot system (Biorad) for 2 hours at 400 mA in Transfer Buffer (6 g/l Tris base, 28.8 g/l glycine and 0.5% SDS plus 20% methanol) or 15 hours at 30 V using Tris-Glycine Buffer (3 g/l Tris base and 14.32 g/l glycine).

ECL developing: Proteins were transferred to a nitrocellulose membrane (Hybond-ECL, GE Healthcare). Membranes were stained with Ponceau S (0.1% w/v Ponceau (SIGMA) in 5% Acetic acid) to check protein loading and transference. Membranes were blocked with TBS- 0.05% Tween-20 (TBS-T) - 5 % milk or Blocking Buffer (ROCHE) for at least 1 hour and primary antibodies were incubated during 1 to 3 hours at RT as indicated in Table **Table M1** After 3 washes with TBS-T of 10 minutes each, membranes were incubated with the corresponding secondary antibodies conjugated with the horseradish peroxidase for 1 hour and washed again. Finally, SuperSignalR West Pico (Pierce) or Immobilon Western Chemiluminescent HRP Substrate (Millipore) was used for chemiluminescence detection depending on the expected strength of the signal.

Odyssey Scanning: A PVDF membrane with low fluorescence background (Inmobilon-FL, Millipore) was used. This membrane was first activated in methanol for 30 s and equilibrated in transfer buffer before the transference and activated with methanol for 5 s after the procedure. Commercial Odyssey Blocking Buffer (LI-COR Biosciences) was used for blocking the membrane for at least 45minutes at RT. Primary antibody was prepared to the appropriate dilution (See Table **Table M1**) in B.O. + 0,01% Tw*een20* and incubated for 1 to 2 hours. Three washes of 10 minutes were performed with 1x TBS + 0,01% Tw*een20* followed by incubation of 1 hour with IRDye secondary antibodies. Finally, membranes where washed again 3 times, rinsed in 1x TBS and immediately scanned or left predrying. Image acquisition was performed in an Odyssey CLx Imager (LI-COR Biosciences).

18. Miscellanea

Standard molecular biology techniques were carried out following common procedures for bacterial transformation, yeast DNA and RNA extraction and other standard molecular biology techniques were carried out following common and manufacture's procedures

Apendix

Apendix 1.
List of 400 deletion strains that composed the mini-collection used for the screening.

ORF	Name	ORF	Name	ORF	Name
YAL040C	CLN3	YOR038C	HIR2	YKL032C	IXR1
YAL021C	CCR4	YPL167C	REV3	YKL054C	DEF1
YAL015C	NTG1	YPL164C	MLH3	YKL057C	NUP120
YAL011W	SWC3	YPL139C	UME1	YKL068W	NUP100
YAR002W	NUP60	YPL138C	SPP1	YKL110C	KTI12
YAR003W	SWD1	YPL121C	MEI5	YKL114C	APN1
YLL019C	KNS1	YPL116W	HOS3	YKL149C	DBR1
YLR013W	GAT3	YPL101W	ELP4	YKL160W	ELF1
YLR014C	PPR1	YBR188C	NTC20	YGR044C	RME1
YLR016C	PML1	YBR195C	MSI1	YGR057C	LST7
YLR085C	ARP6	YBR223C	TDP1	YGR066C	---
YLR095C	IOC2	YBR228W	SLX1	YGR067C	---
YLR113W	HOG1	YBR233W	PBP2	YGR102C	GTF1
YML081W	TDA9	YBR245C	ISW1	YGR104C	SRB5
YML062C	MFT1	YDR108W	TRS85	YOR111W	---
YMR153W	NUP53	YDR117C	TMA64	YOR123C	LEO1
YML060W	OGG1	YDR121W	DPB4	YOR144C	ELG1
YML011C	RAD33	YDR146C	SWI5	YOR166C	SWT1
YMR044W	IOC4	YDR359C	EAF1	YOR197W	MCA1
YMR019W	STB4	YDR399W	HPT1	YOR213C	SAS5
YML061C	PIF1	YDR419W	RAD30	YOR228C	MCP1
YMR167W	MLH1	YDR423C	CAD1	YOR246C	ENV9
YMR179W	SPT21	YDR432W	NPL3	YOR258W	HNT3
YMR190C	SGS1	YEL037C	RAD23	YOR288C	MPD1
YMR201C	RAD14	YEL056W	HAT2	YJL206C	---
YMR219W	ESC1	YER032W	FIR1	YJL176C	SWI3
YMR224C	MRE11	YER041W	YEN1	YLR381W	CTF3
YMR284W	YKU70	YER045C	ACA1	YLR385C	SWC7
YNL330C	RPD3	YER051W	JHD1	YLR392C	ART10
YNL309W	STB1	YER068W	MOT2	YLR398C	SKI2
YNL253W	TEX1	YER085C	---	YLR401C	DUS3
YOR023C	AHC1	YGR123C	PPT1	YLR407W	---
YOR032C	HMS1	YGR129W	SYF2	YLR418C	CDC73
YPL178W	CBC2	YGR200C	ELP2	YLR247C	IRC20
YOR051C	ETT1	YGR212W	SLI1	YLR318W	EST2
YOR080W	DIA2	YHL009C	YAP3	YDR163W	CWC15
YOR083W	WHI5	YHR087W	RTC3	YDR169C	STB3
YOR339C	UBC11	YHR124W	NDT80	YDR192C	NUP42
YOR344C	TYE7	YHR134W	WSS1	YDR217C	RAD9
YOR346W	REV1	YHR206W	SKN7	YGL222C	EDC1
YOR386W	PHR1	YCL061C	MRC1	YGL241W	KAP114
YOL001W	PHO80	YCR014C	POL4	YGL244W	RTF1

ORF	Name
YOL004W	SIN3
YOL051W	GAL11
YPL230W	USV1
YPL216W	---
YPL213W	LEA1
YPL055C	LGE1
YPL048W	CAM1
YPL042C	CDK8
YPL037C	EGD1
YPL015C	HST2
YPL008W	CHL1
YPL001W	HAT1
YPR135W	CTF4
YPR164W	MMS1
YPR179C	HDA3
YPR196W	---
YDR363W-A	SEM1
YFR034C	PHO4
YGR249W	MGA1
YBL006C	LDB7
YBL008W	HIR1
YBL019W	APN2
YBL032W	HEK2
YBL066C	SEF1
YBL088C	TEL1
YGL025C	PGD1
YGL013C	PDR1
YPL022W	RAD1
YGL043W	DST1
YGL082W	---
YGL087C	MMS2
YGL096W	TOS8
YNL236W	SIN4
YNL230C	ELA1
YNL218W	MGS1
YNL156C	NSG2
YKL213C	DOA1
YKL214C	YRA2
YDR253C	MET32
YDR255C	RMD5
YDR273W	DON1
YDR279W	RNH202
YDR295C	HDA2
YDR307W	PMT7

ORF	Name
YLR451W	LEU3
YLR135W	SLX4
YLR154C	RNH203
YKL009W	MRT4
YKL020C	SPT23
YIL079C	AIR1
YIL084C	SDS3
YIL008W	URM1
YFL049W	SWP82
YFL052W	ZNF1
YGR270W	YTA7
YGR275W	RTT102
YGR288W	MAL13
YIR018W	YAP5
YIR033W	MGA2
YKL033W-A	---
YKR077W	MSA2
YKR080W	MTD1
YMR075W	RCO1
YMR078C	CTF18
YMR080C	NAM7
YMR091C	NPL6
YOL090W	MSH2
YOL100W	PKH2
YOL104C	NDJ1
YER064C	VHR2
YER088C	DOT6
YER092W	IES5
YER095W	RAD51
YGR134W	CAF130
YHR191C	CTF8
YHR193C	EGD2
YLR032W	RAD5
YLR035C	MLH2
YLR052W	IES3
YMR312W	ELP6
YNL250W	RAD50
YML095C	RAD10
YML102W	CAC2
YML103C	NUP188
YML109W	ZDS2
YML113W	DAT1
YML121W	GTR1
YMR106C	YKU80

ORF	Name
YGL251C	HFM1
YGR001C	EFM5
YGR006W	PRP18
YPL086C	ELP3
YPL064C	CWC27
YJL127C	SPT10
YJL115W	ASF1
YJL092W	SRS2
YJL065C	DLS1
YJL056C	ZAP1
YJL049W	CHM7
YJL047C	RTT101
YER063W	THO1
YHR041C	SRB2
YJR078W	BNA2
YJR082C	EAF6
YJR147W	HMS2
YKR092C	SRP40
YKR099W	BAS1
YKR101W	SIR1
YLR435W	TSR2
YML021C	UNG1
YNR063W	---
YJR094C	IME1
YJR124C	---
YJR140C	HIR3
YKL005C	BYE1
YDL200C	MGT1
YDL214C	PRR2
YDR004W	RAD57
YBR271W	EFM2
YBR274W	CHK1
YCR065W	HCM1
YCR066W	RAD18
YCR077C	PAT1
YJR035W	RAD26
YJR043C	POL32
YJR050W	ISY1
YJR060W	CBF1
YNL004W	HRB1
YNL021W	HDA1
YNL025C	SSN8
YNL046W	---
YNR010W	CSE2

ORF	Name
YDR314C	RAD34
YDR317W	HIM1
YDR334W	SWR1
YIL017C	VID28
YIL024C	---
YIL040W	APQ12
YIL072W	HOP1
YIR005W	IST3
YIR009W	MSL1
YIR013C	GAT4
YNL072W	RNH201
YNL076W	MKS1
YNL082W	PMS1
YNL085W	MKT1
YNL090W	RHO2
YNL107W	YAF9
YNL121C	TOM70
YDR174W	HMO1
YGR040W	KSS1
YGR063C	SPT4
YKR023W	---
YKR029C	SET3
YBR131W	CCZ1
YBR150C	TBS1
YBR275C	RIF1
YBR289W	SNF5
YLR394W	CST9
YML027W	YOX1
YML036W	CGI121
YML041C	VPS71
YMR048W	CSM3
YMR137C	PSO2
YIL030C	SSM4
YOR270C	VPH1
YOR274W	MOD5
YOR276W	CAF20
YOR179C	SYC1
YKR095W	MLP1
YLR442C	SIR3
YNR052C	POP2
YPR023C	EAF3
YDR443C	SSN2
YGL151W	NUT1
YGL173C	XRN1

ORF	Name
YMR125W	STO1
YPR018W	RLF2
YPR051W	MAK3
YPR052C	NHP6A
YPR068C	HOS1
YPR070W	MED1
YPR101W	SNT309
YJL184W	GON7
YPR072W	NOT5
YLR226W	BUR2
YNL059C	ARP5
YMR263W	SAP30
YCR081W	SRB8
YBR278W	DPB3
YBR285W	---
YNL146W	---
YBR112C	CYC8
YLL054C	---
YLR011W	LOT6
YLR098C	CHA4
YML080W	DUS1
YMR280C	CAT8
YOR001W	RRP6
YOR006C	TSR3
YOR295W	UAF30
YOR297C	TIM18
YOL028C	YAP7
YOL043C	NTG2
YOL068C	HST1
YPL194W	DDC1
YPL129W	TAF14
YPL096W	PNG1
YBR175W	SWD3
YBR182C	SMP1
YBR184W	---
YDR050C	TPI1
YDR076W	RAD55
YDR078C	SHU2
YDR097C	MSH6
YDR123C	INO2
YDR364C	CDC40
YDR392W	SPT3
YDR408C	ADE8
YDR414C	ERD1

ORF	Name
YNR024W	MPP6
YBL103C	RTG3
YBR065C	ECM2
YNL136W	EAF7
YNL133C	FYV6
YIL128W	MET18
YIL130W	ASG1
YIL149C	MLP2
YIR002C	MPH1
YGR056W	RSC1
YOR141C	ARP8
YOR162C	YRR1
YOR161C	PNS1
YOR172W	YRM1
YOR191W	ULS1
YOR208W	PTP2
YBR033W	EDS1
YCR084C	TUP1
YLR399C	BDF1
YGL240W	DOC1
YGR229C	SMI1
YGR252W	GCN5
YAL027W	SAW1
YNL278W	CAF120
YOR290C	SNF2
YOR363C	PIP2
YOR368W	RAD17
YDR066C	RTR2
YDR069C	DOA4
YDR075W	PPH3
YDR092W	UBC13
YDR099W	BMH2
YDR369C	XRS2
YKL113C	RAD27
YHR034C	PIH1
YHR079C	IRE1
YGL058W	RAD6
YJL103C	GSM1
YHR204W	MNL1
YCL011C	GBP2
YGR036C	CAX4
YOL148C	SPT20
YGR258C	RAD2
YGL070C	RPB9

ORF	Name
YER161C	SPT2
YDL074C	BRE1
YFL013C	IES1
YJR047C	ANB1

ORF	Name
YER035W	EDC2
YGR159C	NSR1
YGR180C	RNR4

ORF	Name
YDR296W	MHR1
YGR262C	BUD32
YJL013C	MAD3

Bibliography

Abe, T. *et al.*, 2011. The Histone Chaperone Facilitates Chromatin Transcription (FACT) Protein Maintains Normal Replication Fork Rates. *Journal of Biological Chemistry*, 286(35): 30504–30512.

El Achkar, E., Gerbault-Seureau, M., Muleris, M., Dutrillaux, B. & Debatisse, M., 2005. Premature condensation induces breaks at the interface of early and late replicating chromosome bands bearing common fragile sites. *Proceedings of the National Academy of Sciences of the United States of America*, 102(50): 18069–74.

Aguilera, A., 2005. mRNA processing and genomic instability. *Nature structural & molecular biology*, 12(9): 737–8.

Aguilera, A., 2002. The connection between transcription and genomic instability. *EMBO Journal*, 21(3): 195–201.

Aguilera, A. & García-Muse, T., 2012. R Loops: From Transcription Byproducts to Threats to Genome Stability. *Molecular Cell*, 46(2): 115–124.

Aguilera, A. & Klein, H.L., 1989. Genetic and molecular analysis of recombination events in *Saccharomyces cerevisiae* occurring in the presence of the hyper-recombination mutation *hpr1*. *Genetics*, 122(3): 503–17.

Aguilera, A. & Klein, H.L., 1988. Genetic control of intrachromosomal recombination in *Saccharomyces cerevisiae*. I. Isolation and genetic characterization of hyper-recombination mutations. *Genetics*, 119(4): 779–90.

Aguilera, A. & Klein, H.L., 1989. Yeast intrachromosomal recombination: long gene conversion tracts are preferentially associated with reciprocal exchange and require the RAD1 and RAD3 gene products. *Genetics*, 123(4): 683–694.

Alcasabas, A. *et al.*, 2001. Mrc1 transduces signals of DNA replication stress to activate Rad53. *Nature cell biology*, 3(11): 958–65.

Alvaro, D., Lisby, M. & Rothstein, R., 2007. Genome-wide analysis of Rad52 foci reveals diverse mechanisms impacting recombination. *PLoS Genetics*, 3(12): 2439–2449.

Alzu, A., Bermejo, R., Begnis, M. & Lucca, C., 2012. Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell*.

Arora, R. *et al.*, 2014. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nature communications*, 5: 5220.

Ashton, T.M. & Hickson, I.D., 2010. Yeast as a model system to study RecQ helicase function. *DNA repair*, 9(3): 303–14.

Aygün, O. *et al.*, 2009. Direct inhibition of RNA polymerase II transcription by RECQL5. *The Journal of biological chemistry*, 284(35): 23197–203.

Aygün, O., Svejstrup, J. & Liu, Y., 2008. A RECQ5-RNA polymerase II association identified by targeted proteomic analysis of human chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, 105(25): 8580–4.

Azvolinsky, A., Giresi, P.G., Lieb, J.D. & Zakian, V. *a*, 2009. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. *Molecular cell*, 34(6): 722–34.

Baharoglu, Z., Lestini, R., Duigou, S. & Michel, B., 2010. RNA polymerase mutations that facilitate replication progression in the rep uvrD recF mutant lacking two accessory replicative helicases. Molecular Microbiology, 77(2): 324–336.

Baldacci, G., Chérif-Zahar, B. & Bernardi, G., 1984. The initiation of DNA replication in the mitochondrial genome of yeast. *The EMBO journal*, 3(9): 2115–20.

Balk, B. *et al.*, 2013. Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nature Structural & Molecular Biology*, 20(10): 1199–1205.

Banerjee, S. *et al.*, 2008. Mph1p promotes gross chromosomal rearrangement through partial inhibition of homologous recombination. *The Journal of cell biology*, 181(7): 1083–93.

Bell, S.P. & Labib, K., 2016. Chromosome Duplication in *Saccharomyces cerevisiae*. , 203(July): 1027–1067.

Bermejo, R. *et al.*, 2009. Genome-Organizing Factors Top2 and Hmo1 Prevent Chromosome Fragility at Sites of S phase Transcription. *Cell*, 138(5): 870–884.

Bermejo, R. *et al.*, 2011. The replication checkpoint protects fork stability by releasing transcribed genes from nuclear pores. *Cell*, 146(2): 233–46.

Bermejo, R. *et al.*, 2007. Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation. *Genes & Development*, 21(15): 1921–1936.

Bermejo, R., Lai, M.S. & Foiani, M., 2012. Preventing replication stress to maintain genome stability: resolving conflicts between replication and transcription. *Molecular cell*, 45(6): 710–8.

Bernstein, K. a, Gangloff, S. & Rothstein, R., 2010. The RecQ DNA helicases in DNA repair. *Annual review of genetics*, 44: 393–417.

Berthelot, K., Muldoon, M., Rajkowitsch, L., Hughes, J. & McCarthy, J.E.G., 2004. Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Molecular microbiology*, 51(4): 987–1001.

Beyer, A., Hollunder, J., Nasheuer, H.-P. & Wilhelm, T., 2004. Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. Molecular & cellular proteomics : MCP, 3(11): 1083–92.

Bhatia, V. *et al.*, 2014. BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*, 511(7509): 362–5.

Blobel, G., 1985. Gene gating: a hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 82(24): 8527–9.

Boeke, J.D., La Croute, F. & Fink, G.R., 1984. A positive selection for mutants lacking orotidine-5□-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *MGG Molecular & General Genetics*, 197: 345–346.

Boque-Sastre, R. *et al.*, 2015. Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America*, 112(18): 5785–90.

Boubakri, H., de Septenville, A.L., Viguera, E. & Michel, B., 2010. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *The EMBO journal*, 29(1): 145–57.

Boulé, J.-B. & Zakian, V.A., 2009. Characterization of the Helicase Activity and Anti-telomerase Properties of Yeast Pif1p In Vitro. In Methods in molecular biology (Clifton, N.J.). pp. 359–376.

Bourgeois, C.F., Mortreux, F. & Auboeuf, D., 2016. The multiple functions of RNA helicases as drivers and regulators of gene expression. Nature reviews. *Molecular cell biology*, 17(7): 426–38.

Le Breton, C. *et al.*, 2008. Srs2 removes deadly recombination intermediates independently of its interaction with SUMO-modified PCNA. *Nucleic acids research*, 36(15): 4964–74.

Brill, S.J., DiNardo, S., Voelkel-Meiman, K. & Sternglanz, R., 1987. Need for DNA topoisomerase activity as a swivel for DNA replication for transcription of ribosomal RNA. *Nature*, 326(6111): 414–416.

Byrne, K.P. & Wolfe, K.H., 2005. The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*, 15(10): 1456–1461.

Cabal, G.G. *et al.*, 2006. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, 441(7094): 770–773.

Cai, W. *et al.*, 2017. Wanted DEAD/H or Alive: Helicases Winding Up in Cancers. *Journal of the National Cancer Institute*, 109(6): *djw278*.

Castellano-Pozo, M. *et al.*, 2013. R loops are linked to histone H3 S10 phosphorylation and chromatin condensation. *Molecular cell*, 52(4): 583–90.

Castellano-Pozo, M., García-Muse, T. & Aguilera, A., 2012. R-loops cause replication impairment and genome instability during meiosis. *EMBO reports*, 13(10): 923–929.

Cerritelli, S.M. & Crouch, R.J., 2009. Ribonuclease H: the enzymes in eukaryotes. *FEBS Journal*, 276(6): 1494–1505.

Chakraborty, P. & Grosse, F., 2011. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA repair*, 10(6): 654–65.

Chan, Y. a *et al.*, 2014. Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS genetics*, 10(4): e1004288.

Chang, E.Y.-C. & Stirling, P.C., 2017. Replication Fork Protection Factors Controlling R-Loop Bypass and Suppression. *Genes*, 8(1): 33.

Chaudhuri, J. & Alt, F.W., 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nature Reviews Immunology*, 4(7): 541.

Chávez, S. *et al.*, 2000. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. *The EMBO journal*, 19(21): 5824–34.

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 & development*, 11(24): 3459–70.

Chavez, S., Garcia-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*. *Molecular and Cellular Biology*, 21(20): 7054–7064.

Chen, Y.-Z. *et al.*, 2004. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *American journal of human genetics*, 74(6): 1128–35.

Chen, Y.H. *et al.*, 2009. Interplay between the Smc5/6 complex and the Mph1 helicase in recombinational repair. *Proceedings of the National Academy of Sciences of the United States of America*, 106(1091–6490; 0027–8424; 50): 21252–21257.

Chu, W.K. & Hickson, I.D., 2009. RecQ helicases: multifunctional genome caretakers. Nature reviews. *Cancer*, 9(9): 644–54.

Cohen, I.R., Lajtha, A., Lambris, J.D. & Paoletti, R., 2013. DNA Helicases and DNA Motor Proteins M. Spies, ed., New York, NY: Springer New York.

Collins, A.R., 1999. Oxidative DNA damage, antioxidants, and cancer. *BioEssays*, 21(3): 238–246.

Croteau, D.L., Popuri, V., Opresko, P.L. & Bohr, V. *a*, 2014. Human RecQ helicases in DNA repair, recombination, and replication. *Annual review of biochemistry*, 83: 519–52.

Deshpande, A.M. & Newlon, C.S., 1996. DNA replication fork pause sites dependent on transcription. Science (New York, N.Y.), 272(5264): 1030–3.

Douglas, H.C. & Hawthorne, D.C., 1964. ENZYMATIC EXPRESSION AND GENETIC LINKAGE OF GENES CONTROLLING GALACTOSE UTILIZATION IN SACCHAROMYCES. *Genetics*, 49(5).

Drolet, M. *et al.*, 1995. Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proceedings of the National Academy of Sciences of the United States of America*, 92(8): 3526–30.

Duch, A. *et al.*, 2013a. Coordinated control of replication and transcription by a SAPK protects genomic integrity. *Nature*, 493(7430): 116–9.

Duch, A. *et al.*, 2013b. Coordinated control of replication and transcription by a SAPK protects genomic integrity. *Nature*, 493(7430): 116–9.

Dutta, A. *et al.*, 2015. Ccr4-Not and TFIIS Function Cooperatively To Rescue Arrested RNA Polymerase II. *Molecular and cellular biology*, 35(11): 1915–25.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E. & Nudler, E., 2011. Linking RNA polymerase backtracking to genome instability in E. coli. *Cell*, 146(4): 533–43.

Errico, A. & Costanzo, V., 2012. Mechanisms of replication fork protection: a safeguard for genome stability. *Critical Reviews in Biochemistry and Molecular Biology*, 47(3): 222–235.

Felipe-Abrio, I., Lafuente-Barquero, J., García-Rubio, M.L. & Aguilera, A., 2014. RNA polymerase II contributes to preventing transcription-mediated replication fork stalls. The EMBO journal: 1–16.

Foltman, M. *et al.*, 2013. Eukaryotic replisome components cooperate to process histones during chromosome replication. *Cell reports*, 3(3): 892–904.

Fridovich-Keil, J.L. & Jinks-Robertson, S., 1993. A yeast expression system for human galactose-1-phosphate uridylyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, 90(2): 398–402.

Fu, W., Ligabue, A., Rogers, K.J., Akey, J.M. & Monnat, R.J., 2017. Human RECQ Helicase Pathogenic Variants, Population Variation and "Missing" Diseases. *Human Mutation*, 38(2): 193–203.

Fuller-Pace, F. V., 2013. DEAD box RNA helicase functions in cancer. *RNA Biology*, 10(1): 121–132.

Gaillard, H. & Aguilera, A., 2016. Transcription as a Threat to Genome Integrity. *Annual Review of Biochemistry*, 85(1): 291–317.

Gaillard, H., García-Muse, T. & Aguilera, A., 2015. Replication stress and cancer. Nature reviews. *Cancer*, 15(5): 276–89.

Gaillard, H., Herrera-Moyano, E. & Aguilera, A., 2013. Transcription-associated genome instability. *Chemical reviews*, 113(11): 8638–61.

Gan, W. *et al.*, 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes & development*, 25(19): 2041–56.

Gangloff, S., Soustelle, C. & Fabre, F., 2000. Homologous recombination is responsible for cell death in the absence of the Sgs1 and Srs2 helicases. *Nature genetics*, 25(2): 192–4.

García-Muse, T. & Aguilera, A., 2016. Transcription-replication conflicts: how they occur and how they are resolved. *Nature reviews. Molecular cell biology*, 17(9): 553–63.

García-Rubio, M. *et al.*, 2008. Different physiological relevance of yeast THO/TREX subunits in gene expression and genome integrity. *Molecular Genetics and Genomics*, 279(2): 123–132.

García-Rubio, M., Huertas, P., González-Barrera, S. & Aguilera, A., 2003. Recombinogenic effects of DNA-damaging agents are synergistically increased by transcription in *Saccharomyces cerevisiae*. New insights into transcription-associated recombination. *Genetics*, 65(2): 457–66.

García-Rubio, M.L. *et al.*, 2015. The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS genetics*, 11(11): e1005674.

Gavaldá, S., Gallardo, M., Luna, R. & Aguilera, A., 2013. R-loop mediated transcription-associated recombination in *trf4*Δ mutants reveals new links between RNA surveillance and genome integrity. S. D. Fugmann, ed. *PloS one*, 8(6): e65541.

Gavaldá, S., Santos-Pereira, J.M., García-Rubio, M.L., 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.

Gietz, R.D., Schiestl, R.H., Willems, A.R. & Woods, R.A., 1995. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast*, 11(4): 355–60.

Ginno, P.A., Lim, Y.W., Lott, P.L., Korf, I. & Chédin, F., 2013. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. TL - 23. *Genome research*, 23 VN-r(10): 1590–1600.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I. & Ch??din, F., 2012. R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Molecular Cell*, 45(6): 814–825.

Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D.L., 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature structural & molecular biology*, 15(1): 71–8.

Gómez-González, B., García-Rubio, M., *et al.*, 2011. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *The EMBO journal*, 30(15): 3106–19.

Gómez-González, B. & Aguilera, A., 2007. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proceedings of the National Academy of Sciences of the United States of America*, 104(20): 8409–14.

Gómez-González, B., Ruiz, J.F. & Aguilera, A., 2011. Genetic and molecular analysis of mitotic recombination in *Saccharomyces cerevisiae*. H. Tsubouchi, ed. Methods in molecular biology (Clifton, N.J.), 745: 151–72.

González-Barrera, S., García-Rubio, M. & Aguilera, A., 2002. Transcription and double-strand breaks induce similar mitotic recombination events in *Saccharomyces cerevisiae*. *Genetics*, 162(2): 603–14.

Gottipati, P., Cassel, T.N., Savolainen, L. & Helleday, T., 2008. Transcription-Associated Recombination Is Dependent on Replication in Mammalian Cells. *Molecular and Cellular Biology*, 28(1): 154–164.

Greger, I.H. & Proudfoot, N.J., 1998. Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of *Saccharomyces cerevisiae*. *The EMBO journal*, 17(16): 4771–9.

Groh, M., Lufino, M.M.P., Wade-Martins, R. & Gromak, N., 2014. R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome A. Aguilera, ed. *PLoS Genetics*, 10(5): e1004318.

Gros, J. *et al.*, 2015. Post-licensing Specification of Eukaryotic Replication Origins by Facilitated Mcm2-7 Sliding along DNA. *Molecular Cell*, 60(5): 797–807.

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

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

Hanahan, D., 1983. Studies on transformation of Escherichia coli with plasmids. *Journal of molecular biology*, 166(4): 557–80.

Hanawalt, P.C. & Spivak, G., 2008. Transcription-coupled DNA repair: two decades of progress and surprises. *Nature Reviews Molecular Cell Biology*, 9(12): 958–970.

Hartwell, L.H. & Smith, D., 1985. Altered fidelity of mitotic chromosome transmission in cell cycle mutants of *S. cerevisiae*. *Genetics*, 110(3): 381–95.

Hatchi, E. *et al.*, 2015. BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair. *Molecular Cell*, 57(4): 636–647.

Hecht, A. & Grunstein, M., 1999. Mapping DNA interaction sites of chromosomal proteins using immunoprecipitation and polymerase chain reaction. *Methods in enzymology*, 304: 399–414.

Heller, R.C. & Marians, K.J., 2006. Replication fork reactivation downstream of a blocked nascent leading strand. *Nature*, 439(7076): 557–562.

Helmrich, A., Ballarino, M. & Tora, L., 2011. Collisions between Replication and Transcription Complexes Cause Common Fragile Site Instability at the Longest Human Genes. *Molecular Cell*, 44(6): 966–977.

Herrera-Moyano, E., Mergui, X., García-Rubio, M.L., Barroso, S. & Aguilera, A., 2014. The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. *Genes & development*, 28(7): 735–48.

Hill, S.J. *et al.*, 2014. Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. *Genes & development*, 28(17): 1957–75.

Hu, Y. *et al.*, 2007. RECQL5/Rec*ql5* helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. *Genes & development*, 21(23): 3073–84.

Hu, Y. *et al.*, 2005. Rec*ql5* and Blm RecQ DNA Helicases Have Nonredundant Roles in Suppressing Crossovers. *Molecular and cellular biology*, 25(9): 3431–3442.

Hu, Y., Lu, X., Zhou, G., Barnes, E.L. & Luo, G., 2009. Rec*ql5* plays an important role in DNA replication and cell survival after camptothecin treatment. *Molecular biology of the cell*, 20(1): 114–23.

Huertas, P. & Aguilera, A., 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular cell*, 12(3): 711–21.

Im, J.-S. *et al.*, 2014. ATR checkpoint kinase and CRL1βTRCP collaborate to degrade ASF1a and thus repress genes overlapping with clusters of stalled replication forks. *Genes & development*, 28(8): 875–87.

Islam, M.N. *et al.*, 2012. A variant of the breast cancer type 2 susceptibility protein (BRC) repeat is essential for the RECQL5 helicase to interact with RAD51 recombinase for genome stabilization. *The Journal of biological chemistry*, 287(28): 23808–18.

Islam, M.N., Fox, D., Guo, R., Enomoto, T. & Wang, W., 2010. RecQL5 promotes genome stabilization through two parallel mechanisms--interacting with RNA polymerase II and acting as a helicase. *Molecular and cellular biology*, 30(10): 2460–72.

Ito, H., Fukuda, Y., Murata, K. & Kimura, A., 1983. Transformation of intact yeast cells treated with alkali cations. *Journal of bacteriology*, 153(1): 163–8.

Ivessa, A.S. *et al.*, 2003. The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Molecular cell*, 12(6): 1525–36.

Ivessa, A.S., Zhou, J.-Q., Schulz, V.P., Monson, E.K. & Zakian, V.A., 2002. Saccharomyces Rrm3p, a 5' to 3' DNA helicase that promotes replication fork progression through telomeric and subtelomeric DNA. *Genes & development*, 16(11): 1383–96.

Ivessa, A.S., Zhou, J.Q. & Zakian, V.A., 2000. The Saccharomyces Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell*, 100(4): 479–89.

Jenjaroenpun, P., Wongsurawat, T., Yenamandra, S.P. & Kuznetsov, V. a., 2015. QmRLFS-finder: a model, web server and stand-alone tool for prediction and analysis of R-loop forming sequences. *Nucleic Acids Research*: 1–8.

Jimeno, S., Rondón, A.G., Luna, R. & Aguilera, A., 2002. The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *The EMBO journal*, 21(13): 3526–35.

Kanagaraj, R. *et al.*, 2010. RECQ5 helicase associates with the C-terminal repeat domain of RNA polymerase II during productive elongation phase of transcription. *Nucleic acids research*, 38(22): 8131–40.

Kanagaraj, R., Saydam, N., Garcia, P.L., Zheng, L. & Janscak, P., 2006. Human RECQ5beta helicase promotes strand exchange on synthetic DNA structures resembling a stalled replication fork. *Nucleic acids research*, 34(18): 5217–31.

Kassube, S. a, Jinek, M., Fang, J., Tsutakawa, S. & Nogales, E., 2013. Structural mimicry in transcription regulation of human RNA polymerase II by the DNA helicase RECQL5. *Nature Structural & Molecular Biology*, 20(April): 1–10.

Katou, Y. *et al.*, 2003. S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature*, 424(6952): 1078–83.

Katou, Y., Kaneshiro, K., Aburatani, H. & Shirahige, K., 2006. Genomic approach for the understanding of dynamic aspect of chromosome behavior. *Methods in enzymology*, 409: 389–410.

Kawabe, T. *et al.*, 2000. Differential regulation of human RecQ family helicases in cell transformation and cell cycle. *Oncogene*, 19(41): 4764–72.

Kawauchi, J., Mischo, H., Braglia, P., Rondon, A. & Proudfoot, N.J., 2008. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes & development*, 22(8): 1082–92.

Keskin, H. *et al.*, 2014. Transcript-RNA-templated DNA recombination and repair. Nature.

Kim, N. & Jinks-Robertson, S., 2012. Transcription as a source of genome instability. *Nature Reviews Genetics*, 13(3): 204–14.

Kitao, S. *et al.*, 1998. Cloning of Two New Human Helicase Genes of the RecQ Family: Biological Significance of Multiple Species in Higher Eukaryotes. *Genomics*, 54(3): 443–452.

Kizer, K.O. *et al.*, 2005. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Molecular and cellular biology*, 25(8): 3305–16.

Kogoma, T., 1997. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. Microbiology and molecular biology reviews : MMBR, 61(2): 212–38.

Köhrer, K. & Domdey, H., 1991. Preparation of high molecular weight RNA. *Methods in enzymology*, 194: 398–405.

Koyama, H., Ito, T., Nakanishi, T. & Sekimizu, K., 2007. Stimulation of RNA polymerase II transcript cleavage activity contributes to maintain transcriptional fidelity in yeast. Genes to cells : devoted to molecular & cellular mechanisms, 12(5): 547–59.

Koyama, H., Ueda, T., Ito, T. & Sekimizu, K., 2010. Novel RNA polymerase II mutation suppresses transcriptional fidelity and oxidative stress sensitivity in *rpb9*Δ yeast. *Genes to Cells*, 15(2): 151–159.

Krejci, L. *et al.*, 2003. DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature*, 423(6937): 305–309.

de la Loza, M.C.D., Wellinger, R.E. & Aguilera, a, 2009. Stimulation of direct-repeat recombination by RNA polymerase III transcription. *DNA repair*, 8(5): 620–6.

Labib, K. & Hodgson, B., 2007. Replication fork barriers: pausing for a break or stalling for time? EMBO reports, 8(4): 346–53.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259): 680–5.

León Ortiz, A.M., Reid, R.J.D., Dittmar, J.C., Rothstein, R. & Nicolas, A., 2011. Srs2 overexpression reveals a helicase-independent role at replication forks that requires diverse cell functions. *DNA repair*, 10(5): 506–17.

Li, M., Pokharel, S., Wang, J.-T., Xu, X. & Liu, Y., 2015. RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. *Nature communications*, 6: 6720.

Li, X., Wang, J. & Manley, J.L., 2005. Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. *Genes & Development*, 19(22): 2705–2714.

Lindahl, T., 1993. Instability and decay of the primary structure of DNA. *Nature*, 362(6422): 709–15.

Linskens, M.H. & Huberman, J.A., 1988. Organization of replication of ribosomal DNA in *Saccharomyces cerevisiae*. *Molecular and cellular biology*, 8(11): 4927–35.

Lisby, M., Barlow, J.H., Burgess, R.C. & Rothstein, R., 2004. Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell*, 118(6): 699–713.

Lisby, M. & Rothstein, R., 2015. Cell Biology of Mitotic Recombination. *Cold Spring Harbor Perspectives in Biology*, 7(3): a016535.

Lisby, M., Rothstein, R. & Mortensen, U.H., 2001. Rad52 forms DNA repair and recombination centers during S phase. *Proceedings of the National Academy of Sciences of the United States of America*, 98(15): 8276–82.

Liu, B. & Alberts, B.M., 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* (New York, N.Y.), 267(5201): 1131–7.

Lopes, M., Foiani, M. & Sogo, J.M., 2006. Multiple Mechanisms Control Chromosome Integrity after Replication Fork Uncoupling and Restart at Irreparable UV Lesions. *Molecular Cell*, 21(1): 15–27.

Lu, X., Lou, H. & Luo, G., 2011. A Blm-Re*cql5* partnership in replication stress response. *Journal of molecular cell biology*, 3(1): 31–8.

Luke, B. *et al.*, 2008. The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in *Saccharomyces cerevisiae*. *Molecular cell*, 32(4): 465–77.

Luna, R., Rondón, A.G. & Aguilera, A., 2012. New clues to understand the role of THO and other functionally related factors in mRNP biogenesis. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1819(6): 514–520.

Ma, W.K., Cloutier, S.C. & Tran, E.J., 2013. The DEAD-box Protein Dbp2 Functions with the RNA-Binding Protein Yra1 to Promote mRNP Assembly. *Journal of Molecular Biology*, 425(20): 3824–3838.

Maizels, N., 2005. Immunoglobulin Gene Diversification. *Annual Review of Genetics*, 39(1): 23–46.

Mankouri, H.W., Craig, T.J. & Morgan, A., 2002. SGS1 is a multicopy suppressor of *srs2*: functional overlap between DNA helicases. *Nucleic acids research*, 30(5): 1103–13.

Maréchal, A. & Zou, L., 2013. DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harbor perspectives in biology*, 5(9): a012716.

McElhinny, S.A.N. *et al.*, 2010. Genome instability due to ribonucleotide incorporation into DNA. *Nature Chemical Biology*, 6(10): 774–781.

Merrikh, H., Zhang, Y., Grossman, A.D. & Wang, J.D., 2012. Replication-transcription conflicts in bacteria. Nature reviews. *Microbiology*, 10(7): 449–58.

Meryet-Figuiere, M. *et al.*, 2014. Temporal separation of replication and transcription during S-phase progression. *Cell Cycle*, 13(20): 3241–3248.

Mirkin, E. V, Castro Roa, D., Nudler, E. & Mirkin, S.M., 2006. Transcription regulatory elements are punctuation marks for DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 103(19): 7276–81.

Mischo, H.E. *et al.*, 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Molecular cell*, 41(1): 21–32.

Mnaimneh, S. *et al.*, 2004. Exploration of essential gene functions via titratable promoter alleles. *Cell*, 118(1): 31–44.

Monnat, R.J., 2010. Human RECQ helicases: roles in DNA metabolism, mutagenesis and cancer biology. *Seminars in cancer biology*, 20(5): 329–39.

Moreira, M.-C. *et al.*, 2004. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nature Genetics*, 36(3): 225–227.

Nguyen, V.C. *et al.*, 2010. Replication stress checkpoint signaling controls tRNA gene transcription. *Nature Structural & Molecular Biology*, 17(8): 976–981.

Nudler, E., 2012. RNA Polymerase Backtracking in Gene Regulation and Genome Instability. *Cell*, 149(7): 1438–1445.

Orphanides, G., Lagrange, T. & Reinberg, D., 1996. The general transcription factors of RNA polymerase II. *Genes & development*, 10(21): 2657–83.

Pardo, B., Aguilera, A., Lupski, J., Schwartz, E. & Ehmsen, K., 2012. Complex Chromosomal Rearrangements Mediated by Break-Induced Replication Involve Structure-Selective Endonucleases M. Lichten, ed. *PLoS Genetics*, 8(9): e1002979.

Parker, D. *et al.*, 1996. Phosphorylation of CREB at Ser-133 induces complex formation with CREB-binding protein via a direct mechanism. *Molecular and cellular biology*, 16(2): 694–703.

Paul, B. & Montpetit, B., 2016. Altered RNA processing and export lead to retention of mRNAs near transcription sites and nuclear pore complexes or within the nucleolus. Molecular Biology of the Cell, 27(17): 2742–2756.

Pefanis, E. *et al.*, 2015. RNA Exosome-Regulated Long Non-Coding RNA Transcription Controls Super-Enhancer Activity. *Cell*, 161(4): 774–789.

Peña, Á. *et al.*, 2012. Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. *The EMBO Journal*, 31(6): 1605–1616.

Pfeiffer, V., Crittin, J., Grolimund, L. & Lingner, J., 2013. The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *The EMBO journal*, 32(21): 2861–71.

Piruat, J.I. & Aguilera, a, 1998. A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *The EMBO journal*, 17(16): 4859–72.

Piruat, J.I., Chávez, S. & Aguilera, A., 1997. The yeast HRS1 gene is involved in positive and negative regulation of transcription and shows genetic characteristics similar to SIN4 and GAL11. *Genetics*, 147(4): 1585–94.

Poli, J. *et al.*, 2016. Mec1, INO80, and the PAF1 complex cooperate to limit transcription replication conflicts through RNAPII removal during replication stress. *Genes & development*, 30(3): 337–54.

Pomerantz, R.T. & O'Donnell, M., 2008. The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature*, 456(7223): 762–6.

Popuri, V., Tadokoro, T., Croteau, D.L. & Bohr, V. *a*, 2013. Human RECQL5: Guarding the crossroads of DNA replication and transcription and providing backup capability. *Critical reviews in biochemistry and molecular biology*, 9238: 1–11.

Prado, F. *et al.*, 1997. Recombination between DNA repeats in yeast *hpr1*delta cells is linked to transcription elongation. *The EMBO journal*, 16(10): 2826–35.

Prado, F. & Aguilera, A., 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *The EMBO journal*, 24(6): 1267–76.

Prado, F. & Aguilera, A., 1995. Role of reciprocal exchange, one-ended invasion crossover and single-strand annealing on inverted and direct repeat recombination in yeast: different requirements for the RAD1, RAD10, and RAD52 genes. *Genetics*, 139(1): 109–23.

Prakash, R. *et al.*, 2009. Yeast Mphl helicase dissociates Rad51-made D-loops: Implications for crossover control in mitotic recombination. *Genes and Development*, 23(1): 67–79.

Proudfoot, N.J., 2011. Ending the message: poly(A) signals then and now. *Genes & Development*, 25(17): 1770–1782.

Proudfoot, N.J., 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science*, 352(6291).

Putnam, A.A. & Jankowsky, E., 2013. DEAD-box helicases as integrators of RNA, nucleotide and protein binding. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(8): 884–893.

Ramamoorthy, M. *et al.*, 2012. RECQL5 cooperates with Topoisomerase II alpha in DNA decatenation and cell cycle progression. *Nucleic acids research*, 40(4): 1621–35.

Rando, O.J. & Winston, F., 2012. Chromatin and Transcription in Yeast. *Genetics*, 190(2): 351–387.

Ray, A., Machin, N. & Stahl, F.W., 1989. A DNA double chain break stimulates triparental recombination in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 86(16): 6225–9.

Reyes, A. *et al.*, 2013. Mitochondrial DNA replication proceeds via a "bootlace" mechanism involving the incorporation of processed transcripts. *Nucleic Acids Research*, 41(11): 5837–5850.

Robert, T., Dervins, D., Fabre, F. & Gangloff, S., 2006. Mrc1 and Srs2 are major actors in the regulation of spontaneous crossover. *The EMBO journal*, 25(12): 2837–46.

Roberts, R.W. & Crothers, D.M., 1992. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science*, 258(5087): 1463–1467.

Rocak, S. & Linder, P., 2004. DEAD-box proteins: the driving forces behind RNA metabolism. *Nature Reviews Molecular Cell Biology*, 5(3): 232–241.

Rondón, A.G. *et al.*, 2003. Molecular evidence for a positive role of Spt4 in transcription elongation. *The EMBO journal*, 22(3): 612–20.

Rong, L. & Klein, H.L., 1993. Purification and characterization of the SRS2 DNA helicase of the yeast *Saccharomyces cerevisiae*. *The Journal of biological chemistry*, 268(2): 1252–9.

Salvi, J.S. *et al.*, 2014. Roles for Pbp1 and caloric restriction in genome and lifespan maintenance via suppression of RNA-DNA hybrids. *Developmental cell*, 30(2): 177–91.

Santos-Pereira, J.M. *et al.*, 2013. The Npl3 hnRNP prevents R-loopmediated transcription-replication conflicts and genome instability. *Genes and Development*, 27: 2445–2458.

Santos-Pereira, J.M. & Aguilera, A., 2015. R loops: new modulators of genome dynamics and function. *Nature Reviews Genetics*, 16(10): 583–597.

Santos-Pereira, J.M., Garcia-Rubio, M.L., Gonzalez-Aguilera, C., Luna, R. & Aguilera, A., 2014. A genome-wide function of THSC/TREX-2 at active genes prevents transcription-replication collisions. *Nucleic Acids Research*, 42(19): 12000–12014.

Saponaro, M. *et al.*, 2014. RECQL5 controls transcript elongation and suppresses genome instability associated with transcription stress. *Cell*, 157(5): 1037–49.

Sarkar, M. & Ghosh, M.K., 2016. DEAD box RNA helicases: crucial regulators of gene expression and oncogenesis. Frontiers in bioscience (Landmark edition), 21: 225–50.

Saunders, A., Core, L.J. & Lis, J.T., 2006. Breaking barriers to transcription elongation. *Nature Reviews Molecular Cell Biology*, 7(8): 557–567.

Schiestl, R.H. & Gietz, R.D., 1989. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Current genetics*, 16(5–6): 339–46.

Schürer, K.A., Rudolph, C., Ulrich, H.D. & Kramer, W., 2004. Yeast MPH1 gene functions in an error-free DNA damage bypass pathway that requires genes from Homologous recombination, but not from postreplicative repair. *Genetics*, 166(4): 1673–86.

Schwab, R.A. *et al.*, 2015. The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Molecular Cell*, 60(3): 351–361.

Sekedat, M.D. *et al.*, 2010. GINS motion reveals replication fork progression is remarkably uniform throughout the yeast genome. *Molecular systems biology*, 6: 353.

Selth, L. a, Sigurdsson, S. & Svejstrup, J.Q., 2010. Transcript Elongation by RNA Polymerase II. *Annual review of biochemistry*, 79: 271–93.

De Septenville, A.L., Duigou, S., Boubakri, H. & Michel, B., 2012. Replication fork reversal after replication-transcription collision. W. F. Burkholder, ed. *PLoS genetics*, 8(4): e1002622.

Shearwin, K.E., Callen, B.P. & Egan, J.B., 2005. Transcriptional interference – a crash course. *Trends in Genetics*, 21(6): 339–345.

Shermann, F., Fink, G. & Hicks, J., 1986. Methods in yeast genetics, Cold Spring Harbord, NY.

Shimamoto, A., Nishikawa, K., Kitao, S. & Furuichi, Y., 2000. Human RecQ5beta, a large isomer of RecQ5 DNA helicase, localizes in the nucleoplasm and interacts with topoisomerases 3alpha and 3beta. *Nucleic acids research*, 28(7): 1647–55.

Silva, S. *et al.*, 2016. Mte1 interacts with Mph1 and promotes crossover recombination and telomere maintenance. *Genes & Development*, 30(6): 700–717.

Singh, B.N. & Hampsey, M., 2007. A transcription-independent role for TFIIB in gene looping. *Molecular cell*, 27(5): 806–16.

Singh, D.K., Ghosh, A.K., Croteau, D.L. & Bohr, V. *a*, 2011. RecQ helicases in DNA double strand break repair and telomere maintenance. *Mutation research*: 1–10.

Skourti-Stathaki, K., Kamieniarz-Gdula, K. & Proudfoot, N.J., 2014. R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*.

Skourti-Stathaki, K., Proudfoot, N.J. & Gromak, N., 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular cell*, 42(6): 794–805.

Sollier, J. *et al.*, 2014. Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Molecular Cell*, 56(6): 777–785.

Sparks, J.L. *et al.*, 2012. RNase H2-Initiated Ribonucleotide Excision Repair. *Molecular cell*, 47(6):890-986.

Sperling, A.S., Jeong, K.S., Kitada, T. & Grunstein, M., 2011. Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 108(31): 12693–8.

Sridhara, S.C. *et al.*, 2017. Transcription Dynamics Prevent RNA-Mediated Genomic Instability through SRPK2-Dependent DDX23 Phosphorylation. *Cell Reports*, 18(2):334-343

Srivatsan, A., Tehranchi, A., MacAlpine, D.M. & Wang, J.D., 2010. Co-orientation of replication and transcription preserves genome integrity. N. A. Moran, ed. *PLoS genetics*, 6(1): e1000810.

Stirling, P.C. *et al.*, 2012. R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. *Genes & development*, 26(2): 163–75.

Strässer, K. *et al.*, 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature*, 417(6886): 304–8.

Strathern, J. *et al.*, 2013. The fidelity of transcription: RPB1 (RPO21) mutations that increase transcriptional slippage in *S. cerevisiae*. *The Journal of biological chemistry*, 288(4): 2689–99.

Strauss, E.J. & Guthrie, C., 1994. PRP28, a "DEAD-box" protein, Is required for the first step of mRNA splicing in vitro. *Nucleic Acids Research*, 22(15): 3187–3193.

Stuckey, R., García-Rodríguez, N., Aguilera, A. & Wellinger, R.E., 2015. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proceedings of the National Academy of Sciences of the United States of America*, 112(18): 5779–84.

Sung, P. & Klein, H., 2006. Mechanism of homologous recombination: mediators and helicases take on regulatory functions. Nature reviews. *Molecular cell biology*, 7(10): 739–50.

Tehranchi, A.K. *et al.*, 2010. The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell*, 141(4): 595–605.

Teytelman, L., Thurtle, D.M., Rine, J. & van Oudenaarden, A., 2013. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 110(46): 18602–7.

Tkach, J.M. *et al.*, 2012. Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nature Cell Biology*, 14(9): 966–976.

Torres, J.Z., Bessler, J.B. & Zakian, V.A., 2004. Local chromatin structure at the ribosomal DNA causes replication fork pausing and genome instability in the absence of the *S. cerevisiae* DNA helicase Rrm3p. *Genes & development*, 18(5): 498–503.

Tresini, M. *et al.*, 2015. The core spliceosome as target and effector of non-canonical ATM signalling. *Nature*, 523(7558): 53–8.

Tu, W.-Y., Huang, Y.-C., Liu, L.-F., Chang, L.-H. & Tam, M.F., 2011. Rpl12p affects the transcription of the PHO pathway high-affinity inorganic phosphate transporters and repressible phosphatases. Yeast (Chichester, England), 28(6): 481–93.

Tuduri, S. *et al.*, 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature cell biology*, 11(11): 1315–24.

Wach, A., Brachat, A., Pöhlmann, R. & Philippsen, P., 1994. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. Yeast (Chichester, England), 10(13): 1793–808.

Wahba, L., Amon, J.D., Koshland, D. & Vuica-Ross, M., 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Molecular cell*, 44(6): 978–88.

Wahba, L., Costantino, L., Tan, F.J., Zimmer, A. & Koshland, D., 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes & development*, 30(11): 1327–38.

Wahba, L., Gore, S.K. & Koshland, D., 2013. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. eLife, 2: e00505–e00505.

Watson, J.D. & Crick, F.H., 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356): 737–8.

Wei, X. *et al.*, 1998. Segregation of transcription and replication sites into higher order domains. Science (New York, N.Y.), 281(5382): 1502–6.

Wellinger, R.E., Prado, F. & Aguilera, A., 2006a. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Molecular and cellular biology*, 26(8): 3327–34.

Wellinger, R.E., Prado, F. & Aguilera, A., 2006b. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Molecular and cellular biology*, 26(8): 3327–34.

Whitby, M.C., 2010. The FANCM family of DNA helicases/translocases. *DNA Repair*, 9(3): 224–236.

Winkler, D.D. & Luger, K., 2011. The Histone Chaperone FACT: Structural Insights and Mechanisms for Nucleosome Reorganization. *Journal of Biological Chemistry*, 286(21): 18369–18374.

Winzeler, E. a., 1999. Functional Characterization of the *S. cerevisiae* Genome by Gene Deletion and Parallel Analysis. *Science*, 285(5429): 901–906.

Yan, H., Gibson, S. & Tye, B.K., 1991. Mcm2 and Mcm3, two proteins important for ARS activity, are related in structure and function. *Genes & development*, 5(6): 944–57.

Yeo, C.Q.X. *et al.*, 2016. p53 Maintains Genomic Stability by Preventing Interference between Transcription and Replication. *Cell Reports*, 15(1): 132–146.

Yu, G., Wang, L.-G. & He, Q.-Y., 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* (Oxford, England), 31(14): 2382–3.

Yu, T.-Y., Kao, Y. & Lin, J.-J., 2014. Telomeric transcripts stimulate telomere recombination to suppress senescence in cells lacking telomerase. *Proceedings of the National Academy of Sciences of the United States of America*, 111(9): 3377–82.

Zeman, M.K. & Cimprich, K.A., 2013. Causes and consequences of replication stress. *Nature Cell Biology*, 16(1): 2–9.

Zhao, J., Bacolla, A., Wang, G. & Vasquez, K.M., 2010. Non-B DNA structure-induced genetic instability and evolution. *Cellular and Molecular Life Sciences*, 67(1): 43–62.

Zheng, L. *et al.*, 2009. MRE11 complex links RECQ5 helicase to sites of DNA damage. *Nucleic acids research*, 37(8): 2645–57.